

Scale development: Behavioural Couples Therapy Scale for Depression

Isabelle Rudolf von Rohr

June, 2017

Submitted in partial fulfilment of the requirements for the degree of Doctor in
Clinical Psychology (DClinPsy), Royal Holloway, University of London.

Acknowledgements

I would like to thank the many people who have helped me through the completion of my thesis. First and foremost, I would like to say a huge thank you to Dr. Kate Muse and Dr. Helen Pote for their professional and personal support and encouragement. I am particularly grateful to Kate for believing in me and giving me the confidence to do the same. I would also like to acknowledge with much appreciation the role of the BCT team: Dr. Sarah Corrie, Michael Worrell, Don Baucom and Melanie Fischer who made the research possible within their BCT postgraduate Diploma training course.

I am grateful to the many BCT supervisors and trainees who took time out of their busy schedules to participate in my research.

I am lucky to have had a very supportive cohort at University and a great deal of encouragement from my friends and family, which has gone a long way to helping me finish my thesis and keeping my sanity. My mum (Monique Rudolf von Rohr) and sister (Marie-Thérèse Rudolf von Rohr) for always being there to morally support me. And my dad (René Rudolf von Rohr) and my brothers (Tobias and Lukas Rudolf von Rohr) for all their humour and help wherever they could. I would also like to thank my niece and nephew for providing me with much needed distraction from time to time with beautiful and funny pictures. Last, but certainly not least, I am incredibly grateful to Mathew Mahon and his parents (Anita and Keith Blake) who made it possible for me to take the time I needed to do my work and for their patience, understanding and moral support. And finally thank you to my little boy Oliver Mahon, who's always brightened up my day and made me keep the bigger picture in mind.

Abstract

This study outlines an expert review and a contribution to scale evaluation for the further development of the Behavioural Couple Therapy Scale for Depression (BCTS-D). The BCTS-D aims to provide a novel assessment framework to deliver formative and summative feedback regarding therapists' performance within observed behavioural couple therapy (BCT) treatment sessions. The expert review and scale evaluation was conducted for two versions: the BCTS-D v1 (N = 14) with a BCT supervisor sample and the BCTS-D v2 (N =20) with a mixture of BCT supervisors and trainees within a BCT training context.

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. Scores on both versions of the BCTS-D demonstrate good internal consistency and overall inter-rater reliability, which were comparable to the Revised Cognitive Therapy Scale (CTS-R). However inter-rater reliability for single items is a weak point that will need exploring in future research. The studies also provided insight into areas for refinement and a number of modifications were undertaken to improve the BCTS-D v1. Additional modifications will be needed to respond to the feedback for the BCTS-D v2. In summary, the BCTS-D is an appropriate and useful measure of BCT competence that can be used to promote self-reflection and provide therapists with formative and summative feedback within a BCT training context.

List of Tables

Table 1: Overview of three key stages in scale development.....	31
Table 2: Participants' demographics - study 1.....	44
Table 3: Participants' demographics - study 2.....	54
Table 4: Overview of descriptive themes to improve and develop the BCTS-D v1	57
Table 5: Cronbach's alpha if item was deleted for the BCTS-D v1.....	68
Table 6: BCTS-D v2 compared to the CTS-R	72
Table 7: Content validity results for each item in the BCTS-D v2.....	76
Table 8: Usability results for the BCTS-D v2.....	80
Table 9: Qualitative feedback on the BCTS-D v2: themes identified to improve the overall scale	82
Table 10: Themes identified to improve the three domains.....	90
Table 11: Cronbach's alpha if item was deleted for the BCTS-D v2.....	104
Table 12: Analysis of inter-rater reliability for individual items within the BCTS-D v2.....	108

List of Figures

<i>Figure 1: A framework for therapist competence measures, based on Miller's (1990) clinical skills hierarchy (Muse & McManus, 2013)</i>	21
---	----

Table of Contents

Introduction	8
Overview	8
Evidence-based couple therapy	9
Importance of assessing competence within psychotherapy	11
What is competence?.....	14
Assessment of competence	18
Assessment of competence within couple therapy.....	22
Behavioural Couple’s Therapy (BCT).	22
Couple Therapy for Depression Competency Adherence Scale.....	23
Behavioral Couple Therapy Competence Rating Scale.	24
Summary.	25
Rational of development of BCTS-D.....	26
Scale development theory.....	28
Establishing validity and reliability of a new scale	31
Aims	32
Research questions	32
Methods	34
Overview	34
Ethical Considerations.....	35
Confidentiality.	35
BCTS-D version 1	36
Scale development.....	36
Expert Review.....	37
Scale Evaluation.	40

Scale development: BCTS-D

Procedure.....	42
Participants.....	43
BCTS-D v2	45
Information on Measure Development.	45
Expert review.	45
Scale evaluation.	47
Procedure.....	50
Participants.....	51
Results	55
BCTS-D v1	55
Expert Review.....	55
Scale Evaluation.	67
BCTS-D v2	69
Scale development.....	69
Expert Review.....	74
Scale evaluation.	103
Discussion	110
Summary of Results	110
BCTS-D v1.	110
Scale revisions.	112
BCTS-D v2.	113
Limitations and future directions.....	123
Clinical and Scientific implications	131
Conclusion	132

Scale development: BCTS-D

References 134

Appendices 149

Introduction

Overview

This study focuses on the development and piloting of a tool measuring therapist competence in delivering Behavioural Couple Therapy (BCT). Recent National Institute for Health and Care Excellence (NICE) guidelines for depression have included BCT as an evidence-based treatment (NICE, 2009) and it's being rolled out as part of the Increasing Access to Psychological Therapies program (IAPT; Clark et al., 2009). The success of such a dissemination program depends on effective training of BCT therapists to the point at which they are "competent" (Sharpless & Barber, 2009). Therefore, reliable and valid measures of therapist competence are essential to assess the training of new therapist and to monitor the quality of treatment provision within routine clinical practice (Fairburn & Cooper, 2011; Muse & McManus, 2013). Second, competence measures allow targeted feedback regarding a therapist's strengths and weaknesses which can be effective in improving competence (McManus, Westbrook, Vazquez-Montes, Fennell, & Kennerley, 2010; Muse & McManus, 2013). Third, competence assessment is essential to interpret outcomes of effectiveness studies. Fourth, there may be a relationship between treatment outcome and therapist competence. Hence, by assessing therapist competence it can be ensured that BCT is optimally effective (Muse & McManus, 2013). Yet no measure currently exists for BCT couple therapy. The measure was therefore developed to fill a gap in BCT research and training contexts.

The review of the literature provided in this chapter will consider the concepts and literature relevant to scale development in general and the

Scale development: BCTS-D

competence assessment tools within psychotherapy, and specifically within couple therapy. Initially, the importance of assessing therapist competence in delivering evidence-based psychological treatments such as BCT will be discussed. Second, the challenges of defining and measuring a construct such as 'therapist competence' will be examined. Then, existing competence measures within couple therapy will be reviewed and their relevance to measuring competence of BCT therapists will be explored. Finally, theoretical and methodological concerns which must be considered in the development of any novel measure such as reliability and validity will be reviewed. The chapter will conclude with a rationale for the current study and aims will be presented.

Evidence-based couple therapy

BCT has recently been included in the NICE guidelines for depression as an evidence-based treatment (2009) and within couple's therapy significant progress has been made in developing evidence-based psychological treatments for a variety of disorders and problems. Couple-based interventions such as Cognitive-Behavioural Couple Therapy (CBCT; Epstein & Baucom, 2002) have been tailored to target individual psychopathology (i.e. depression), while at the same time improving relationship distress. CBCT has evolved out of the mainly behaviourally based approach 'Traditional Behavioural Couple Therapy' (TBCT; Jacobson & Addis, 1993) and was modified to include cognitions and emotions alongside behaviours. CBCT and other behavioural couple therapies have been found more effective than wait list conditions and as effective as individual treatment of depression with the added benefit of improving

Scale development: BCTS-D

relationship adjustment when provided to martially distressed couples with a depressed partner (Baucom, Shoham, Mueser, Daiuto, & Stickle, 1998; Beach & Daniel O’Leary, 1992; Bodenmann et al., 2008; Fischer, Baucom, & Cohen, 2016; Gupta, Coyne, & Beach, 2003). Shadish and Baldwin (2003) found a mean effect size of 0.84 comparing couple therapy to no treatment across all couple therapies when conducting a meta-analysis. Studies comparing different couple therapies have found very similar outcomes with regards to the effectiveness of the different couple-based approaches (Gurman, 2015). A meta-analysis by Barbato & D’Avanzo (2008) suggests that couple therapy appears to be comparable to individually oriented treatment in reducing depressive symptoms ($d = -.12$) and more effective than individually oriented treatment in improving relationship satisfaction ($d = -.60$), thereby addressing a risk factor for relapse of depression. The American Psychological Association Division 12 Task Force (Chambless & Hollon, 1998) have listed BCT as a ‘well established’ treatment intervention and in the U.K. the NICE guidelines for depression have included BCT as an evidence-based treatment for depression (2009).

Therefore, couple therapy for depression is recommended as part of the stepped care programme within the Increasing Access to Psychological Therapies programme (IAPT; Clark et al., 2009). There are two strands of couple-based interventions that were chosen to provide empirically supported interventions to the public as part of IAPT (Clark et al., 2009). One is Cognitive Behavioural Couples Therapy (CBCT, referred to as BCT

Scale development: BCTS-D

throughout this report¹), which is the main focus of this thesis, (Baucom, Epstein, Kirby, & LaTaillade, 2015) and the other is an integrative approach to Couple Therapy for depression (Hewison, 2011). Both approaches have evolved beyond the traditional Behavioural Couples Therapy (TBCT) and are based on the evidence gathered in the NICE database (2009). As part of IAPT, BCT effectiveness will be routinely evaluated. It's a high intensity approach, denotes formal face-to-face psychological therapy delivered by a relatively specialist psychological therapist and lasts between 15 to 20 sessions over 5 to 6 months, and offers an alternative or supplement to established individual treatments such as cognitive behaviour therapy (CBT) or medication.

Importance of assessing competence within psychotherapy

Assessing therapist competence in delivering BCT is important to the continued progression of the field for several reasons. First, the success of dissemination programmes like IAPT (Clark et al., 2009), which purpose is to increase access to evidence-based treatment, depends on effective training of BCT therapists to the point at which they are “competent” (McHugh & Barlow, 2010; Rakovshik & McManus, 2010; Sharpless & Barber, 2009). Achieving a certain competence level as a therapist is the inherent goal of training therapists (Rakovshik & McManus, 2010). Assessment of training progress usually examines both didactic knowledge and skills related to

¹ The terms CBCT and BCT are often used interchangeably. For this study, the term BCT will be used, rather than CBCT, as the research is carried out within the Central London training centre for BCT in England.

Scale development: BCTS-D

competencies using validated competency measures (McHugh & Barlow, 2010). In Clinical Psychology training programmes, for example, a key feature of the new accreditation standards is doing in vivo assessments of therapists' competence (British Psychological Society, 2010) and other courses such as CBT training courses routinely assess therapist competence as part of the accreditation process. Such measures of competence are therefore essential to determine effectiveness of training (Blackburn et al., 2001). Especially since effectiveness of evidence-based treatment protocols (e.g. CBT or BCT) are not guaranteed when implemented in a routine setting and taken out of the research environment (Rakovshik & McManus, 2010). Elkin (1999) highlighted the fact that different emphasis on training, supervision and therapist competence might impact service delivery and effective practice. This means that implementing an evidence-based therapy could be undermined if the training and supervision components are neglected and therapist competence is not guaranteed (Clulow, 2010b). Furthermore, competency based ratings provide a framework for delivering formative feedback about a therapist's strengths and weaknesses, promote on going self-reflection of practitioners (Muse & McManus, 2013) and guide future learning (Bennett-Levy, 2006), which is useful for personal development, supervision and examining skill acquisition.

Second, in recent years awareness has increased that therapist competence may be important to psychotherapy outcomes (Barber, Sharpless, Klostermann, & McCarthy, 2007; Blow, Sprenkle, & Davis, 2007; Davidson et al., 2004; Elkin, 1999). However, research on how therapist competence influences treatment outcome has yielded inconsistent findings.

Scale development: BCTS-D

Delivering individual CBT with higher levels of competence has been associated with a greater amount of change in clients' depression, anxiety and global clinical symptoms (Davidson et al., 2004). In addition a relationship between client outcomes and therapist competence when treating depression with CBT was reported, as long as the therapist adhered to the treatment protocol (Shaw et al., 1999). Within behavioural marital therapy, therapists, who had couples who responded positively within treatment, were asked to rate themselves on a checklist. The majority of therapists rated themselves as being effective in inducing collaboration (Holtzworth-Munroe, Jacobson, DeKlyen, & Whisman, 1989). This indicates that therapist competence is also important to outcomes within couple therapy. However, there is surprisingly little research regarding therapist variables as contributors to outcome, especially in the area of marital and couple therapy (Blow et al., 2007). Most recent research has focused on the relationship between CBT competence and outcome for individual treatment for depression (Davidson et al., 2004; Shaw et al., 1999; Strunk, Brotman, DeRubeis, & Hollon, 2010; Trepka, Rees, Shapiro, Hardy, & Barkham, 2004). However, associations between outcome and competence vary from no relationship (Webb, DeRubeis, & Barber, 2010) to a strong relationship ($r=-.47$; Trepka et al., 2004). Possible explanations suggested for such varied results are not considering patient variables or therapeutic alliance, having a restricted range of competence amongst therapists within a research trial or poor reliability of existing methods measuring competence (Crits-Christoph et al., 1991; Webb et al., 2010; Whisman, 1993). Therefore, reliable and valid measures of therapist competence are essential to assess the training of new therapist and to

Scale development: BCTS-D

monitor the quality of treatment provision within routine clinical practice to ensure treatment is optimally effective for patients (Muse & McManus, 2013).

The BCT training course, therefore, needs to ensure therapist competence and include a formal competence rating within the accreditation process.

Third, therapist competence assessment is essential to interpret the outcomes of effectiveness studies. For 'between-model' comparisons, competence must be equated across the two treatments to be able to draw valid conclusions regarding the efficacy of specific treatments. Competence cannot be assumed on the basis of the experience level of the therapist (Jacobson & Addis, 1993). Elkin (1999) suggests that reporting the amount of variance within treatment outcome due to therapist competence, ensures that therapist variables are considered in reports of treatment efficacy. Hence, improved assessment of therapist competence may improve much needed future research examining the association between therapist competence and outcome and could provide insight into the 'active ingredients of a therapy model' (Dobson & Singer, 2005).

Despite the importance of measuring therapist competence in psychotherapy, there is a lack of consensus regarding what is meant by competence or how to measure it (Elkin, 1999) and is often neglected in scientific practice (Weck, Bohn, Ginzburg, & Stangier, 2011).

What is competence?

The Oxford dictionary defines competence as 'the ability to do something successfully or efficiently' ("oxford dictionary," n.d.). Within the psychotherapy literature definitions of therapist competence have included a broad range of concepts.

Scale development: BCTS-D

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). Kaslow (2004) discusses eight different domains of psychologist competencies (e.g. assessment, intervention, ethics, supervision, etc.). Competence within the intervention domain includes global therapeutic knowledge and skills (i.e. therapist's ability to independently assess patient's well-being; manage many clinical problems) and limited-domain competency which is specific knowledge and skills according to the therapeutic domain (Barber et al., 2007; Sharpless & Barber, 2009; Waltz et al., 1993). This makes it necessary to define model-specific competencies (Kazantzis, 2003; Sharpless & Barber, 2009). This is particularly important since what might be considered competent for one psychotherapeutic domain may be classed as incompetent or even unethical in another (Sharpless & Barber, 2009). Roth and Pilling (2007) were the first to map 'core competencies' required to deliver effective CBT. The framework includes over 50 competencies and is grouped into five domains: generic competences (e.g. engaging and relating to a client), basic CBT competencies (e.g. knowledge of CBT principles such as use of homework), specific CBT competences (e.g. Socratic questioning, exposure techniques), problem-specific competences (e.g. disorder-specific interventions) and metacompetences (e.g. procedures used to guide practice). The framework is restricted to the diagnosis of depression and anxiety disorder and for people at working age. Roth and Pilling (2008) argue that the framework could easily be extended to other disorders, but are uncertain how it would apply to different client populations (e.g. children,

Scale development: BCTS-D

couples). The authors note that the framework is not feasible to assess competence in its current form (i.e. over 50 core competencies), but suggest that therapists will be rated on a chosen subset of core competencies (Roth & Pilling, 2008). Several well-validated instruments for assessing adherence and competence in CBT already exist (e.g. Cognitive Therapy Scale – Revised; Blackburn et al., 2001) and map onto a subset of those competencies.

Further, an important distinction between competence and treatment adherence is suggested, with adherence demonstrating that the therapist delivers the intervention as outlined using the right psychotherapeutic procedures and competence referring to the skill with which the intervention was implemented (Barber et al., 2007; Fairburn & Cooper, 2011; Waltz et al., 1993). Adherence is a prerequisite for competence, however adherence by itself does not necessarily imply competence (Muse & McManus, 2013). Fairburn and Cooper (2011) suggest abandoning these two distinctions when assessing the overall standard of treatment provided within routine care due to a lot of overlap. However, within research (i.e. treatment trials) both concepts are usually considered separately and frequently checks of therapists' adherence to a manual are made to minimise therapist variability (Elkin, 1999).

Within couple therapy the Department of Health commissioned a research group to define specific competences required to deliver effective couple therapy for partners with depression (Clulow, 2010a). The competency model was based on the CBT competence model (Roth & Pilling, 2007; Roth & Pilling, 2008) and brings together the competences and techniques

Scale development: BCTS-D

identified from a range of manuals that are evidence-based and likely to be effective in treating depression (Clulow, 2010b). The framework has the same five domains as Roth and Pilling's model (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. behavioural couple therapy) and metacompetences (e.g. using clinical judgement when implementing the therapy). The framework describes activities to carry out couple therapy for depression effectively, but does not prescribe exactly what to do (Clulow, 2010b). The separate competencies are mainly theory-based at this point and more research is needed determining what sub-competencies are considered important for treatment success (Dobson & Singer, 2005). The next step is to develop and refine appropriate psychometric measures of couple therapist competence (Clulow, 2010b).

Finally, another conceptual issue in defining therapist competence is whether competence is a trait-like construct or if it varies across time and situation (Dobson & Singer, 2005). Kaslow (2004) states that competence is context dependent and developmental, which means that competence varies depending on the setting, the environment and the individual's stage of professional functioning (i.e. trainee versus expert). If competence is seen as variable and context-dependent, even experienced clinicians need continuous supervision and competence assessments (Kazantzis, 2003). Some experts in the field have even warned of the potential for 'therapist drift' if a therapist does not sufficiently monitor and maintain their clinical knowledge and skills

Scale development: BCTS-D

over time (Waller, 2009).

Competence within therapy is multi-dimensional and even though there is some agreement of what it consists of, more systematic efforts are needed to ascertain what key sub-competencies are important for effective treatment (Kaslow, 2004) and how competencies vary for different settings and environment (Roth & Pilling, 2008). Clulow (2010a, 2010b) has taken the step to start filling the gap for couple therapy for depression.

Assessment of competence

Miller's (1990) hierarchy of clinical skills provides a useful framework for therapist competence assessments and proposes four different levels of clinical skill. Muse and McManus (2013) have used this framework and reviewed the various assessment methods of competence within Psychology within each level of clinical skill (see figure 1). Level 1 is "knows" and refers to conceptual knowledge of a therapist and is assessed through multiple-choice questions or essays. Level 2 "knows how" refers to practical knowledge of how to apply theory, which can be assessed by short answer clinical vignettes or case reports. Level 3 "shows how" refers to competence in demonstrating the ability to apply skills in clinical situations and can be assessed through standardised role-plays. Level 4 "does" refers to how therapists apply skills in clinical practice, which reflects therapist quality and is typically assessed through rating treatment sessions.

Muse and McManus (2013) suggest that several levels of those clinical skills within Miller's framework need to be considered when assessing competence within CBT. The use of multi-method approaches is however costly and time-consuming (Perepletchikova, Hilt, Chereji, & Kazdin, 2009)

Scale development: BCTS-D

and the most commonly used approach within routine practice is ratings of therapists' in session performance by an observer (Muse & McManus, 2013) using a competence rating scale. However, there is no consensus regarding how to assess competence in psychotherapy or in CBT. Muse and McManus (2013) highlight that standardised scales are useful to establish whether therapists have a recognised standard of competence, providing detailed feedback on a therapist's strengths and weaknesses and allowing comparison between different training courses or research trials. However, many standardised scales have issues undermining their utility (Barber et al., 2007; Muse & McManus, 2013; Waltz et al., 1993). There are either trans-diagnostic scales focusing on competence which underpin most CBT interventions, which might neglect key competencies for certain disorders (e.g. exposure technique for phobia), or disorder-specific scales, which focus on competence required to deliver disorder-specific protocols, but might put less focus on more generic CBT skills (i.e. Socratic questioning; Muse & McManus, 2013). The biggest drawback of many scales is that the ability to provide reliable and valid measures of competence outside of controlled research settings has not yet been established. Inter-rater reliability is often poor and is associated with difficulties conceptualizing competence and striking the right balance of the number of items included in the scale. Further, the used competence threshold scores are usually arbitrary and not empirically grounded (Muse & McManus, 2013). Other issues concern the implementation of these scales such as what level of training is needed to use a scale or how many ratings are needed to reliably assess a therapist's competence (Muse & McManus, 2013). All these issues make it challenging

Scale development: BCTS-D

to choose assessment methods that are feasible and reliable within the resource constraints in routine clinical practice and need to be considered when developing a competency scale within BCT.

Scale development: BCTS-D

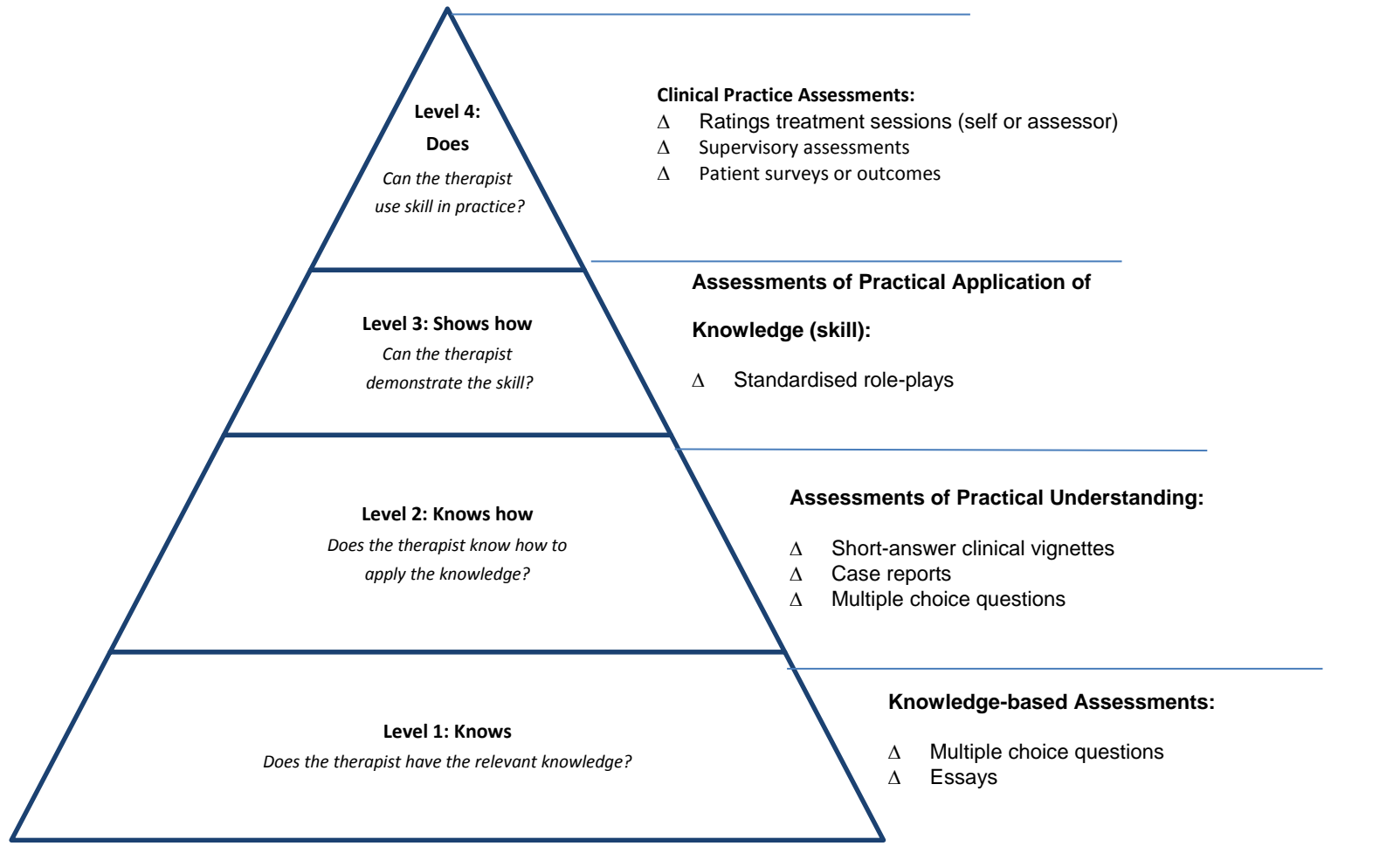


Figure 1: A framework for therapist competence measures, based on Miller's (1990) clinical skills hierarchy (Muse & McManus, 2013)

Assessment of competence within couple therapy

Within couple therapy, competency rating scales are scarce. This section explores what competence rating instruments already exist and if they could feasibly be used within a BCT training context. First, the BCT approach and its techniques will be described, followed by the competence scales' descriptions and purposes to see if they fit a BCT training context assessing therapist competence.

Behavioural Couple's Therapy (BCT). The goals of traditional BCT are to promote positive change in couples through direct instruction and skill training (Christensen et al., 2004). During treatment, the therapist relies on three primary treatment strategies: behavioural exchange, communication training, and problem-solving training. BCT incorporates these aspects to treatment, but has evolved way beyond it over the last two decades. The two major influences are Cognitive Psychology and Social Cognition research as well as integrating core themes the couple brings (e.g. desired levels of closeness), and considering influences of personality and the environment (Baucom et al., 2015). The principles and techniques employed in BCT can be adapted and broadened to assist couples who have one member of the couple experiencing some form of psychopathology, such as clinical depression (Baucom, Whisman, & Paprocki, 2012; Bodenmann et al., 2008). In the UK, the focus is on BCT therapists assisting couples who have the comorbid conditions of relationship distress and depression. Initially, the therapy aims to enhance relationship well-being. This is then supplemented by an understanding of clinical depression, how it exists in an interpersonal context, and how the couple's relationship can be employed in the alleviation

Scale development: BCTS-D

of depression, drawing from cognitive behavioural theory and research on individual treatment of depression (Baucom & Boeding, 2013). In an extensive literature search, only two competency rating scales for a couple therapy context were found (see Appendix A).

Couple Therapy for Depression Competency Adherence Scale.

One scale is called the 'Couple Therapy for Depression Competency Adherence Scale' and it was developed by the Tavistock Relationship workgroup who deliver couple therapy training within IAPT (Hewison, 2011). This scale is used for supervisors to rate a trainee therapist against a list of competencies for couple therapy for depression and can also be used as a self-assessment tool by practitioners. The scale only covers the 'specific couple therapy techniques' domain out of the five competence domains described in Clulow's (2010a) framework. The scale includes a comprehensive list of techniques used across all couple therapy for depression approaches (Clulow, 2010b). The scale includes nine domains (e.g. techniques that engage the couple) and amounts to 41 items (e.g. an ability to form and develop a collaborate alliance with the couple as a unit). Each item is rated on a scale from 0 – 4 (i.e. 0 = not present, 1 = possibly present, 2 = briefly present, 3 = moderately present, 4 = extensively present). At the end of the scale each trainee needs to fill in five questions justifying their choice of technique selection (e.g. technique chosen, what is the fit with your basic couple training?). Not all competencies are expected to be observed during one session "as some relate to stages of therapy and some are mutually contradictory because of the differences between the therapeutic interventions found in the evidence base" (p.56; Hewison, 2011). The

Scale development: BCTS-D

Tavistock team have set the standard of acceptable competence at 80% of the list. This threshold was chosen by the Tavistock team and is intended to reflect that the practitioners will not use all the listed techniques, which can come from different, even contrasting, modalities (Hewison, 2011).

The 'Couple Therapy for Depression Competency Adherence Scale' is very long consisting of 41 items. However, it only covers one out of the five suggested competence domains (i.e. specific couple therapy techniques; Clulow, 2010a) and covers all evidence-based couple therapies for depression. This would suggest the assessor needs knowledge in all the different approaches to couple therapy to be able to rate the trainee in a reliable manner. Using this scale would not be feasible within a BCT specific training course assessing BCT competence. Finally, no psychometric data could be found on how the scale was developed or on how reliable and valid the scale is.

Behavioral Couple Therapy Competence Rating Scale. The second scale found is the 'Behavioral Couple Therapy Competence Rating Scale' (Jacobson, Christensen, Prince, Cordova, & Eldridge, 2000). This scale was used as part of two larger RCTs (Christensen et al., 2004; Jacobson et al., 2000) comparing traditional behavioural couples therapy (TBCT; Jacobson & Margolin, 1979) and integrative behavioural couple therapy (IBCT; Christensen & Jacobson, 1998). This scale solely assesses TBCT competences and was developed for those research trials to ensure that the therapist did not display a new treatment bias by being less enthusiastic and less competent delivering TBCT. The competency scale describes 10 skills essential to good TBCT (see Appendix B). 9 items (e.g. agenda, structuring,

Scale development: BCTS-D

etc) are rated on 6-point scales (0 – poor, 2 – mediocre, 4 – good, 6 – excellent) and 1 is rated on a 12-point scale (i.e. implementation of techniques), because it is deemed more important, resulting in a maximum score of 66. A score of 40 indicates an average rating of "good" and is considered the cut-off for competent performance of TBCT. The scale has short anchor points describing the score on the scoresheet, but no accompanying manual describing how to use the scale and only a specialist consultant observing several of the therapist's sessions is able to judge if competency is achieved on this scale. In this trial, the expert rater viewed the initial four sessions for case conceptualization purposes and then completed competency ratings on all the completed sessions to ensure the expert could see the continuity between sessions (Christensen et al., 2004; Jacobson et al., 2000). Within routine practice a more practical approach is needed, allowing to measure competence of one single therapy session and allowing others than specialists within the field to use the scale. Further, this scale does not cover all aspects of BCT (Epstein & Baucom, 2002). Finally, no psychometric data on the scale's validity or reliability was found. Adherence to the treatment manual was rated separately.

Summary. 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 requires a unique measure capturing the general skills of couple therapy, 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

Scale development: BCTS-D

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 & McManus, 2013).

Rational of development of BCTS-D

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. No previously developed scale was found to be fit to assess BCT competence specifically and therefore an adequate competence measure needed to be developed. As the competency framework for couple therapy for partners with depression (Clulow, 2010a) was based on the CBT competence model (Roth & Pilling, 2007; Roth & Pilling, 2008) and due to the limited research on competence assessments for couple-based interventions, it was decided to focus on previous research of competence assessments within a CBT model to inform the development of a BCT competency measure. Competence in CBT involves both adherence to the model as well as skilful application of treatment methods in caring for patients. Despite certain limitations, which are discussed below, the Cognitive Therapy Scale-Revised (CTS-R; Blackburn et al., 2001) is the most widely used tool for measuring CBT competence with adults (Keen & Freeston, 2008; Rakovshik & McManus, 2010) and considered to provide a comprehensive overview of the generic skills required to competently practise CBT with adults (Keen & Freeston, 2008). Further,

Scale development: BCTS-D

BCT therapists, as a BCT course requirement, have already been trained in CBT with a lot of clinical experience, and are therefore already familiar with its use and purpose. Thus, to avoid high cost while developing a new competence assessment measure for BCT and make it a time and labour efficient tool, it was decided to build upon and be consistent with the CTS-R (Blackburn et al., 2001), while adapting it to a couple's context. The CTS-R has on a couple of occasions been successfully adapted for different client groups such as people suffering with psychosis or children and adolescents (Haddock et al., 2001; Stallard, Myles, & Branson, 2014). To help avoid and improve on the known criticism of the CTS-R, its psychometric properties were explored in more detail before developing the new competence measure.

The CTS-R is a revised version of the original Cognitive Therapy Scale (CTS; (Vallis, Shaw, & Dobson, 1986; Young & Beck, 1980). The CTS has demonstrated a high degree of internal consistency (α range = .84 - .95 [Dobson, Shaw, & Vallis, 1985]). Typically, an α range=.80 - .85 represents satisfactory agreement (Bland & Altman, 1997). Factor analysis revealed only one major factor with most of the items loading highly on it (64.8% of variance, Vallis et al., 1986). This factor reflected overall cognitive therapy quality, composed of both nonspecific factors and specific cognitive therapy factors (Vallis et al., 1986). Hence the CTS appears to measure one construct, with rationally defined subscales which are not independent. The CTS-R was reported to be internally consistent as well (α range= .75–.97; Blackburn et al., 2001) and appears to measure one construct like the CTS. However, some might interpret this as certain items being redundant or

Scale development: BCTS-D

multiple concepts being addressed by single items (Whisman, 1993) and therefore, undermining the content validity of the scale. Further, the developers report good inter-rater reliability for the initial 13-item CTS-R (average ICC=.63; Blackburn et al., 2001). An effect of training on the raters using the CTS-R was found increasing correlations between raters from moderate without rater training ($r=.44$) to good following rater training ($r=.67$; Blackburn et al., 2001). Overall, the training of supervisors on the CTS-R leads to a significant improvement of their inter-rater agreement (Reichelt, James, & Blackburn, 2003). However, the inter-rater reliability for individual items ranged from $r=.84$ (ICC for a pair of raters) to $r=-.14$ and remained poor for some items even following rater training (range $r=.26-.62$; Reichelt et al., 2003). Other limitations to the CTS-R reported are its arbitrary cut-off for competence (a score above 39 is commonly considered a minimum competency score with every item needing a score of at least 2, but has not been validated; Muse & McManus, 2013). Despite these limitations, the CTS-R (Blackburn et al., 2001) is the most widely used tool for measuring CBT competence with adults (Keen & Freeston, 2008; Rakovshik & McManus, 2010) and all BCT therapists are, per course requirement, already familiar with its use and purpose.

Scale development theory

The development of the BCTS-D follows the Classical Test Theory (CTT) approach outlined by DeVellis (2012). The development of any measure is a process of finding indicators describing a theoretical construct and thereby operationalizing this theoretical construct that can't directly be observed or assessed (Barker et al., 2002; DeVellis, 2012). CTT and item

Scale development: BCTS-D

response theory (IRT) are the most commonly used psychometric approaches. CTT is trying to capture an underlying construct, *latent variable*, by measuring an observed score. The observed score is the result of a respondent's true score in addition to a single error term (DeVellis, 2012). IRT, on the other hand, differentiates between different types of error, particularly error with respect to item characteristics. IRT is therefore more focused on individual items whereas CTT is more concerned with the overall item score. IRT methods are demanding, need a higher degree of expert judgment and are still in its active stage of development. To ensure that developed items are being assessed independently of the characteristics of the sample being studied, IRT requires repeated testing of large and heterogeneous samples (Barker et al., 2002; DeVellis, 2012). Therefore, in most cases, CTT is considered the favourable approach except for the development of scales with inherently hierarchical items or scales concerned with differential item functioning, and used to develop the BCTS-D (DeVellis, 2012). This approach involves three key stages: (1) scale creation, (2) expert review, and (3) scale evaluation (DeVellis, 2012). However, scale development is an iterative process with using data collected in later stages to improve certain previous steps (DeVellis, 2012).

The initial process of scale creation can be further broken down into three steps (see in Table 1): step one- define scale scope (i.e. describe the construct the scale aims to measure), step two- create an item pool (i.e. develop theory-driven scale items), step three- create a response scale (i.e. decide if items should be scored on a continuum). A BCT expert reference group conducted the first stage of scale creation, which is described in more

Scale development: BCTS-D

detail in the methods section. Stage two includes step four of reviewing the item pool by experts (i.e. check the validity and usability of the scale and the items). Finally, the third stage can be broken into two further steps: step five – administer items to a development sample (i.e. get an idea of how the scale will perform), and step six - evaluate the items (i.e. examine the scale's reliability).

Scale development: BCTS-D

Table 1:

Overview of three key stages in scale development

Stage 1	Define scale scope
	Create an item pool
	Create a response scale
Stage 2	Reviewing item pool by experts
Stage 3	Administer items to a development sample
	Evaluate the items

Establishing validity and reliability of a new scale

Within CTT, reliability (i.e. reproducibility of measurement) and validity (i.e. meaning of measurement) are specific concepts used to evaluate developed measurement instruments. Reliability is an essential concept in scale development and refers to the degree of reproducibility of the measurement (Barker et al., 2002). DeVellis (2012, p.27) describes reliability as “the proportion of variance attributable to the true score of the latent variable”. There are different ways to assess reliability depending on the type of measure (e.g. self-report or observation), but they share that fundamental definition. Validity is commonly described as “whether a test measures what it is intended to measure” (Howitt & Cramer, 2011, p.272). Whereas reliability indicates whether a scale measures in a consistent way, validity indicates whether the scale measures the underlying construct (e.g. does a depression scale truly measure depression). However, a reliable scale does not guarantee validity. There are three main approaches of how to measure validity of a scale: content validity, criterion validity and construct validity

Scale development: BCTS-D

(DeVellis, 2012). How these concepts apply to the current project is explored further within the methods section.

Aims

An expert reference group has adapted the CTS-R (Blackburn et al., 2001) and drafted an observation-based rating scale, Behavioural Couple Therapy Scale for Depression (BCTS-D, Corrie, Fischer, Worrell, & Baucom, n.d.), which is intended be used to provide formative and summative feedback regarding therapists' performance within observed BCT treatment sessions. The aim of this project is to develop and improve the scale, and to use a pilot project to determine the psychometric properties of the BCTS-D. The objectives to achieve these study aims are: i) usability study to determine if the use of the BCTS-D v1 (based on the CTS-R) is feasible within a couple context and to further develop and improve the scale (BCTS-D v2) ii) Assess and improve content and face validity of the second version using qualitative and quantitative data. The process by which these aims were carried out will be discussed in the following chapter

Research questions

1.1: What is the BCTS-D v1's reliability (internal consistency and inter-rater reliability) and is it comparable to the CTS-R?

1.2: What is the BCTS-D v1's validity and usability?

1.3: Explore the qualitative feedback about the scale (written feedback) to further develop and improve the scale.

2.1: What is the BCTS-D v2's reliability (internal consistency and inter-rater reliability) and is it comparable to the CTS-R?

2.2: What is the BCTS-D v2 validity and usability?

Scale development: BCTS-D

2.2: Explore the qualitative feedback (written and oral feedback) to further develop and improve the scale.

Methods

Overview

The development of the BCTS-D, a rating scale assessing therapists' competence in delivering BCT, consists of three different stages: (1) scale creation, (2) expert review and (3) scale evaluation. The development of a rating scale should be an iterative process that employs a range of qualitative and quantitative techniques to test and revise scale items (Brewer & Hunter, 2006; DeVellis, 2012). The first stage – scale creation – was a precursor to the current study and included a BCT expert group adapting the CTS-R (Blackburn et al., 2001) for use within the context of cognitive therapy for couples (see below). This study was led by a researcher independent from the BCT team and conducted the second stage of measure development - expert review - and contributed to the third stage - scale evaluation by conducting an assessment of the BCTS-D psychometric properties. The study therefore used a mixed-methodology design aiming to further develop the BCTS-D.

This project did stage 2 - an expert review - and stage 3 - scale evaluation – of measure development twice, one for each version of the BCTS-D (v1 [see Appendix D] and v2 [see Appendix F]; for an overview of the studies and the analyses planned see Appendix C). Additionally, feedback on the BCTS-D v1's usability and relevance within a couple context was collected. The psychometric and usability data were used to make further refinements and improve the scale. After the first study the scale was revised in consultation with the expert group using the quantitative and qualitative data. Then the BCTS-D v2 was examined. Analysing the psychometric

Scale development: BCTS-D

properties of a rating scale is useful in highlighting any items which require modification or omission (DeVellis, 2012) and thus identifies any further revisions that are required.

Finally, the BCTS-D v2's usability and utility was examined by obtaining in depth qualitative feedback from people who have experience of using the scale in practice (i.e. individual written feedback and oral feedback within a focus group).

The study recruited the participants and collected its data from the postgraduate diploma course in BCT at the Central London training centre for CBT, a course accredited by the British Association for Behavioural Cognitive Psychotherapy (BABCP; Central London CBT Training Centre, 2016). The competence of the BCT trainee therapists' (referred to as trainees throughout this report) is routinely assessed within the BCT training course on three different occasions during their year-long training. Each trainee has to submit three 60-minute long video tapes of different treatment sessions. The video tapes have to be of different clients at various stages of their treatment (assessment and ending session excluded).

Ethical Considerations

Ethical permission to collect data within the Central London Training Centre was obtained from the Health Research Authority (HRA, see Appendix Q) and the Royal Holloway University of London Ethics Committee (see Appendix R).

Confidentiality. Standard procedures to ensure confidentiality and anonymity were followed in the study. The study had no access to any patient data and each participant was allocated a unique number, ensuring that all

Scale development: BCTS-D

materials related to the participation (e.g. completed questionnaires) do not relate to any personal information. All data are stored in a locked filing cabinet and all computerised data are stored on an encrypted and password protected USB stick. For audit purposes, all data for the study will be securely stored for 5 years and will be destroyed after this time. Data will be accessed only by members of the research team. However, individuals from Royal Holloway University of London and other regulatory authorities may require access to relevant data for audit and monitoring purposes only. Personal identifying information are stored only in the form of consent forms and are kept separate in a locked cupboard. These will be stored for two years and then destroyed.

BCTS-D version 1

Scale development. An expert group consisting of the founder of BCT (Baucom; Epstein & 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 v1 (Corrie, Fischer, Worrell, & Baucom, n.d.; see Appendix D). The CTS-R was modified to be appropriate for rating the therapist's degree of competence in BCT working within a couple and depression context. The expert reference group focussed the measure on clients suffering with depression as the evidence-base for BCT is strong for this area of pathology (Whisman, 1993, 1999, 2001; Whisman & Baucom, 2012). Each item was rephrased to fit a couple's setting (e.g. assesses whether the therapist encouraged both members of the couple to participate appropriately, rather than just an individual patient) and one item (focus on depression in context) was added capturing the therapist's

Scale development: BCTS-D

focus during the treatment session (i.e. either on patient's depression or on the couple's relationship [see item for item comparison with CTS-R in Appendix E]). Like in the CTS-R each item was described in more depth on the scoresheet describing what should be rated as part of that item. Further, each item is rated on a 7-point Likert scale like in the CTS-R and based on Dreyfus and Dreyfus (1986) defining development of competence (0-1: incompetent, 1-2: novice, 2-3: advanced beginner, 3-4: competent, 4-5: proficient, 5-6: expert). However, in this first version of the BCTS-D no in-depth descriptions of the anchor points of each competence level were developed. The total BCTS-D v1 score came to 78 and the overall passing threshold was set at the same level as the CTS-R (i.e. a total score of 50% or more). Finally, the BCTSD-v1 was like the CTS-R designed to assesses both audio and video recordings of active-treatment sessions (i.e. excluding assessment and ending sessions). At this stage, the items were briefly described on the scoresheet but no comprehensive manual was created yet.

Expert Review.

Measures. Participants were asked to complete a short feedback questionnaire (see Appendix J) about the scale to assess aspects of the face and content validity and the usability of the scale. In the feedback questionnaire participants were asked whether there were any important aspects of BCT competence they felt were missing from the scale (yes / no), if the scale gave ample opportunity for feedback (yes / no) and to rate the overall style, appearance and layout of the scale (1- *poor*, 2- *fair*, 3- *good*, 4- *very good*) and how easy they thought the scale was to use (1- *not easy*, 2- *somewhat easy*, 3- *quite easy*, 4- *very easy*). Participants were also provided

Scale development: BCTS-D

with space for qualitative feedback after each question to state what revisions could improve the scale.

Analysis Plan and data management.

Validity and usability. Content validity assesses whether the scale covers the relevant characteristics of the construct being measured. Lynn (1986) has described a quantitative approach using expert feedback to determining content validity, which will be discussed below. Examining criterion validity is beyond the scope of this study. Concurrent validity cannot be established as no appropriate measure with established reliability and validity (gold standard) currently exists and which is why the BCTS-D was developed. Finally, construct validity explores whether the patterns of relationship the scale has, is consistent with the theoretical expectations (Barker et al., 2002). The construct validity of the BCTS-D cannot currently be used as an approach to exploring its validity as this is outside the scope of this study. Another form of validity which was considered within this study however is face validity. This reflects a lay person's acceptance of an instrument, rather than it being a true psychometric assessment technique. This will be explored asking people using the scale if it appears to be sound and relevant (Lynn, 1986).

The validity and usability questions from the questionnaire were represented by calculating either the percentage of a certain response (yes/no questions) or the mean and standard deviation for questions rated on a scale from 1 - 4.

Qualitative analysis. The written feedback provided by participants in the questionnaire was analysed qualitatively using thematic analysis.

Scale development: BCTS-D

Thematic analysis is a widely used qualitative analytic method within Psychology, which identifies, analyses and reports themes within data (Braun & Clarke, 2006). Thematic analysis was chosen because it is flexible and can be used as a realist method (i.e. descriptive approach) to describe and reflect the reality of participants (Braun & Clarke, 2006). Other approaches such as Interpretative Phenomenological Analysis (IPA; (Smith, 2015) and Grounded Theory were considered (GT; Strauss & Corbin, 1997). However, the aim of the analyses was neither to develop a theoretical framework (GT) nor to explore personal experiences (IPA), but to identify and describe common strengths and weaknesses of the scale, and potential revision points to the BCTS-D. Therefore, thematic analysis with a pragmatic approach to the extraction of themes which could determine themes across participants was deemed most appropriate. This thematic analytic process followed the guidelines outlined by Braun and Clarke (2006) and 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. First, the questionnaire feedback provided by each participant was read in detail, to provide familiarity with the data. Second, initial codes were generated. This was achieved by summarising the key issues highlighted in each comment provided in the feedback questionnaires (where a comment involved multiple issues, each issue was outlined separately). Third, codes were collated into themes (i.e. codes with similar meanings were combined to form overarching themes). The analytic process 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

Scale development: BCTS-D

themes or to relate these themes to previous literature), as the intention of the analysis was simply to identify 'surface level' meaning of the participants' comments; i.e. problems and suggested scale improvements (Braun & Clarke, 2006). Further, even if a code was only mentioned once, it was still included in the overall analysis. This was to ensure that no weaknesses or potential for improvement was missed, which could be essential to further develop a scale. The same process was followed by a second rater, Dr Kate Muse, doing an independent validation of the analytic process. Thereafter analyses were compared and any discrepancies were discussed and common ground was found by reviewing the data and explaining the reasoning behind the coding made. Further, the qualitative data, coding and emerging themes were independently reviewed and judged acceptable by the researcher's academic supervisor.

Scale Evaluation.

Measures. Participants were asked to provide demographic information (i.e. age, gender, profession, experience with the CTS-R, in CBT and BCT; see Appendix I). Then the BCTS-D v1 (see Appendix D) was used for everyone to rate an audio recording of a therapy session individually.

Analysis Plan and data management.

For all statistical calculations, the programme IBS SPSS Statistics 21 was used.

Internal consistency. Internal consistency is usually measured using Cronbach's alpha (α) and ranges from 0 = *items independent* to 1 = *items identical*, with a minimum of .80 - .85 typically representing satisfactory agreement (Bland & Altman, 1997; Cronbach, 1951). However, there needs

Scale development: BCTS-D

to be a balance between very high correlation (possibly indicating item overlap/redundancy) and low correlation (indicating the items do not assess the same underlying construct). When initially developing a measurement scale, a higher α is advisable ($>.90$ may be necessary) as this will allow for alpha to deteriorate somewhat when used in a new research context (DeVellis, 2012). Simply analysing alpha does not guarantee that all items underlie one single latent variable. Instead a factor analysis can be used to empirically determine the underlying constructs of a set of items (DeVellis, 2012). Doing a factor analysis was beyond the scope of this study as usually a sample of about 5 to 10 subjects per item is needed (resulting in a sample between 65 - 130 subjects; DeVellis, 2012). An assumption was therefore made that competence in BCT is a single latent variable underlying diverse skill. This decision was based on the fact that the Cognitive Therapy Scale (CTS; Young & Beck, 1980) revealed in a factor analysis that most items loaded highly on one major factor (64.8% of the variance). This factor was thought to reflect overall cognitive therapy quality (i.e. composed of nonspecific therapeutic factors and specific cognitive therapy factors; Vallis et al., 1986). Hence the CTS was found to measure one construct, with two rationally defined subscales which are not independent (i.e. general skills and cognitive therapy skills). The BCTS-D is based on the CTS-R and on the same underlying principles as CBT. The CTS-R (Blackburn et al., 2001) is a revised version of the Cognitive Therapy Scale (CTS; Young & Beck, 1980), which has demonstrated a high degree of internal consistency (α range = .92 - .97; Blackburn et al., 2001). Therefore, it was assumed that the BCTS-D is based on one underlying construct as well. Further, the Cronbach's alpha (α)

Scale development: BCTS-D

if the item was deleted was calculated for each item to explore if removing an item would increase internal consistency.

Inter-rater reliability. Inter-rater reliability refers to the extent to which the raters using the scale agree or co-vary with each other (Barker et al., 2002) and will be examined as the scale is an observer – based competence scale. The ‘gold standard’ approach to examining inter-rater reliability (IRR) is to calculate Intra class correlation coefficients (ICCs: Shrout & Fleiss, 1979). ICC is considered more suitable than Pearson’s correlations when several judges rate the same targets and where assignment of raters is arbitrary (Shrout & Fleiss, 1979). A two-way mixed effects approach ICC with absolute agreement is appropriate where several raters assess the same target.

Single items were not investigated at this point as the aim of this study was to check if using an adapted version of the CTS-R in a couple context has overall acceptable IRR. Calculating IRR for single items would not be very reliable due to not having a written manual to assist raters.

Procedure. The data collection happened as part of a BCT supervision workshop (for their continued professional development) and participants rated an audio recording of a therapy session on the BCTS-D v1 after being introduced to the BCTS-D v1. The chosen therapy session was a mid-treatment session (i.e. no assessment or ending session) with a couple of which one suffered with depression. The session was conducted by a trainee therapist allowing for a greater range of ratings to ensure reliability analyses were possible and not skewed (i.e. if an expert tape was rated there would be a high risk of a ceiling effect causing artificially high reliability). After using the scale, supervisors were asked for their feedback on the usability and

Scale development: BCTS-D

relevance of the BCTS-D v1 within the context of BCT. At the beginning of the training and data collection, all participants were informed about the research aspect and given an information sheet to read (see Appendix G). After having the opportunity to ask any questions they all signed a written informed consent form (see Appendix H).

Participants. The data was collected within a routine training workshop on “BCT supervision” organised by the London training centre for CBT as part of their continuing professional development for BCT supervisors. An email invitation was sent to 19 supervisors of which 14 supervisors (73.68%) attended.

All supervisors were BABCP accredited CBT therapists who had been practicing CBT for between two and 20 years ($M = 10.57$ years, $SD = 6.15$) and BCT between one and five years ($M = 2.14$ years, $SD = 1.35$). Four had already achieved BABCP accreditation in BCT, while the other ten were still in the process of becoming accredited. Eleven supervisors had treated over 200 CBT cases and three had treated between 50 and 200 CBT cases. All 14 supervisors had treated less than 50 BCT cases. Seven supervisors were clinical psychologists, three were CBT therapists, one was a nurse, one was an occupational therapist, one was a counselling therapist and one person did not indicate their professional background. All fourteen supervisors were familiar with the use of the CTS-R (Blackburn et al., 2001) and one was using it on a weekly basis, four were using it monthly and eight were using it less than once a month.

Scale development: BCTS-D

Table 2:

Participants' demographics - study 1

	Experts
	N = 14
Age – Mean (SD)	42.43 (7.71)
[range]	[31- 54]
CBT accredited	14
Years practicing CBT - Mean (SD) [range]	10.57 (6.15)
	[2 – 20]
Number of CBT cases:	
Treated < 50 cases	0
Treated 50 – 200 cases	3
Treated > 200 cases	11
BCT accredited	4
Years practicing BCT – Mean (SD) [range]	2.14 (1.35)
	[1 – 5]
Number of BCT cases:	
Treated < 50 cases	14
Treated 50 – 200 cases	0
Treated > 200 cases	0
Familiar with CTS-R:	14
Profession:	
Clinical Psychologist	7
CBT therapist	3
Nurse	1
Occupational therapist	1
Counselling therapist	1

Scale development: BCTS-D

BCTS-D v2

Information on Measure Development. Using information from the evaluation of BCTD v1 the expert group refined the scale and created an accompanying manual (see Appendix K; for more details on how this was done see results section).

Expert review.

Measures.

Feedback questionnaire. The aim of the feedback questionnaire was to assess validity and the utility of the scale. The questionnaire (see Appendix N) was based on a feedback questionnaire used in a study by Muse and colleagues (2016) and is in line with recent guidelines (Anastasi & Urbina, 1997; DeVellis, 2012). Participants were asked whether the scale was useful to help judge competency (1-not, 2- somewhat, 3- quite, 4- very useful), and to rate the overall style, appearance and layout of the scale (1- *poor*, 2- *fair*, 3- *good*, 4- *very good*). Further, they were asked how easy it was to use the scale (1- *not easy*, 2- *somewhat easy*, 3- *quite easy*, 4- *very easy*), if the scale gives the opportunity for useful feedback (1-not, 2- somewhat, 3- quite, 4- very useful) and how appropriate they found the scoring system (1-not, 2- somewhat, 3- quite, 4- very appropriate). If participants circled a three or below they were asked how the scale could be improved. Additionally, each item of the scale was rated for its relevance and clarity (1- not, 2- somewhat, 3- quite, 4- very relevant/clear).

Focus group. The aim of the focus group was to collect in depth feedback on the scale's strengths and weaknesses and provided a different perspective taking group consensus and dynamics into account (Krueger &

Scale development: BCTS-D

Casey, 2009). A semi-structured interview schedule was used to facilitate the discussion (see Appendix O). Within the schedule, emphasis was placed on facilitating discussion of both positive and negative responses to the BCTS-D v2. Where negative issues were mentioned, participants were asked whether the issue could be resolved and, if so, how. The interview was designed to be nondirective (Kvale, 1996). As such, questions were open-ended, thus allowing individuals in the group to comment on areas they thought were most important. The schedule was used flexibly and participants were encouraged to explore issues as they arose within the discussion. The discussion was facilitated by Michael Worrell (external research supervisor and BCT training lead) and lasted 45 minutes. The discussion was audio-recorded and transcribed according to Braun and Clarke's recommendations (2006), and used for the subsequent qualitative analysis (see Appendix P).

Analysis and data management.

Validity and usability. To confirm content validity an index of content validity was examined (Lynn, 1986) by calculating the percentage of participants who rated the item as both relevant and clear (i.e. a rating of three or four on the four-point scale). This was done separately for experts and novices as the calculation of the content validity index is usually based on expert feedback. Further, for the questions on how useful the scale was to help judge competency and whether any important aspects of BCT competence were missing the frequency of responses (yes/no) was looked at. For the other questions: the overall style, appearance and layout of the scale, ease of use of the scale, if the scale gives opportunity for useful feedback and

Scale development: BCTS-D

how appropriate the scoring system is, the mean and standard deviation was calculated.

Qualitative analysis. Qualitative feedback obtained in the questionnaire and through the focus group were analysed and represented together using thematic analysis (see p. 39 - 40 for the rationale for this analysis method). The feedback questionnaire gave an individual account of people's experiences and opinions of the scale and the focus group provided a more in depth account of the usability and the relevance of the scale. The analytical approach taken followed the guidelines for thematic analysis by Braun and Clarke (2006) and is described in the analysis of study one in more detail. Thereafter, a respondent validation was completed, evaluating all feedback given within the feedback questionnaire. Following this analytic process, revisions to the scale were made in response or an explanation and justification was provided as to why no changes were made.

Scale evaluation.

Measures. Participants filled in a demographic information sheet explaining their professional background and expertise (see Appendix I). The BCTS-D v2 and its accompanying manual was then used to rate an example therapy session.

Analysis and data management.

Internal consistency. Internal consistency was calculated, which is an assessment of inter-item reliability based on the correlation amongst individual items. Further, the Cronbach's alpha (α) if the item was deleted was calculated for each item to explore if removing an item would increase internal consistency.

Scale development: BCTS-D

Inter-rater reliability. Due to time and resource constraints it was only possible for participants to rate one therapy session using the BCTS-D v2. There are no clear guidelines regarding the most appropriate statistical approach for examining IRR when many raters rate only one session. The 'gold standard' usually is ICC when several judges rate the same targets and where assignment of raters is arbitrary (Shrout & Fleiss, 1979). This approach was taken to calculate the ICC for the overall scale. However, to calculate IRR for individual items, it was decided to employ three analytic strategies in order to 'triangulate' the results and to ensure the results are not skewed by a high degree of error². This was decided as calculating IRR for individual items includes less data and may therefore be more error-prone especially when using such a small sample. Firstly, as there was only one therapy session ICC values could only be established after randomly assigning participants as 'rater 1' and 'rater 2' and then calculating the ICC values, following the approach taken by Reichelt, James and Blackburn (2003). For each item a new SPSS database was created and five possible combinations of pairs of rater were entered into the database. Then the ICC was calculated for each item separately. Different allocations might yield different ICCs, which is why a range of stimulations was used (Dracup, 1997). Secondly, examining the range of scores and the standard deviation provides an indication of the degree of variation in the scores assigned for each item. Lastly, a P-bar analysis measures the reliability across all possible pairs of ratings for each

² This analytic approach was developed through consultation with statistical experts with experience of scale development and evaluation.

Scale development: BCTS-D

item and establishes the proportion of absolute agreement between the BCTS-D v2 ratings (Kalton & Stowell, 1979). P-bar is a more stringent method of calculating inter-rater agreement as it relates to absolute agreement and in this case, might be less error prone due to using all possible pairs of raters rather than a random sample of it.

Scale development: BCTS-D

Procedure. A second data collection day was organised where the BCTS-D version 2 (see Appendix F) was introduced and explained (about an hour presentation on the scale and how it's used). Thereafter, supervisors and trainees were asked to individually rate the competency of a therapist on the BCTS-D v2 by listening to an audio recording of a single therapy session. The chosen therapy session was a mid-treatment session (i.e. no assessment or ending session) with a couple of which one suffered with depression. The session was conducted by a trainee therapist allowing for a greater range of ratings to ensure a meaningful statistical analysis of the data (i.e. if an expert tape was rated there would be a high risk of a ceiling effect causing artificially high reliability). It was the same therapist conducting BCT with the same couple as in study one, however a different session was chosen. Participants were sent the manual and the BCTS-D v2 beforehand and asked to familiarize themselves with its content and were advised to use the corresponding manual for further support and guidance during the rating. After using the scale, all participants filled in a detailed feedback questionnaire exploring the BCTS-D v2's content and face validity, and its usability. Further, qualitative feedback about the scale was collected to further improve and refine the scale.

In the last part of the data collection day all participants attended a focus group to discuss content validity, the usability and the feasibility of the BCTS-D v2 and what changes or refinements would improve the scale. Focus groups offer the opportunity for collective discussion and hence provide a deeper understanding of how a group of people feel about an issue, providing a different perspective to the already individual written qualitative feedback

Scale development: BCTS-D

(Krueger & Casey, 2009; Vaughn, Schumm, & Sinagub, 1996). Therefore, the intention of the group was to identify what people thought worked well and what did not work so well using the scale to rate therapist competence and any areas which needed improvement or refinement.

Participants were invited to the training day as part of their continuous professional development. An information sheet about the study (see Appendix L) was included. Due to limited time on the day, participants were provided with the BCTS-D v2 and its manual beforehand and were asked to familiarise themselves with the material before the workshop. After having an opportunity to ask further questions at the beginning of the workshop participants signed an informed consent form (see Appendix M).

Participants. DeVellis (2012) suggests that, even though a general sample of 300 ratings has been described necessary in research for a scale evaluation study, practical experience has shown it possible with smaller samples. This is definitely true within the development of CBT competency rating scales: N = 102 for the CTS-R (Blackburn et al., 2001) and N = 111 for the ACCS (Muse, McManus, Rakovshik, & Thwaites, 2017). Further, the optimal number of participants for an expert review of a scale regarding its content validity is dependent on several different factors such as length and style of the scale and practical considerations such as availability of experts (Haynes, Richard, & Kubany, 1995). 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 experts (i.e. BCT supervisors), which is an essential ingredient of efforts to scale development (Brewer & Hunter, 2006), and novices (i.e.

Scale development: BCTS-D

BCT trainees) as it is useful to gain feedback from the target population (Campanelli, Martin, & Rothgeb, 1991). Three of the experts had also taken part in the first study. An email invitation was sent to 19 supervisors and 44 novices of which six supervisors (32%) and 14 trainees (32%) attended.

Of the six experts, four were BABCP accredited CBT therapists and two were not (simply due to practising in the United States of America) who had been practicing CBT for between seven and 20 years ($M = 14.5$ years, $SD = 6.16$) and BCT between two and nine years ($M = 5.67$ years, $SD = 2.34$). Four experts had treated less than 50 cases and two experts had treated between 50 -200 cases with BCT. Five experts had treated over 200 CBT cases and one had treated between 50 and 200 CBT cases. Their professions were Clinical Psychologist ($N = 4$), CBT therapist ($N = 1$) and nurse ($N = 1$). Four experts were familiar with the CTS-R, of which one was using it weekly, one was using it monthly and one less than once a month. Two experts were not familiar with the CTS-R due to practising in the United States of America.

Twelve novices had been BABCP accredited CBT therapists and two were in the process of becoming accredited. They had had been practising CBT between one and 10 years ($M = 7.0$ years, $SD = 2.51$) and BCT between one and three years ($M = 1.14$ years, $SD = .86$). Ten novices had treated more than 200 cases with CBT, two had treated between 50 – 200 cases, one less than 50 cases and one omitted to answer this question. All 14 novices had treated less than 50 cases with BCT at this point. Their professions were Clinical Psychologist ($N = 4$), CBT therapist ($N = 9$) and Counselling Psychologist ($N = 1$). Thirteen novices were familiar with the CTS-R, one

Scale development: BCTS-D

participant omitted their answer, and two were using it weekly, two were using it monthly and nine were using it less than once a month. There were no significant differences between experts and novices with regards to age³. However, the experts had been practicing CBT ($U = 11, p = .009$) and BCT ($U = 2, p = < .001$) for significantly longer than novices.

³ As the data was not normally distributed, non-parametric Mann-Whitney U -tests were performed to examine the difference between the age and years of experience held by novices and experts.

Scale development: BCTS-D

Table 3:

Participants' demographics – study 2

	Experts	Novices
	N = 6	N = 14
Age – Mean (SD)	44.67 (11.84)	38.64 (6.34)
[range]	[31 -61]	[28 - 50]
CBT accredited	4/6	12/14
Years practicing CBT - Mean (SD)	14.5 (6.16)	7.0 (2.51)
[range]	[7 – 20]	[1-10]
Number of CBT cases:		
Treated < 50 cases	0	1
Treated 50 – 200 cases	1	2
Treated > 200 cases	5	10
BCT accredited		
Years practicing BCT – Mean (SD)	5.67 (2.34)	1.14 (.86)
[range]	[2 - 9]	[1 -3]
Number of BCT cases:		
Treated < 50 cases	4	14
Treated 50 – 200 cases	2	
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

Results

BCTS-D v1

In the first study, supervisors ($N = 14$) used the BCTS-D v1 to rate a therapy session. Following that they were asked for their feedback on the usability and relevance of the BCTS-D v1 within the context of BCT.

Expert Review.

Validity and usability ratings. Seven participants out of 14 (50 % of the total sample, all experts) circled yes that there were important aspects of BCT competence they felt were missing from the scale (i.e. key competences which the scale neglected). The qualitative feedback on competences the participants felt were missing are outlined in Table 3. The refinements made in response to the qualitative feedback are also noted in the Table. Most experts felt the scale offered ample opportunity to give feedback (78.6%; yes = 11, no = 2, omitted = 1). Experts judged the scale's style and appearance to be good (Rating from 1 (poor) - 4 (very good), $M = 3.38$ and $SD = .51$), and rated the ease of the scale's use as between 'somewhat easy' and 'quite easy' (Rating from 1 (not easy) - 4 (very easy), $M = 2.46$ and $SD = .88$).

Qualitative feedback. A summary of the overall themes derived from the Thematic Analysis indicating how to improve and change the BCTS-D v1 is given in Table 3. The codes are organised in relation to themes found in the feedback: (1) Need to capture competence better, (2) Manual needed (i.e. a manual with guidance and anchor points regarding the competency levels is needed), and (3) Improve usability (revisions need to be made to the scale or manual to improve usability). The revisions to the BCTS-D v1 made in

Scale development: BCTS-D

response to each theme are outlined in the Table 3. A more in depth description of the development of the BCTS-D v2 is below (section 'scale development', p.68).

Scale development: BCTS-D

Table 4:

Overview of descriptive themes to improve and develop the BCTS-D v1

Theme	Codes	No. endorsed (out of N =14)	Solutions or Responses
Need to capture competence better	Assess with each partner whether there's an appropriate balance of interventions	1	In the BCTS-D v2 item 9 (selection of intervention strategy) is part of the scale. This item captures (1) if the therapist has selected clearly recognisable BCT intervention(s), (2) if the choice of intervention was appropriate for each individual and the couple considering environmental factors and (3) if the therapist introduced the selected strategy in a clear and appropriate way.
	Assess whether the interventions chosen are suitable for this couple	1	
	Assess application of BCT principles	2	
	Environmental factor missing in scale	1	In the BCTS-D v1 item 8 (formulation of depression in context) was added to the scale. This item assesses

Need to capture competence better

		<p>how the therapist is able to identify and focus on relevant individual, couple and environmental factors that appear implicated in the client's depression. This has been made clear in the now created manual.</p>
Assess appropriate flexibility in agenda setting	1	<p>The item 1 (agenda setting) gives flexibility in agenda setting by not being too descriptive. This reflects the BCT approach allowing the therapist a certain flexibility to adapt to the couple's style (i.e. needing very clear and detailed agenda versus having a couple of vague topics that will need discussing).</p>
Assess delivery of Psychoeducation	1	<p>Whether the therapist uses Psychoeducation appropriately is measured in several items: in item 8 (formulation of depression in context), where the therapist is supposed to use Psychoeducation to</p>

explain how depression can affect the client and the client's partner. Then it is measured in the three intervention items too, when the intervention itself is set-up and explained: item 10 (emotion-focused interventions), item 11 (cognitive interventions) and item 12 (behavioural interventions).

The expert group decided against creating a single Psychoeducation item as this might discriminate against later sessions, where appropriately not so much Psychoeducation is used.

Scale development: BCTS-D

Need to capture competence better

<p>Assess appropriate interruption of the couple – maybe part of therapist interpersonal effectiveness</p>	<p>1</p>	<p>In the BCTS-D v2 item 4 (facilitating couple communication) is added to the scale. This item captures whether the therapist was effective at facilitating couple communication. This involves making moment-to-moment choices about the communication needs of the couple such as being highly directive and interrupting one or both partners to enable an effective communication strategy.</p>
<p>Emphasise ‘containment of emotions’ in sessions</p>	<p>1</p>	<p>In the BCTS-D v2 item 4 (facilitating couple communication) is part of the scale. This item captures if the therapist was effective at facilitating couple communication. This involves making moment-to-moment choices about the communication needs of the couple including managing problematic emotional</p>

Need to capture competence better

			states and reactions in the session.
	Add a rating of couple complexity	1	<p>A formal rating of couple complexity was not added.</p> <p>The expert group discussed this issue highlighting that complexity can mean different things either that one partner has a complex diagnosis (i.e. physical health problems, which impact the session) or the therapeutic process with the couple is complex. To mitigate any potential effects of “couple complexity” however an informal summary sheet (see Appendix S) giving qualitative information (i.e. treatment stage, formulation of the couple) to the rater was added and a rule to rate items that are absent for appropriate reasons (i.e. the session was slow due to couple conflict and only included one intervention) is given a pass mark (i.e.</p>

given the score 3). The CTS-R does not take any
quantitative account of patient complexity.

Need to capture competence better

<p>Have a space for the person being rated to provide any important contextual information</p>	<p>1</p>	<p>Within the BCT postgraduate diploma course an informal summary sheet (see Appendix S) giving qualitative information about treatment stage and the formulation for the couple was requested before assessing the therapists. However, it was not made a formal part of the scale. Rakovshik and McManus (2010) discuss the importance of considering patient related factors when assessing therapist competence as these would impact therapy outcome and if not considered would ‘punish’ therapist treating more complex clients. However, the most widely used CBT competence assessment scales (i.e. CTS or CTS-R) do not take these into account.</p>
--	----------	--

Scale development: BCTS-D

Manual needed	A manual describing specific BCT competencies and how these can be applied well, is needed	4	A manual was developed accompanying the BCTS-D v2 rating scale describing how to use the scale and what each item consists of. Further, within each item all the different competency levels were described and anchor points of what to look out for during a session are given.
	Clear anchor descriptions for scoring needed	11	
	Scale felt subjective and dependent on one's personal experience	1	By developing a manual with a clear description of the items and guidance on how to rate each item, the expert group tried to make the ratings more objective.
Improvements to usability	Have expandable boxes for qualitative feedback	1	The space for qualitative feedback in the rating scale was made bigger.
	Make headings bold and put bullet points underneath	1	Item titles are now in italics and there are some anchor points underneath – bullet points were not added to avoid people using the scale more like a checklist

Scale development: BCTS-D

			rather than a rating scale.
Familiarisation with the scale	Need more practice to be familiar with the scale	1	The expert group is very clear that training is needed to be able to use the BCTS-D competently. It was thought that a one-day training workshop will be needed. However, this will be further thought about after completion of the second study (i.e. initial training workshop for the BCTS-D v2).
	Rating competence of a live treatment session is very complex	1	Doing observation-based competence ratings are very complex. However, in practice raters will have access to the therapy recording and will be able to go back and review certain sections if necessary. Further, it will need training and experience to be able to use the BCTS-D well.
	Use the tool for self-reflection first	1	The tool was developed with the intention of being used

both for assessor- and self-ratings. However, the expert group again highlight that for trainees using the scale and rating oneself will be a learning process supported through supervision and training.

Scale Evaluation.

Internal consistency. To assess inter-item reliability the correlation between all individual items in the scale was analysed. Cronbach's alpha for the BCTS-D v1 was $\alpha = .93$, which indicates that there was more than satisfactory agreement between the scale items for the BCTS-D v1. Bland and Altman (1997) state that an α with a minimum of .80-.85 is ideal. Table 5 outlines the Cronbach's alpha (α) if the item was deleted for each item in the BCTS-D v1. The α if item deleted ranged from .91 to .94. These results indicate that none of the BCTS-D v1 items would significantly increase the scale α if they were deleted and thus none were removed from the scale in this instance to increase internal consistency.

Scale development: BCTS-D

Table 5:

Cronbach's alpha if item was deleted for the BCTS-D v1

	α if item deleted	N ^a
Item 1 (Agenda setting)	.93	13
Item 2 (Feedback)	.92	13
Item 3 (Collaboration)	.92	13
Item 4 (Pacing and efficient use of time)	.92	13
Item 5 (Interpersonal effectiveness)	.93	13
Item 6 (Focus on depression in context)	.92	13
Item 7 (Eliciting appropriate emotional expression)	.92	13
Item 8 (Eliciting key cognitions)	.91	13
Item 9 (Eliciting and planning behaviours)	.92	13
Item 10 (Guided discovery)	.94	13
Item 11 (Conceptual integration)	.92	13
Item 12 (Application of change methods)	.92	13
Item 13 (Homework setting)	.92	13

^a One case was excluded due to missing value

Scale development: BCTS-D

Inter-rater reliability. Intra-class correlations (ICC) was used to examine inter-rater reliability between all raters ($N=14$) on the training day (Shrout & Fleiss, 1979). The following scale guideline was used to interpret the agreement coefficient: $< 0.20 = poor$, $0.21 - 0.40 = fair$, $0.41 - 0.60 = moderate$, $0.61 - 0.80 = good$, and $0.81 - 1.0 = very good$ (Gwet, 2010). The agreement between raters for the BCTS-D v1 total score was very good at $ICC = .84$.

BCTS-D v2

In this study, supervisors and trainees were asked to individually rate the competency of a therapist on the BCTS-D v2 by listening to an audio recording of a therapy session. After using the scale, all participants ($N = 20$) filled in a detailed feedback questionnaire exploring the BCTS-D v2's content and face validity, and its usability. Further, written and oral qualitative feedback about the scale was collected to further improve and refine the scale.

Scale development. After completion of the initial scale evaluation in study one, a second version of the BCTS-D was developed (BCTS-D version2; Appendix F). Further, the BCT expert reference group, with Sarah Corrie taking the lead, wrote up a comprehensive manual (see Appendix K) entailing detailed descriptions of each competence level for every item to support assessors use the BCTS-D v2 reliably (see Appendix F). The development of the manual and scale was informed by BABCP Minimum Training Standards (Holland, 2006), the Couple Therapy for Depression Competence Framework (Clulow, 2010a) and work on enhanced cognitive-behavioural therapy for couples (Epstein & Baucom, 2002). Some changes

Scale development: BCTS-D

were based on the results gained from study one, while others developed from further discussion within the expert reference group. The scale was divided into three broad domains (see Table 6): (1) Structure of the session (i.e. is there a clear sequence in the session?) (2) Interaction with the couple and management of the therapeutic process (i.e. how is the therapeutic process managed within the session? Is the therapist interrupting, guiding, etc. the couple within the session appropriately?) (3) Interventions selected and employed (i.e. have appropriate BCT interventions been chosen? How well are therapists conducting them?). These domains were chosen by the expert reference group as they are present in any session. These domains have not been confirmed with a factor analysis at this point. In the rating scale the items were placed in order of occurrence (i.e. agenda setting at the beginning, ending the session at the end). One new item was added to the BCTS-D v2 (item 4 facilitating couple communication) and the homework item was split into two different items: item 2 (review of homework from previous session) and item 14 (setting homework for next session). This meant that the scale consists of 15 items overall. Moreover, a scoring rule was introduced allowing assessors to give an item a pass mark (i.e. scored a three) when absent for appropriate reasons (e.g. one of the intervention items). This is very different to the CTS-R and was meant to make the BCTS-D v2 more reliable when rating treatment sessions with complex clients. Further, an informal summary sheet (see Appendix S) was added for trainees to fill in before having their therapy session rated by their supervisors. This sheet however, was not included as a formal aspect of the scale. Further, all items were changed to better fit a couple context and become BCT for depression

Scale development: BCTS-D

specific. These changes were agreed within the expert reference group and were based on BCT and relationship functioning research (Baucom & Boeding, 2013; Baucom et al., 2015, 1998; Baucom et al., 2012; Holtzworth-Munroe et al., 1989; Whisman, 1993, 1993, 2001, 2007; Whisman & Baucom, 2012) within which one member of the BCT expert reference group for the BCTS-D (Don Baucom) was a driving force (Epstein & Baucom, 2002).

Scale development: BCTS-D

Table 6:

BCTS-D v2 compared to 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) ^a	-
Item 15	Ending the Session	Overlap with feedback item
<hr/> BCTS-D: Domain 2. Interaction with the couple and management of the therapeutic process <hr/>		
Item 3	Collaboration	Collaboration
Item 4	Facilitating Couple Communication ^b	-
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

^a homework item split into two items

^b added to BCTS-D v2

Scale development: BCTS-D

Item 13 Dyadic Conceptualisation Conceptual integration

BCTS-D: Domain 3. Interventions

selected and employed

Item 8	Formulation of Depression in Context ^c	-
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

^c added to BCTS-D v1 and not present in CTS-R

Expert Review.

Content validity. Content validity results for each item in BCTS-D v2 are presented in Table 7. Data was looked at separately for experts and novices for content validity as the guidelines suggest that at least five experts for a meaningful analysis of the content validity index (CVI) is needed (see below). Mean relevance scores for individual items ranged from 3.67 to 4.00 for experts and from 3.57 to 3.93 for novices (1 – not relevant to 4 – very relevant), indicating that both novices and experts found all items in the scale at least ‘quite’ relevant. Mean clarity scores for individual items ranged from 3.17 to 4.00 for experts and from 3.33 to 3.86 for novices (1 – not clear to 4 – very clear), indicating that both novices and experts found all the domains in the scale at least ‘quite’ clear.

The CVI (i.e. the percentage of participants who rated the item as three or four on the four-point scale for both relevance and clarity, [Lynn, 1986]) ranged from 83.33 % to 100 % for experts and from 92.86 % to 100.00 % for novices, with the CVI being above the suggested threshold of 70 % required to establish content validity for all items in the scale.

Mann-Whitney tests⁸ revealed no significant differences between the scores for relevance or clarity assigned by novices and experts, indicating that novices and experts did not differ significantly in their views. Thus, all items in the scale were viewed as having acceptable content validity by both novices

⁸ As the data was not normally distributed, non-parametric Mann-Whitney *U*-tests were performed to examine the difference between the scores assigned by novices and experts.

Scale development: BCTS-D

and experts. In summary, none of the validity analysis indicated that any of the items should be removed from the scale. Thus, no items were removed on this basis.

Only one expert out of six (16.7% of the expert sample and 5 % of the total sample of 20 participants) and none of the novices indicated that there were important aspects of BCT competence they felt were missing from the scale (i.e. key competences which the scale neglected). The refinements made in response to the qualitative feedback indicating which aspects of competence participants felt were missing is outlined in the qualitative analysis section below.

Scale development: BCTS-D

Table 7:

Content validity results for each item in the BCTS-D v2

	CVI ^a		Relevance (1-4)		Mann – Whitney U ^b	Clarity (1-4)		Mann – Whitney U ^b
	Novices (N=14)	Experts (N=6)	Novices (N=14)	Experts (N=6)		Novices (N=14)	Experts (N=6)	
	%	%	M (SD)	M (SD)		M (SD)	M (SD)	
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)	100	91.67	3.93 (.27)	4.0 (.0)	p = .841	3.64 (.5)	3.33 (.82)	p = .494

^a CVI: 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.

^b As the data was not normally distributed, non-parametric Mann-Whitney *U*-tests were performed to examine the difference between the scores assigned by novices and experts.

Scale development: BCTS-D

communication)								
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	100	100	3.93 (.27)	4.0 (.0)	p = .841	3.78 (.43)	3.83 (.41)	p = .904
interpersonal								
effectiveness)								
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	96.43	100	3.77 (.44)	4.0 (.0)	p = .467	3.46 (.66)	3.67 (.52)	p = .639
depression in context)								
9 (selection of	96.43	100	3.92 (.28)	4.0 (.0)	p = .831	3.62 (.65)	3.83 (.41)	p = .639
intervention strategy)								
10 (emotion-focused	100	100	3.92 (.29)	4.0 (.0)	p = .820	3.33 (.49)	3.83 (.41)	p = .102
interventions)								
11 (cognitive	96.43	100	3.79 (.43)	4.0 (.0)	p = .494	3.36 (.63)	3.83 (.41)	p = .153
interventions)								
12 (behavioural	100	100	3.93 (.27)	4.0 (.0)	p = .841	3.5 (.52)	3.83 (.41)	p = .274

Scale development: BCTS-D

interventions)								
13 (dyadic	96.15	91.67	3.77 (.44)	4.0 (.0)	p = .467	3.54 (.88)	3.5 (.84)	p = .898
conceptualisation)								
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

Usability. Usability results for the BCTS-D v2 are presented in Table 8. These results indicate that both experts and novices felt the scale was at least 'quite useful' to judge BCT competence (1 – not useful to 4 – very useful), the way the scale helps give feedback was at least 'quite' useful (1 – not useful to 4 – very useful), the scale would be at least 'quite' easy to use (1 – not easy to 4 – very easy), had at least 'good' style, appearance and layout (1 – poor to 4 – very good), and had at least a 'quite' appropriate scoring system (1 – not appropriate to 4 – very appropriate).

Mann-Whitney U tests revealed no significant differences between the scores for style, appearance and layout or appropriateness of the scoring system assigned by novices and experts, indicating that novices and experts did not differ significantly in their views¹¹.

On the basis of those ratings no changes were made to the scale. However, the qualitative feedback provided by participants was used to revise the scale and implement changes to overcome limitations in relation to the ability of the BCTS-D v2 to judge BCT competence, its ability to give useful feedback using its scoring system and increasing its ease of use.

¹¹ As the data was not normally distributed, non-parametric Mann-Whitney *U*-tests were performed to examine the difference between the scores assigned by novices and experts.

Scale development: BCTS-D

Table 8:

Usability results for the BCTS-D v2

	Novices N = 14	Experts N = 6	Mann- Whitney U ^a
Useful to judge competency (1-4) Mean (SD)	3.5 (.52)	3.83 (.41)	p = .274
Opportunity for useful feedback (1-4) Mean (SD)	3.29 (.47)	4.0 (.0)	p = .051
Style, appearance and layout (1-4) Mean (SD)	3.5 (.52)	3.83 (.41)	p = .274
Appropriate scoring system (1-4) Mean (SD)	3.43 (.51)	3.8 (.45)	p = .257
Ease of Use (1-4) Mean (SD)	3.36 (.5)	3.5 (.84)	p = .494

^a As the data was not normally distributed, non-parametric Mann-Whitney *U*-tests were performed to examine the difference between the scores assigned by novices and experts.

Qualitative feedback. Thematic analysis of rater feedback identified major themes in the written and oral feedback on how to improve the scale and are presented in Tables 9 and 10 below. These Tables outline the themes relating to the whole scale, followed by the themes specific to one of the three domains. Out of the coding and the thematic analysis of the written feedback and the focus group data four main themes emerged: (1) Need to capture competence better, (2) complexity of competence assessment ratings, (3) improve clarity on how to use the scale (i.e. revisions needed to make the rating of the scale easier), (4) overlap of items (i.e. aspects of competence assessed in one item overlaps with the aspects of competence assessed in another item). Finally, many participants commented positively on the development of the scale. Participants' felt that the scale would be useful for self-reflection, personal development and can be used as an objective review of one's competencies, while having a reminder of what competencies are part of BCT. Further, it was stated that the scale is very thorough and gives the opportunity to rate the therapist's competency while accommodating the couple's needs (i.e. being very directive if needed).

The revisions to the BCTS-D v2 made in response to each coding, or an explanation as to why the issue in the theme was not resolved, are also outlined in the Tables 9 and 10.

Scale development: BCTS-D

Table 9:

Qualitative feedback on the BCTS-D v2: themes identified to improve the overall scale

Themes	Codes	No endorsed			Response or solution
		Super- visors	Novices	Focus group	
Need to capture BCT competence better	Assess the essence of the session and overall mark for pass/fail; this could be rated by both supervisors and trainees	2	1	1	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). Further, the scale is designed to define the competence during that session only as competence might fluctuate over time and situation (Dobson & Singer, 2005).

Scale development: BCTS-D

Need to capture BCT competence better	Assess how the therapist is teaching the couple skills	1	Teaching the couple skills is part of Item 9 (selection of intervention strategy), possibly item 10 (emotion-focused intervention), 11 (cognitive intervention), and 12 (behavioural intervention) depending on what the chosen intervention includes (i.e. communication skills training, problem-solving skills).
	Assess how well the therapist engages the couple in the process	1	This is already covered in the BCTS-D v2 with either Item 3 (collaboration) which assesses if the therapist ensured that each party present had a chance to contribute to the session and that the therapist created a working environment that fostered collaboration. Additionally, item 6 (therapist's interpersonal effectiveness) which

Scale development: BCTS-D

				assesses if the therapist engaged the couple and created an emotional climate that increases the likelihood of an effective session.
	Add a contextual framework of the overall therapy to the scale	2	2	The raters do receive an informal summary sheet (Appendix S) about the treatment. However, this could be included in the official rating pack.
Complexity of competence assessment ratings	Restricted ratings of 2,3 and 4's of BCT trainees on the BCTS-D v2	1		This was expected as all the ratings are for trainee therapists in their first year of BCT training and indicates the scale is working as it should. Higher end ratings of 5/6 are for levels of excellence.
	Complex and long manual which carries the risk that people won't use it	1	6	6

				<p>However, the expert group are quite clear that the scale should always be used in combination with the manual to ensure a reliable competency assessment. Further, 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.</p>
<p>Complexity of competence</p>	<p>assessment ratings</p>	<p>Uncertainty how to rate some items as more flexibility and possibly too subjective</p>	<p>2 3</p>	<p>The BCTS-D v2 is purposely more flexible in its use compared to the CTS-R. This is to reflect the nature of the treatment as BCT is a principle driven approach. This flexibility ensures that therapists are not scored down for adjusting their approach for potentially more complex couples (i.e. slowing</p>

complexity of competence assessment ratings

		session down when a lot of arguing between couple). The manual was developed to make the rating of the BCTS-D v2 more objective. More research on IRR is needed.
Unclear how to rate when the performance is mixed	1	The manual of the BCTS-D v2 provides clear anchor points in the manual and examples of how to rate certain performances (even if mixed). More research needed to see what could be changed.
Uncertainty of how much training needed before able to use the BCTS-D v2 confidently.	2	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.

Scale development: BCTS-D

Improve clarity on how to rate items/ use the scale	Clarification on scoring needed - how to rate items that are not present for appropriate reasons (i.e. item 10, 11, 12 (emotion- focused, cognitive and behavioural interventions)	4	6	The BCTS-D v2 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.
	Clarify the purpose of Item 4 (facilitating couple communication) compared to item 12 (behavioural intervention)	1		The item 4 (facilitating couple communication) focuses on the process of communication throughout a session (i.e. how to interrupt someone, how give someone time to find their words) and it's not about teaching a skill. This is covered in item 12 (behavioural intervention). This has been highlighted in the manual for both items

Improve clarity on how to rate items/ use the scale

		now.
The scoring sheet is unclear	2	The descriptions of each item of the BCTS-D v2 scoresheet should only provide a small indication what the item is about. However, to actually rate the item the assessor needs to use the manual for clarification.
Clarify how to differentiate between a score 3 and 4 with more examples	2	There are some examples in the manual for the BCTS-D v2 already. The expert group discussed that they would not like to add more examples and make the manual longer and more chaotic. Further, they highlighted the risk of the scale becoming too prescriptive if there are too many examples in the scale.

Scale development: BCTS-D

<p>Improve clarity on how to rate items/ use the scale</p>	<p>Clarify difference between 1 guided discovery versus cognitive intervention items by providing more examples</p>	<p>In the BCTS-D v2 item 7 (guided discovery) underpins an entire session while a cognitive intervention (item 11) can be present in some parts of the session, but might not always be. Therefore, in some sessions there might be some overlap, but in others there won't be. The expert group did not want to add more examples to not become too descriptive or confusing at this stage.</p>
<p>Overlap of item</p>	<p>There is overlap of items, making rating difficult</p>	<p>1 This comment isn't very clear as to which items it applies and is therefore difficult to respond to.</p>

Table 10:

Themes identified to improve the three domains

Themes	Codes	No endorsed	Response or solution
		Super- visors	Novices Focus group
Domain 1: Structure of the session			
Improve clarity on how to rate items/ use the scale	Difficulty rating Item 2	1	Generally, in BCT it would be very unusual for there not to be any homework. The homework might be less structured and formally agreed compared to CBT, but usually there was something agreed in the previous session (except after assessment sessions). If it is obvious why homework was omitted, it would be scored as a pass (score 3), which is in the manual. More research needed to
	(review of homework) if no prior homework set for appropriate reasons		

				see if that needs clarifying.
	Clarify differences of the agenda setting (item 1) compared to individual CBT	1	1	The expert group highlight that item 1 (agenda setting) is not necessarily different from individual CBT. The way an agenda is set up might be more flexible as there is no need to set timings or have an exact list of topics.
		Super- visors	Novices	Focus group
Domain 2: Interaction with the couple and management of therapeutic process				
Need to capture BCT competence better	Clarify the description of item 3 (collaboration)	1		The manual describes how item 3 (collaboration) is all about balance between the therapist, the couple and each partner equally. The expert group are currently reviewing the instructions in the manual and considering adapting those.

Scale development: BCTS-D

Clarify how item 5 (pacing and flow) is different in couple therapy compared to individual therapy

1

The expert group discussed how the pace and flow of a BCT session isn't necessarily different to an individual CBT session. The main difference lies in how to manage the communication with the couple and the therapist, which is why item 4 (facilitating couple communication) was added to the scale.

Scale development: BCTS-D

Need to capture BCT competence better

<p>Incorporate attending to cultural differences into item 6 (therapist's interpersonal effectiveness) and make scoring more objective</p>	1	1	1	<p>Item 6 (therapist's interpersonal effectiveness) is part of assessing how the process of a therapy session went. The expert group highlighted that assessors often find this domain the most difficult to rate and that it might be the domain at risk of a bias such as personal liking. However, by using the accompanying manual of the BCTS-D v2 and attending training on how to use the scale any such bias is hoped to be minimised. More research will need to explore this issue further.</p>
<p>Clarify how Item 7 (guided discovery) is different from CBT Socratic questioning and define the construct</p>	1			<p>The expert group discussed how Socratic questioning is part of the guided discovery item, which defines a style of enquiry and exploration, and is not seen as a technique on its own. The</p>

Scale development: BCTS-D

	more clearly.			manual mentions how BCT therapists are less likely to use the more traditional, Socratic-style questioning typical of individual CBT, and it is <i>unlikely</i> that the therapist will seek to ‘expose’ inaccuracies in one partner.
Improve clarity on how to rate items/ use the scale	More specific examples for item 13 (dyadic conceptualisation) would be helpful	1	1	There are some examples in the manual for the BCTS-D v2 already. The expert group discussed that they would not like to add more examples and make the manual longer and more chaotic. Further, they highlighted the risk of the scale becoming too prescriptive if there are too many examples in the scale.
	More examples of item 3 (collaboration) on scoring		2	The expert group highlighted how having too many examples on the scoring sheet could lead to people

Scale development: BCTS-D

	sheet				not using the manual. To avoid this no further examples will be added to the scoring sheet.
Improve clarity on how to rate items/ use the scale	Item 4 (facilitating couple communication) needs more behavioural descriptions and clarity on how to rate it (especially when little communication) - highlight that sharing thoughts and feelings does not belong to it.	1	2	2	The expert group highlighted how no communication can still be a facilitation of communication. The same counts if the intervention of 'sharing thoughts and feelings' was used a lot. There would still be communication (i.e. how the intervention was introduced and talked about it / did both clients have an opportunity to talk?). However, need to highlight that 'sharing thoughts and feelings' is part of item 10 (emotion-focused intervention).
	Difficulties rating item 7 (guided discovery)	1			This comment was quite vague and the expert group felt to need further research in order to

Scale development: BCTS-D

			helpfully change the rating of item 7 (guided discovery).
Overlap of item	Consider cultural influences as its own item or part of interpersonal effectiveness, rather than being part of Item 3 (collaboration).	1	The expert group discussed how cultural influences would normally be considered in item 6 (interpersonal effectiveness), but certainly not as its own item. The expert group are considering rewording item 6 to make this more obvious in the manual.
	Overlap of Item 3 (collaboration) with other items. This makes it difficult to rate.	1	This comment is tricky to respond to as it is not clear with what item it is supposed to overlap. However, the item was reviewed by the expert group anyway (see above).

Overlap of item	Overlap of Item 4 (facilitating couple communication) with other items such as item 9 (selection of intervention strategy). This makes it difficult to rate.	1	The expert group discussed how item 4 (facilitating couple communication) might overlap with other items (i.e. item 9 (selection of intervention) or item 11 (cognitive intervention)). The manual explains however that item 4 is about how the therapist manages the process of communication throughout a session (i.e. introducing an intervention or homework), rather than what intervention was selected or if the intervention was conducted well. There might be an area of overlap. This is outlined in the manual and will need further research to see how it could be helpfully changed.
	<hr/>		

Scale development: BCTS-D

Overlap of item	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)	1		The expert group considered this issue and decided to do more research on IRR to see if item 7 is rated reliably between assessors or not before possibly omitting it from the scale.
	Collapse item 8 (formulation of depression) and item 13 (dyadic conceptualisation) into one item about psychoeducation	1	1	The manual describes how item 8 (formulation of depression) assesses the extent to which the therapist is able to identify and focus the session on the relevant individual, couple, and environmental factors that appear implicated in the patient's depression. The therapist needs to be clear as to whether their primary objective is to (1) target the

disorder or (2) improve the relationship during a given session. However, item 13 (dyadic conceptualisation) taps something different and addresses the therapist's competence in helping the couple view and address their problems with a dyadic perspective. Many couples enter therapy exhibiting specific, and well-documented patterns of negative interaction, and it is essential, therefore, that the therapist knows how to identify and conceptualise their impact as a precursor to intervening. Therefore, the expert group decided that collapsing them into one item would make rating the item a lot more difficult. Further, often later in treatment there is not much use of Psychoeducation which would make a

psychoeducation item difficult to rate.					
		Super- visors	Novices	Focus group	
Domain 3: Interventions selected and employed					
Need to capture BCT competence better	More examples on scoring sheet for item 8 (formulation of depression), item 10 (emotion-focused), item 11 (cognitive) and item 12 (behavioural interventions).	2	3	1	As already mention the expert group do not want the score sheet to replace the manual or encourage assessors to not use the manual. Therefore, no further description will be added to the score sheet. Additionally, the expert group wouldn't want to give a list of interventions within the manual to avoid the BCTS-D v2 becoming to prescriptive, which would clash with the BCT ethos of being principle driven.

Scale development: BCTS-D

Improve clarity on how to rate items/ use the scale	Clarify rating of item 10,11 and 12 (emotion-focused, cognitive and behavioural interventions) if an intervention includes elements of each	1		The manual of the BCTS-D v2 provides clear anchor points of how to rate each intervention (even with mixed performance). Further, the expert group highlighted that it is expected that the different interventions might cause a change within a different modality (i.e. cognitive intervention might cause an emotional shift within the couple). The assessor would however still rate the intervention rather than the outcome.
	Clarify how to rate item 8 (formulation of depression) if depression is not present.	2	1	The BCTS-D v2 was specifically designed to assess BCT competence for therapists treating patients suffering with depression. Therefore, even if the patient might not seem as depressed during a session, depression would still be part of the

Scale development: BCTS-D

formulation.

Scale development: BCTS-D

Scale evaluation.

Internal consistency. Cronbach's alpha (α) was calculated to assess internal consistency. Cronbach's alpha (α) ranges from 0 = *items independent* to 1 = *items identical*, with a minimum of .80 - .85 typically representing satisfactory agreement (Bland & Altman, 1997). Cronbach's alpha for the BCTS-D v2 was $\alpha = .94$ ($N = 20$, 2 missing), which is as high as in the BCTS-D v1. These results indicate that there was more than satisfactory agreement between scale items for the BCTS-D v2. Table 11 outlines the Cronbach's alpha (α) if an item was deleted for each item in the BCTS-D v2. The Cronbach's alpha if an item was deleted ranged from $\alpha = .93$ to .94 for the BCTS-D v2. These results indicate that none of the BCTS-D v2 items would significantly increase the scale's internal consistency, if they were deleted and thus none were removed from the scale.

Scale development: BCTS-D

Table 11:

Cronbach's alpha if item was deleted for the BCTS-D v2

	α if item deleted	N ^a
Item 1 (Agenda setting)	.94	18
Item 2 (Review of homework)	.93	18
Item 3 (Collaboration)	.93	18
Item 4 (Facilitating couple communication)	.93	18
Item 5 (Pacing and Flow)	.93	18
Item 6 (Therapist's interpersonal effectiveness)	.93	18
Item 7 (Guided discovery)	.93	18
Item 8 (Formulation of depression in context)	.94	18
Item 9 (Selection of intervention strategy)	.93	18
Item 10 (Emotion-focused interventions)	.93	18
Item 11 (Cognitive interventions)	.94	18
Item 12 (Behavioural interventions)	.93	18
Item 13 (Dyadic Conceptualisation)	.93	18
Item 14 (Setting homework)	.93	18
Item 15 (Ending the session)	.93	18

^a N=2 were excluded from these calculations due to missing values

Scale development: BCTS-D

Inter-rater reliability. To establish inter-rater reliability for the total score of the scale a two-way mixed effects approach ICC with absolute agreement was calculated. The following benchmarking guidelines were used to interpret agreement coefficients: $< 0.20 = \textit{poor}$, $0.21 - 0.40 = \textit{fair}$, $0.41 - 0.60 = \textit{moderate}$, $0.61 - 0.80 = \textit{good}$, and $0.81 - 1.0 = \textit{very good}$ (Gwet, 2010). Agreement between raters for the BCTS-D v2 for the overall scale was very good with $ICC = .83$. *Intra-class correlation for individual items.* To establish inter-rater agreement for individual items participants were randomly assigned as rater 1 and rater 2, this is in line with recent research on the CTS-R (Reichelt et al., 2003). Five possible combinations of pairs of raters were used to calculate the ICC. Different allocations might yield different ICC values, which is why a range of simulations was used (Dracup, 1997). The following benchmarking guidelines were used to interpret agreement coefficients: $< 0.20 = \textit{poor}$, $0.21 - 0.40 = \textit{fair}$, $0.41 - 0.60 = \textit{moderate}$, $0.61 - 0.80 = \textit{good}$, and $0.81 - 1.0 = \textit{very good}$ (Gwet, 2010). The single item ICC values ranged from $ICC = .40$ to $.00$ (see Table 12). Four out of the individual items fell in the range of fair agreement (item1, 3, 4 and 8) and 11 of the individual items fell in the range of poor agreement of which some had negative values. Negative reliability values can occur when the sample size is small and only a small number of items is looked at (Magnusson, 1966; Nichols, 1999). A low true reliability score and random disturbances can generate a negative rather than positive average covariance and hence a negative alpha and ICC. As negative scores make no statistical sense, it is recommended to equal the negative value to zero (Magnusson, 1966; Nichols, 1999).

Scale development: BCTS-D

Variance for individual items. The variance of the scores provides an indication of the degree of variation in the scores assigned for each item (see Table 12) by examining the range of scores and the standard deviation (i.e. how scores spread around the mean). The standard deviation was low for three items (below .54): item 4, item 8 and item 11, indicating a higher agreement between raters. All the other item scores had a high degree of variation (between .70 and 1.25), indicating low rater-agreement.

P-bar analysis for individual items. P-bar analysis measures the reliability across all possible pairs of ratings for each item and establishes the proportion of absolute agreement between the BCTS-D v2 ratings (Kalton & Stowell, 1979)^a. Within this study there were 18 participants who fully rated a therapy session on the BCTS-D v2, which results in 153 different pairwise comparisons of the ratings for each item. The p-bar values of five items (see Table 12) lie between .41 and .55, which shows good inter-rater agreement. Most items' p-bar value was between .31 and .41 indicating low rater agreement. Further, two items were below that: Item1 and item 6 demonstrated very low absolute agreement (.26 and .28 respectively). On the other hand, item 11 demonstrated very high absolute agreement at .76.

Summary of IRR. All three different approaches analysing IRR gave a similar picture. IRR on the BCTS-D v2 for individual items is generally very

^a An aid to the interpretation of the P coefficients is provided by noting that if 17 out of 18 coders agree on the coding response, then $P = 88.9\%$ for that response, if the coders split into 10 in agreement and 8 in disagreement then $P = 47.7\%$.

Scale development: BCTS-D

low with six items, depending on the analysis, demonstrating more promising IRR (item1, item 2, item 4, item 8, item 11 and item 12).

Scale development: BCTS-D

Table 12:

Analysis of inter-rater reliability for individual items within the BCTS-D v2

	M (SD)	Range	Intra-class correlation ^a		P-bar ^b
			ICC	95% CI	
Item 1 (Agenda setting)	1.56 (1.25)	0 – 3	.24	[-.377-.581]	.26
Item 2 (Review of homework)	3.00 (.77)	1 – 4	0	[-1.39-.30]	.41
Item 3 (Collaboration)	3.00 (.84)	1 – 4	.38	[-.06-.64]	.33
Item 4 (Facilitating couple communication)	3.17 (.51)	2 – 4	.40	[-.09-.67]	.55
Item 5 (Pacing and Flow)	2.83 (.79)	1 – 4	0	[-2.16-.004]	.35
Item 6 (Therapist's interpersonal effectiveness)	3.50 (.92)	2 – 5	0	[-2.91- -.11]	.28
Item 7 (Guided discovery)	2.83 (.79)	2 – 4	.14	[-.58-.53]	.31
Item 8 (Formulation of depression in context)	3.10 (.54)	2 – 4	.28	[-.30-.60]	.54
Item 9 (Selection of intervention strategy)	3.10 (.80)	2 – 5	.16	[-.52-.53]	.35

^a Five simulations of random pairings were done to calculate ICC values for individual items

^b The P-bar value shows the proportion of absolute agreement between pairs of raters

Scale development: BCTS-D

Item 10 (Emotion- focused interventions)	3.10 (.80)	2 – 5	0	[-.78-.43]	.35
Item 11 (Cognitive interventions)	3.00 (.37)	2 – 4	.04	[-.79-.49]	.76
Item 12 (Behavioural interventions)	3.11 (.83)	1 – 5	0	[-2.07 - .09]	.45
Item 13 (Dyadic Conceptualisation)	3.39 (.70)	2 – 5	.01	[-.70-.44]	.39
Item 14 (Setting homework)	2.78 (.81)	2 – 4	0	[-1.34-.31]	.32
Item 15 (Ending the session)	2.78 (.88)	1 – 4	0	[-1.35-.27]	.34

Discussion

Summary of Results

BCTS-D v1. The results of the first study indicate that the BCTS-D v1 demonstrates good usability, validity and reliability and is comparable to the CTS-R.

Validity and usability. The missing manual and missing BCT specific items were highlighted with half of all BCT supervisors (7 out of 14) stating that some aspects of BCT competence were missing. This information was collected as written feedback as part of the expert review to examine the usability and face validity of the BCTS-D v1. The majority of BCT supervisors felt the scale gave ample opportunity for feedback and the scale itself was user-friendly. The qualitative feedback about the BCTS-D v1 reflected a similar picture with BCT supervisors commenting to improve the content validity of the scale: 1) certain items were missing from the scale (i.e. appropriate balance of interventions with each partner, whether BCT principles were applied, interpersonal effectiveness within a couple setting), 2) a rating on couple complexity was needed, 3) a space for contextual information about the therapy session was needed, and 4) a manual with clear anchor descriptions and instructions on how to use the BCTS-D v1 was needed. In terms of usability BCT supervisors felt the scale was easy to use and was visually appealing (i.e. had good style, appearance and layout) with maybe needing more space for qualitative feedback. Further, BCT supervisor commented on needing more time to become familiar with the scale and its use.

Scale development: BCTS-D

Taken together, these results suggest that the BCTS-D v1 demonstrated comparable reliability to the CTS-R and good validity and was a good first step in the direction of developing an observer-based rating scale to assess competence within BCT.

Reliability. The Cronbach's alpha was very high with $\alpha = .93$, which is comparable to the CTS-R (α range = .75 - .97; Blackburn et al., 2001; James et al., 2001; Reichelt et al., 2003). This is in line with DeVellis' (2012) recommendation of a higher α ($>.90$) when developing a new scale to allow for α to deteriorate when used in a new research context.

Agreement between the supervisors' rating of the one therapy session on the BCTS-D v1 was very good (ICC = .84). This is comparable or higher than inter-rater reliability achieved with the CTS-R (average ICC = .63 [Blackburn et al., 2001], $r = .67$ with training and .44 without training [Reichelt et al., 2003], ICC = .38 [Gordon et al., 2006]). The inter-rater reliability for individual items was not examined in this study. These results are very encouraging and confirm that an adapted version of the CTS-R is feasible to be used within a couple context.

Previous research has shown that a large amount of assessor training is necessary to achieve adequate inter-rater reliability on the CTS-R (Gordon, 2006; Reichelt et al., 2003). The data on the BCTS-D v1 were collected after a one day workshop on 'BCT supervision' with an hour introduction on the BCTS-D v1. However, all supervisors in this sample had on average 10 years CBT experience with the majority having treated over 200 CBT cases. This meant they were all familiar with the CTS-R and its use. Further, no manual existed yet describing the competences for BCT in more detail, which might

Scale development: BCTS-D

mean that the BCTS-D v1 focused more on the general CBT skills (similarly to the CTS-R [Blackburn et al., 2001]) underlying BCT rather than the specific techniques used within BCT.

Scale revisions. In light of the results and the feedback received from the BCT supervisors, a number of areas of how to improve and further develop the BCTS-D v1 were identified. Other changes were based on further discussion within the BCT expert group. These included adding two items and restructuring existing items on the BCTS-D v1 further. A comprehensive manual (Appendix K) was created and the BCT expert group decided to divide the scale into three broad domains: (1) Structure of the session (2) Interaction with the couple and management of the therapeutic process (3) Interventions selected and employed. These were selected as they reflect the cornerstones of every BCT session. However, these domains will need confirming by doing a factor analysis of the scale. Finally, trainee therapists being rated on the scale were asked to hand in a written summary (see Appendix S) providing the assessor with a context in which to assess the tape overview (i.e. couple's presenting problem and gender, formulation, how many sessions so far and the aim of the session). This approach was used in other recent developments of CBT competency scales (Muse et al., 2017) to mitigate any potential effects of couple complexity. Further, if an item was absent for appropriate reasons (e.g. one of the three intervention items) the item would be given a pass mark. The revised version, BCTS-D v2 and its manual, can be seen in Appendix F and K.

When creating the new manual, the expert panel tried to counteract some of the limitations of the CTS-R. Specific, behavioural anchor

Scale development: BCTS-D

descriptions of how to competently deliver BCT were included to develop a mutual understanding between the raters using the scale and improve inter-rater reliability.

BCTS-D v2. The results of the second study indicate that the BCTS-D v2 demonstrates good validity and usability, and good internal-consistency, but low inter-rater reliability especially for individual items.

Validity and usability. The BCTS-D v2 received encouraging feedback from raters when examining the usability, face validity and content validity of the BCTS-D v2. The majority of experts and novices 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 expert out of the 20 participants indicated that important aspects BCT competence were missing in the scale. Further, experts and novices indicated that they found the scale useful and easy to use with an appropriate scoring system. Thus, the BCTS-D v2 appears to have good face validity, content validity and usability.

Qualitative feedback. During the second study, qualitative written and oral feedback was obtained from BCT experts and novices who had experience using the BCTS-D v2 in order to examine assessors' views about the usability and utility of the BCTS-D v2 in more depth.

Strengths. A number of strengths of the BCTS-D v2 were outlined by participants in the written feedback as well as the focus group. Participants' felt that the scale would be useful for self-reflection, personal development and can be used as an objective review of one's competencies. Further, it was stated that the scale is very thorough and gives the opportunity to rate

Scale development: BCTS-D

the therapist's competency while accommodating the couple's needs (i.e. being very directive if needed). Whilst some participants found the comprehensiveness of the BCTS-D v2 and its manual a key strength, others felt that the amount of detail in the manual was challenging. Consequently, participants reported that the opportunity to practice and become familiar with the scale before using it was important. Thus, the importance of training on how to use the scale was highlighted.

Flexibility of scale. Another issue that was raised by the focus group in particular, was the fact that the use of the BCTS-D v2 appeared more flexible compared to the CTS-R (with which most participants were familiar). One flexible criteria which caused a lot of confusion was that the scale allows for certain items that are appropriately omitted to be given a pass mark (scored a three). However, some participants missed that in the manual and stated that even though they remembered certain items could be omitted from the short presentation at the beginning of the training day, they were confused as how to rate those. To solve this issue, the expert group has decided that this important rule will be added in bold to the scoring sheet. A question remains however, if items, which do not occur during every session, need to be part of the scale or could be dropped. And if that rule is making the ratings too complicated and unreliable. The CTS-R does not apply such a rule. To get more clarity on this further research is needed exploring inter-rater reliability for two separate groups, one applying that rule to rating the scale and one that does not.

Room for interpretation. Additionally, some participants commented that they felt there was generally more room for interpretation when rating

Scale development: BCTS-D

most items compared to the CTS-R. At this point it is not clear if this is due to the manual not being clear enough or due to participants not having read the manual properly and not having gone through it in detail during the short presentation. Another explanation could be that BCT is a principle driven approach rather than a protocol driven one, and many participants had only recently started their BCT training after practising CBT within IAPT for several years, which often is very treatment protocol led. This was highlighted with some of the qualitative feedback when novices asked for more examples on how particular items were different within a couple's context (i.e. agenda setting, pacing and flow). These items however are not fundamentally different within a couple context, and could explain why some participants struggled to rate certain items.

Assessor expertise. This brings up another question of how much expertise an assessor needs to be able to use the BCTS-D v2 effectively. Certain researchers argue that experts are needed to decide if an intervention was delivered competently (Barber & Critis-Christoph, 1996; Waltz, Addis, Koerner, & Jacobson, 1993), which means that an assessor using the competence measure needs to be sufficiently skilled in the therapy being rated in order to judge competence. This is supported by the notion that trained expert raters achieved better inter-rater reliability on the CTS-R, (ICC = .63; Blackburn et al., 2001) than experienced CBT practitioners with a certain knowledge of the CTS-R and within a naturalistic setting (ICC = .38; Gordon, 2006).

Use of manual. Some confusion on how to use the BCTS-D v2 was highlighted in the feedback received, this could either be due to the manual

Scale development: BCTS-D

and scale not being very clear and too complicated to use or due to participants not having read the manual before the workshop and then only having had time to briefly read the BCTS-D v2 manual before completing the rating on the BCTS-D v2 for the therapy session provided for the first time. Even if the confusion is due to not having read the manual properly, the findings from this study may be representative in that busy assessors within research, training and routine practice settings may find it difficult to find the time or motivation to read the BCTS-D v2 manual, which incorporates 98 pages. One solution to this problem may be to shorten or remove the specific guidance provided in the manual. However, this is likely to lead to increased subjectivity of ratings and thus reduced inter-rater reliability. Furthermore, participants repeatedly asked for more item-specific examples or rating guidelines. Another solution highlighted by participants in this study was to add further guidance about the anchor points on the scoring sheet. However, adding all of the detail included in the manual would defeat the purpose of a manual. An alternative solution is to offer assessors training in how to use the BCTS-D v2. Attending training may be less of a barrier, even though time-consuming as well, than finding time in a busy schedule to read a manual. Thus, a useful avenue for further exploration may be the development of a standardized training in how to use the BCTS-D v2 and evaluating its effect on reliability and perceived validity and usability of the scale.

Therapeutic context. Another area of feedback from raters was to add a contextual framework for the treatment and a couple complexity rating to the BCTS-D v2. Not taking therapeutic context into account when measuring competence has been criticised in the research (Stiles, Honos-Webb, &

Scale development: BCTS-D

Surko, 1998; Waltz et al., 1993), stating that it undermines the scales' validity. It could lead to therapists being penalized and rated as less competent for treating more complex cases (Rakovshik & McManus, 2010). Waltz and colleagues (1993) argue that competence includes "the extent to which the therapists conducting the interventions took the relevant aspects of the therapeutic context into account and responded to these contextual variables appropriately" (p.620). The therapeutic contexts that should be considered when rating competence are 1) stage in therapy, 2) client difficulty, and 3) client presenting problems (Waltz et al., 1993). Stage in therapy includes both number of sessions completed and the progress already made to ensure the assessor can adequately judge if the therapist has used the appropriate techniques. For example, in CBT exploring underlying core beliefs in the first session might be considered inappropriate and incompetent, but if the same technique was used during later sessions the therapist would be judged as highly competent. Waltz (1993) therefore argues that using the same scale with the same items for each session (i.e. no matter at what stage in treatment), like for example the CTS-R does, does not consider stage of therapy. The developers of the BCTS-D v2 request an informal summary sheet (i.e. couple's age, gender and presenting problem; diagnosis; formulation; number of sessions; what was done previously and the aim of the session; see Appendix S) about the therapeutic context when rating a therapy session within the BCT training course, but agreed that it should be part of the official BCTS-D v2 pack. The manual however does not specify how this summary sheet should inform the competence rating and would need clarifying in the future to ensure reliable use of it. Additionally, the BCTS-D v2

Scale development: BCTS-D

manual specifies that “the scale is optimally suited to assessing performance on mid-treatment sessions where the therapist is *demonstrating the use of specific BCT interventions*” (p.6; Corrie, Fischer, Worrell, & Baucom, n.d.).

Further, when doing a summative rating on the BCTS-D v2, it is recommended in the manual to exclude ending sessions, review sessions or session that mostly focus on psychoeducation. Another aspect of the BCTS-D v2 that supports the notion of needing to take therapeutic context into consideration is the flexibility that certain items can be rated with. For example, if a couple is very hostile to each other and the therapist has to take a lot of time to control that hostility in the room and focus the couple, the therapist would not be penalised for not using any intervention techniques during that session. If the interventions were omitted for appropriate reasons they can be given a pass mark (i.e. scored a three).

The second aspect to affect ratings is client difficulty, which is described as a multidimensional construct including the degree of functional impairment, presence of collateral problems, number of previous episodes (i.e. depressive episodes) or duration of the problem (Waltz et al., 1993). A study found that supervisors’ ratings of therapist performance decreased with greater client difficulty. This was also true for the therapists’ self-ratings of effectiveness (Foley, O’Malley, Rounsaville, Prusoff, & Weissman, 1987). The BCTS-D v2 covers some of these aspects in the summary sheet (see Appendix S), for example in the formulation section therapist can include any hypotheses about patients’ expectations and struggles, but how this should influence the competence ratings on the BCTS-D v2 needs to be specified within the manual. Other scale’s such as the Manual-Assisted Cognitive

Scale development: BCTS-D

Behavior Therapy Rating Scale (MACT-RS) have included an item assessing client difficulty (Davidson et al., 2004). However, the item was described to confound patient difficulty and therapist competence, with scores ranging from 1 - therapist is ineffective not due to client to 7 – client very difficult but doesn't appear to be due to therapist's effectiveness. This description makes it unclear where, for example, a competent therapist whose patient was not difficult would fit on this item (Muse & McManus, 2013).

The third aspect of therapeutic context (Waltz et al., 1993) is presenting problem. The BCTS-D v2 is specifically measuring competence on how therapists deliver BCT for depression, which targets two elements: treating depression and increasing relationship satisfaction. However, it was mentioned in the rater feedback that the depression specific item on the scale (formulation on depression in context) might not be feasible within routine practice and that in reality depression is not always present and that their client group has a huge comorbidity with anxiety disorders. Within CBT disorder-specific scales were demanded to assess the competence in delivering disorder-specific strategies and procedures thought to be central to that CBT protocol (Fairburn & Cooper, 2011). On the other hand, it's been discussed that the use of such disorder-specific scales makes more sense within research settings, where patients usually have clear and well-defined diagnoses. Within routine clinical practice or even training settings however patients often experience a range of mental health problems and comorbidity making a transdiagnostic scale more relevant (Muse & McManus, 2013). Considering that the BCTS-D v2 was developed to be used within a training setting this issue needs reconsidering in the future.

Scale development: BCTS-D

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.

Internal consistency. Measures of internal consistency indicate that items in the BCTS-D v2 are highly inter-correlated and thus that the BCTS-D v2 measures a single underlying construct ($\alpha = .94$). This internal consistency rating is comparable to the CTS-R (α range = .75 - .97; Blackburn et al., 2001; James, Blackburn, Milne, & Reichelt, 2001; Reichelt, James, & Blackburn, 2003) and in line with DeVellis' (2012) recommendation of a higher α (>.90) when developing a new scale to allow for α to deteriorate when used in a new research context. Further, high internal consistency estimates give some support for criterion validity and suggest that theoretically similar interventions are being delivered together (Barber, Sharpless, Klostermann, & McCarthy, 2007). However, Agbo (2010) argues that a factor analysis is needed to confirm unidimensionality of a scale.

Inter-rater reliability. Agreement between raters for the BCTS-D v2 on this single audio recorded therapy session was good for the overall scale (ICC = .83). Inter-rater reliability for individual items ranged from poor to fair (ICC ranged from 0 to .40). Some items were below zero, which indicates only random agreement between raters. If there are three or more coders some ICC may be less than 0, this can occur when there is a large number of raters and only a small number of tapes, which may make the correlations more error prone and more vulnerable to outliers (Magnusson, 1966; Nichols, 1999; Reichelt et al., 2003). Twelve items fell in the poor range and three fell in the

Scale development: BCTS-D

range of fair agreement. There is a big discrepancy between the ICC value for the overall scale compared to the ICC values for the individual items of the BCTS-D v2. According to Ebel (1972) within psychometric test theory it is possible for the total score on a test to be very reliable while the reliability of individual items is rather low and this is often due to error. Within this study five random stimulations of assigned raters were used to calculate the ICC values for individual items. These random assignments of raters into pairs might be subject to a lot of random error. Moreover, only having the data of one recording makes the data even more vulnerable. Hence, this big discrepancy between the overall ICC value and the individual item ICC values needs to be considered with these drawbacks in mind and might be due to a methodological artefact or a real difference in reliability of the overall scale compared to its individual items. When analysing the CTS-R and its inter-rater reliability for individual items, Blackburn and colleagues (2001) noted a wide fluctuation between pairs of raters, which lead to some individual items having very low inter-rater reliability while the scale displaying good overall reliability. Therefore, this issue will need further exploration for the BCTS-D v2 in future research using several recordings of different therapy sessions that were rated by two different raters.

Other methods used to explore the inter-rater agreement for individual items showed a similar picture with the variance of the scores indicating a high degree of variation for 12 items and low for three items. However, five items had an acceptable level of absolute agreement while the other 10 did not. These results are consistent with other studies of CBT competence

Scale development: BCTS-D

evaluation, which found good inter-reliability for individual items (ICC = -.14 to .84 [Blackburn et al., 2001]) difficult to achieve (Gordon, 2006).

There are several issues that could have led to low inter-rater reliability in the current study. Firstly, only one tape of a therapy session was rated, which means that ratings are statistically more prone to error. Secondly, after the rating session some participants reported being confused about how to rate items that were absent for appropriate reasons with some simply omitting the item or rating it at a zero. This will have negatively impacted on the inter-rater reliability. Thirdly, participants were sent the rating manual before the workshop as there was not enough time to go through the manual in a lot of detail during the workshop. However, most participants commented during the focus group not having been able to read the very comprehensive manual beforehand. The manual is essential in conceptualising what the items mean and what aspect of competence they represent. This could have led to participants having different understandings of competence and therefore rating the items differently. Fourthly, many raters (n=14) were novice therapists in BCT (even though they were experienced CBT therapists). As such their BCT knowledge and what BCT competence looks like might not have been consolidated enough. Finally, it may be that the BCTS-D v2 does not operationalize BCT competence in a helpful way.

To increase the likelihood of obtaining reliable ratings previous research has shown that assessor training is necessary to achieve adequate inter-rater reliability (Barber et al., 2007). This was confirmed for a range of clinician-rated scales (Gaur, Kaviani, Bansal, & Lee, 2010; Kobak, Engelhardt, & Lipsitz, 2006; Kobak, Feiger, & Lipsitz, 2005; Kobak, Opler, & Engelhardt,

Scale development: BCTS-D

2007; Müller et al., 1998; Rosen et al., 2008; Tracy, Adler, Rotrosen, Edson, & others, 1997) and clinical performance assessments (Bachrach, Mintz, & Luborsky, 1971; Lin et al., 2013) as well as CBT competence rating scales (Gordon, 2006; Reichelt et al., 2003). Training enables standardisation of assessors' definitions of competence and their interpretations of specific scale items. It also teaches assessors the specific performance indicators associated with each point on the rating scale, and provide assessors with practice improving familiarisation with the rating scale (Barber et al., 2007). Studies on the CTS-R have concluded that it's not enough for assessors to have general experience of assessing therapy skills and a good working knowledge of the scale (Gordon, 2006; Reichelt et al., 2003). Training on the scale is essential and standardized training across different practice sites is needed to increase inter-rater reliability (Barber et al., 2007) as systemic difference can arise between groups of raters with different training experiences (Jacobson, 1998). More recent research with the CTS-R drew the same conclusion (Gordon, 2006; Reichelt et al., 2003), emphasizing the need for structured training to achieve the reliable use of the scale. These findings are highly relevant to the successful implementation of the BCTS-D v2 and the effect of training upon inter-rater reliability needs to be determined by further research. However, rater training is only one of several measures to improve inter-rater agreement. Other actions should include further revisions of the BCTS-D v2 and its manual.

Limitations and future directions

The results of this study must be viewed in the context of a number of limitations. The sample size, the amount of data and range of data was not

Scale development: BCTS-D

sufficient for a robust analysis of the scale's psychometric properties. All quantitative analysis is based on ratings for one therapy session each for the BCTS-D v1 and BCTS-D v2. The aim of the first study was to establish if the CTS-R adapted to a couple context would be feasible. Therefore, the intention was to simply gain some insight if the BCTS-D v1's psychometrics were comparable to the psychometrics of the CTS-R, but more importantly gain feedback from BCT supervisors if the scale itself represented a valid and usable scale and how to improve it. Thereafter, a comprehensive manual was created and the scale was changed to improve face and content validity (BCTS-D v2). The second study had a similar aim with exploring the scale's reliability and collecting in-depth feedback about the scale's validity, usability and how to improve the scale further. Therefore, the scale validation data will need to be viewed with caution as it is only based on one therapy session (in both parts) which makes it more prone to be influenced by outliers causing more error in the score. Further, there are several risks to using a small sample when evaluating a rating scale (DeVellis, 2012). The covariation among items may not be stable, leading to inflated internal consistency. However, the internal consistency reported in this study was high, thus allowing for alpha deterioration in the future. A small sample also increases the likelihood that the sample may not be representative of the population for which the scale is intended. Further, the sample of experts and novice BCT therapists who participated in this study were all part of the BCT postgraduate diploma programme that has developed the BCTS-D. There were no independent BCT experts that participated in this study. Therefore, more efforts will need to go into exploring validity and usability with independent

Scale development: BCTS-D

experts to ensure the BCTS-D v2 covers all necessary competences of BCT. A further disadvantage to the small sample size is that it meant that the study was underpowered to complete exploratory factor analysis. For the purposes of principal component factor analysis, it has been suggested that a sample size of 5-10 participants is required per item (Osborne & Costello, 2004; Tinsley & Tinsley, 1987). The BCTS-D v2 has 15 items, meaning that between 75 and 150 ratings from *independent* therapists are required. Thus, in order to determine how many latent variables underlie the items and if this is comparable to the CTS (Young & Beck, 1980), which initially reported two underlying factors, but then confirmed one underlying factor (64.8% of the variance, overall cognitive therapy quality) in their analysis, it will be necessary to collect a much larger sample size and to use this to carry out a factor analysis. Doing a factor analysis would solve the question of how many constructs underlie the BCTS-D v2 since having high internal-consistency does not automatically confirm this. Instead it could point to item overlap and redundancy within the scale (Agbo, 2010). In addition to the factor analysis future research should focus on using several recordings, rated by at least two raters to explore reliability of the scale. In order to be able to do this research, some data protection issues within the NHS will need to be overcome (i.e. being able to keep the recording for a second rater to rate). Additionally, the research would need to be conducted on independent sites, which at the moment would be very difficult since BCT is only in its early stages of training therapists up all over England, which means there are a limited amount of BCT practitioners and supervisors in England. Then there

Scale development: BCTS-D

will be already well-known barriers to the research such as constraint time of all clinicians and limited resources.

Rater familiarisation with the scale and manual was another limitation. The BCTS-D v2 has a comprehensive manual including detailed descriptions of individual items and benchmarks for differing competence levels, but opportunities for familiarisation and training were limited. Most participants reported that they did not familiarise themselves with the manual before the rating session and rater training before scale use was limited to an hour, which focused on the purpose and the use of the scale. This is problematic as it may influence between rater agreement, with raters relying on their own interpretation of what an item measures rather than a common and shared understanding of the item. This may have contributed to the low IRR found in this study.

Further, research suggests that familiarisation through reading a manual is not enough to ensure good consistency between raters (Gordon, 2006; Reichelt et al., 2003). Thorough rater training is proposed as key to ensure inter-rater reliability. To explore this issue and get some clarity if reading a manual would be enough or if comprehensive training is needed to achieve adequate IRR on the BCTS-D v2, it would be necessary to study the impact of training. Ideally, a standardized training workshop is developed to train assessors in using the BCTS-D v2. This would include didactic aspects familiarising the assessors with the scale, its purpose and use. Then the training should aim to standardize assessors' interpretations of scale items by providing a clear explanation of each item, drawing on recordings of role-play sessions to provide concrete examples of the competence discussed in each

Scale development: BCTS-D

item, and finally giving the assessors the opportunity to practise conducting ratings on the BCTS-D v2 and discussing items which they have disagreed on. During such calibration meetings assessors seek to increase homogeneity of ratings through feedback and to discuss issues with certain item definitions or ratings (Nichols, 1999). As a second step, assessor agreement would need to be analysed within a naturalistic setting when assessors are provided with training. The study should also employ a control group (i.e. raters not receiving any training) in order to separate the 'learning effects' from the impact of the teaching session (Reichelt et al., 2003). The CTS-R was found to be reliable in a naturalistic setting when training is provided (Reichelt et al., 2003). This still needs to be determined for the BCTS-D v2. Moreover, it is important to note that evidence suggests that continued training is important when aiming to maintain a high level of inter-rater agreement. Many assessors stop using the manual after being more familiar with the scale which can lead to a rater drift (Foster, Bell-Dolan, & Burge, 1988a). Therefore, regular training sessions should be scheduled to prevent a drop in reliability (Reid & DeMaster, 1972; Taplin & Reid, 1973).

Following this recommendation, the BCT expert reference group has implemented continued training on the BCTS-D v2 within supervisor's supervision. The aim is to establish inter-rater reliability across pairs of supervisors, with two supervisors regularly rating the same therapy session on the BCTS-D v2 and comparing and discussing their ratings. This will give raters the opportunity to learn from one another and access more tapes for practice ratings.

Scale development: BCTS-D

Raters limited awareness of the therapeutic context may also have limited the validity of ratings. Therapeutic context is an important factor to consider when measuring competence (Stiles et al., 1998; Waltz et al., 1993) but in the current studies no context was available to the participants. Some information was given orally, but the participants had no opportunity to refer to written information while doing the rating. This might again have influenced the scores on the BCTS-D v2 and resulted in a greater diversity of scores and therefore lower inter-rater reliability. A summary sheet designed to describe the therapeutic context has been designed to accompany and influence rating (see Appendix S) but this was not used in the present study. The manner in which this contextual information should influence ratings is not detailed in the manual and the utility of this contextual information in improving ratings is yet to be determined. Both of these issues require further development and investigation in the refinement of the BCTS-D.

Unfortunately, it was not possible within this study to establish an empirically grounded cut-off point for competence on the BCTS-D v2. Previous research has criticized the CTS-R and other scales for having an arbitrary cut-off (a score about 39 is commonly considered a minimum competency score, with no item allowed to score less than a two; Blackburn et al., 2001). Formulation of threshold for competence is paramount especially since the BCTS-D v2 is going to be used within a postgraduate diploma training for BCT. As part of the accreditation process, trainees will need to achieve competence on the BCTS-D v2. However, to be able to situate this threshold in a useful manner, a reliable measuring tool that possesses sufficient depth, breadth and construct validity (Barber et al., 2007) is needed.

Scale development: BCTS-D

This study provided a step into the direction of a valid and reliable tool to measure BCT competence. The BCT expert reference group have articulated what level of competence is acceptable for the purpose of the postgraduate diploma training course in BCT and set the pass mark at a score of 45 (50%, range of score between 0 and 90). The cut-off for the CTS was calculated by taking the score one standard deviation below the mean of a group of certified cognitive therapists rated on the CTS (Shaw et al., 1999). This is still a somewhat arbitrary approach. Sharpless and Barber (2009) have therefore defined five key developmental stages within a clinical framework based on Dreyfus & Dreyfus' (1986) theory of competence which have been described as a useful basis for further development of competence thresholds (Muse & McManus, 2013). However, when considering a threshold for therapist competence it will need to be done within the context of the assessment. Someone passing an introductory therapy course may need lower scores than someone trying to get accredited (Muse & McManus, 2013). Unfortunately, it was beyond the scope of this study to validate if this threshold of competence is reliable and appropriate within the training of BCT therapists. In future research, it could be checked if the BCTS-D v2 manages to discriminate between BCT trainees and accredited BCT therapist based on the mentioned threshold as a first step towards validating that cut-off point.

It is also worth noting that the written and oral qualitative feedback collected in this study might have been influenced by demand characteristics (e.g. participants not wanting to give negative feedback about the scale which was then read/heard by the scale authors). Steps taken to mitigate against participants providing positively biased feedback (i.e. for the written feedback)

Scale development: BCTS-D

included providing therapists with anonymised questionnaires to return, confirming on the information sheet (see Appendix L) that all feedback would be both anonymous and confidential, and by clearly highlighting to participants that the purpose of the study was to identify potential problems and pitfalls with the scale. Additionally, at the beginning of the focus group the intent of the qualitative feedback was highlighted which was to gain in-depth insight into the opinions of a small group practising BCT to identify any aspects of the BCTS-D v2 which required further improvement and refinement.

Finally, it was beyond the scope of this study to determine discriminant and predictive validity of the BCTS-D v2. Especially within a diploma-level training setting it is expected for therapeutic competence to increase over a year-long course as trainee therapists develop their skills (McManus et al., 2010; Williams, Moorey, & Cobb, 1991). Measuring discriminant validity would indicate if the BCTS-D v2 could provide a useful tool for measuring therapists' progress within a BCT training programme, and this analysis should be included in future research. It will also be important to determine the predictive validity of the BCTS-D v2 - i.e. to examine whether the measure predicts improvements in patients' symptoms. Although all aspects of competence included within the BCTS-D v2 are theoretically important in delivering BCT, it is not actually known whether all of these aspects of competence are, in practice, necessary to achieve good patient outcomes. Therefore, more research needs to focus on dismantling studies identifying active components of BCT. That way training bodies and clinical supervisors would focus their efforts largely on developing those competencies that lead to improved patient

Scale development: BCTS-D

outcome (Keen & Freestone, 2008). Further examination of this issue is vital given that the ultimate goal of delivering competent BCT is to alleviate patients' symptoms.

Clinical and Scientific implications

Despite these limitations, this pilot study shows that the BCTS-D v2 demonstrates comparable internal-consistency and overall inter-rater reliability to the CTS-R, and has good face validity, content validity and usability. Therefore, it provides an important step towards measuring competence within BCT. However, limits of inter-rater reliability are of concern.

To be able to use the scale to measure competence within the training setting it was designed for (i.e. BCT postgraduate diploma training), measure development needs to attend to several unresolved issues. First, the threshold which a therapist is required to reach to demonstrate competence needs to be empirically validated and needs to take the purpose of the assessment into account as for example the threshold to pass an introductory BCT training program may be lower than the requirement for accreditation as a BCT therapist (Muse & McManus, 2013). Gordon (2006) even suggested that safeguards are needed such as lowering the threshold when using the CTS-R as a training outcome measure due to its relatively low inter-rater reliability. Second, in most training courses competence is inferred from rating one or two treatment sessions per therapist. Recent studies suggest however that a larger sample is needed to assess therapist competence reliably. Keen and Freestone (2008) talk about needing between 15 and 24 sessions rated to achieve a reliable assessment. It may be the case that postgraduate

Scale development: BCTS-D

courses can use routinely generated material to increase sampling of trainees' clinical work, although it is a continual challenge for educational institutions to both manage resources and deliver reliable and valid examinations. Third, the material assessed is usually selected by the therapists themselves, which might be a biased selection choosing the sessions that went well or the less complicated clients (Muse & McManus, 2013). Finally, assessors themselves need continued training and supervision when regularly rating therapist competence. Ideally, the training is standardised to increase inter-reliability including regular discussion and feedback session about how to calibrate ratings of certain treatment sessions.

Assessing therapist competence is not only important for the accreditation process of a therapist, but these competence checks can give supervisors information about the strengths and weaknesses of a therapist. They can highlight what types of interventions they are performing competently and in what areas additional training and supervisory feedback is required. This information can then be used to improve the quality of training and, ultimately, the treatment (Waltz et al., 1993).

Conclusion

This initial evaluation of the scale development shows that both versions of the BCTS-D demonstrate good face validity, content validity, and usability and provides a tool that is useful for providing formative feedback and promoting self-reflection. Further results suggest that the BCTS-D v2 has good internal consistency and overall inter-rater reliability, with inter-reliability for single items being a weak point. Hence the BCTS-D v2 appears to be suitable for use in clinical practice, training settings and research studies.

Scale development: BCTS-D

Several issues will need to be addressed in future research to ensure reliable use of the scale within a BCT training context (i.e. factor analysis to find out about underlying construct, validate threshold of the scale, etc). Finally, previous research shows that assessing competence within psychology is a complex process and needs to happen in a multi-trait, multi-method, multi-informant manner to be reliable, as no single method is able to provide a comprehensive assessment of all aspects of competence (Muse & McManus, 2013).

References

- Agbo, A. A. (2010). Cronbach's Alpha: Review of Limitations and Associated Recommendations. *Journal of Psychology in Africa, 20*(2), 233–239.
<https://doi.org/10.1080/14330237.2010.10820371>
- Anastasi, A., & Urbina, S. (1997). *Psychology testing*. New Jersey: Prentice Hall.
- Bachrach, H., Mintz, J., & Luborsky, L. (1971). On rating empathy and other psychotherapy variables: An experience with the effects of training. *Journal of Consulting and Clinical Psychology, 36*(3), 445.
- Barbato, A., & D'Avanzo, B. (2008). Efficacy of couple therapy as a treatment for depression: A meta-analysis. *Psychiatric Quarterly, 79*(2), 121–132.
- Barber, J., & Critis-Christoph, P. (1996). Development of a Therapist Adherence/Competence Rating Scale for Supportive-Expressive Dynamic Psychotherapy: A Preliminary Report. *Psychotherapy Research, 6*(2), 81–94.
<https://doi.org/10.1080/10503309612331331608>
- Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*(5), 493.
- Barker, C., Pistrang, N., & Elliott, R. (2002). *Research methods in clinical psychology: An introduction for students and practitioners*. John Wiley & Sons.
- Baucom, D. H., & Boeding, S. (2013). The Role of Theory and Research in the Practice of Cognitive-Behavioral Couple Therapy: If You Build It,

Scale development: BCTS-D

They Will Come. *Behavior Therapy*, 44(4), 592–602.

<https://doi.org/10.1016/j.beth.2012.12.004>

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 ed., pp. 23–60). New York: The 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(1), 53.

Baucom, D. H., Whisman, M. A., & Paprocki, C. (2012). Couple-based interventions for psychopathology. *Journal of Family Therapy*, 34(3), 250–270. <https://doi.org/10.1111/j.1467-6427.2012.00600.x>

Beach, S. R. H., & Daniel O’Leary, K. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy*, 23(4), 507–528. [https://doi.org/10.1016/S0005-7894\(05\)80219-9](https://doi.org/10.1016/S0005-7894(05)80219-9)

Bennett-Levy, J. (2006). Therapist skills: A cognitive model of their acquisition and refinement. *Behavioural and Cognitive Psychotherapy*, 34(1), 57–78.

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(4), 431–446.

Scale development: BCTS-D

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj*, 314(7080), 572.

Blow, A. J., Sprenkle, D. H., & Davis, S. D. (2007). Is who delivers the treatment more important than the treatment itself? The role of the therapist in common factors. *Journal of Marital and Family Therapy*, 33(3), 298–317.

Bodenmann, G., Plancherel, B., Beach, S. R., Widmer, K., Gabriel, B., Meuwly, N., ... Schramm, E. (2008). Effects of coping-oriented couples therapy on depression: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 76(6), 944.

BPS guidelines to assess competence within clinical psychology. (2010).

Retrieved from

https://www.google.co.uk/search?client=safari&rls=en&q=BPS+guidlines+to+assess+competence+within+clinical+psychology&ie=UTF-8&oe=UTF-8&gfe_rd=cr&ei=6cohWfDFOejW8gfey7GgCg

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.

Brewer, J., & Hunter, A. (2006). Foundations of multimethod research: Synthesizing styles. Sage.

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*, 253–264.

Central London CBT Training Centre. (2016, 2017). Post Graduate Certificate in Behavioural Couples Therapy.

Scale development: BCTS-D

- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7.
- 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*(2), 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: Norton.
- 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*(11), 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
- Corrie, S., Fischer, M. S., Worrell, M., & Baucom, D. H. (n.d.). Behavioural Couple Therapy Scale - Depression (BCTS-D) The Manual. *In Prep*.
- Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., ... Zitrin, C. (1991). Meta-Analysis of Therapist Effects in

Scale development: BCTS-D

Psychotherapy Outcome Studies. *Psychotherapy Research*, 1(2), 81–91. <https://doi.org/10.1080/10503309112331335511>

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests.

Psychometrika, 16(3), 297–334.

Davidson, K., Scott, J., Schmidt, U., Tata, P., Thornton, S., & Tyrer, P. (2004).

Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy Trial: the POPMACT study. *Psychological Medicine*, 34(5), 855–863.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (Vol. 26).

Sage publications.

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(4), 295–300.

Dobson, K. S., & Singer, A. R. (2005). Definitional and practical issues in the

assessment of treatment integrity. *Clinical Psychology: Science and Practice*, 12(4), 384–387.

Dracup, C. (1997). The reliability of marking on a psychology degree. *British*

Journal of Psychology, 88(4), 691–708.

Dreyfus, H., & Dreyfus, S. (1986). Mind over machines. *Mind over Machines*.

Elkin, I. (1999). A Major Dilemma in Psychotherapy Outcome Research:

Disentangling Therapists From Therapies. *Clinical Psychology: Science and Practice*, 6(1), 10–32. <https://doi.org/10.1093/clipsy.6.1.10>

Epstein, N. B., & Baucom, D. H. (2002). *Enhanced cognitive-behavioral*

therapy for couples: A contextual approach. American Psychological

Scale development: BCTS-D

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*(6–7), 373–378. <https://doi.org/10.1016/j.brat.2011.03.005>

Fischer, M. S., Baucom, D. H., & Cohen, M. J. (2016). Cognitive-Behavioral Couple Therapies: Review of the Evidence for the Treatment of Relationship Distress, Psychopathology, and Chronic Health Conditions. *Family Process*, *55*(3), 423–442. <https://doi.org/10.1111/famp.12227>

Foley, S. H., O'Malley, S., Rounsaville, B., Prusoff, B. A., & Weissman, M. M. (1987). The relationship of patient difficulty to therapist performance in interpersonal psychotherapy of depression. *Journal of Affective Disorders*, *12*(3), 207–217.

Foster, S. L., Bell-Dolan, D. J., & Burge, D. A. (1988). Behavioral observation. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook*, 3rd ed (pp. 119–160). Elmsford, NY, US: Pergamon Press.

Gaur, R., Kaviani, H., Bansal, S., & Lee, N. (2010). RATER TRAINING ON PANSS AND SANS. *Schizophrenia Research*, *117*(2), 265–266.

Gordon, P. K. (2007). A Comparison of Two Versions of the Cognitive Therapy Scale. *Behavioural and Cognitive Psychotherapy*, *35*(3), 343–353. <https://doi.org/10.1017/S1352465806003390>

Scale development: BCTS-D

Gupta, M., Coyne, J. C., & Beach, S. R. (2003). Couples treatment for major depression: Critique of the literature and suggestions for some different directions. *Journal of Family Therapy*, 25(4), 317–346.

Gurman. (2015). The Theory and Practice of Couple Therapy. In A. S. Gurman, J. . Lebow, & D. K. Snyder (Eds.), *Clinical Handbook of Couple Therapy* (5th ed., pp. 1–22). New York: Guilford Press.

Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., ... Harris, N. (2001). An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy*, 29(2), 221–233.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238.

Hewison, D. (2011). *Curriculum for Couple Therapy for Depression*. National IAPT Programme Team.

Holland, R. (2006). Minimum training standards for the practice of CBT. *BABCP, Bury: British Association for Behavioural and Cognitive Psychotherapies*.

Holtzworth-Munroe, A., Jacobson, N. S., DeKlyen, M., & Whisman, M. A. (1989). Relationship between behavioral marital therapy outcome and process variables. *Journal of Consulting and Clinical Psychology*, 57(5), 658.

Howitt, D., & Cramer, D. (2011). Introduction to research methods in psychology. Pearson Education.

Scale development: BCTS-D

Jacobson, N. S. (1998). How important are cognitive interventions in cognitive therapy. In *World Congress of Behavioral and Cognitive Therapies, Acapulco, Mexico*.

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(1), 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(2), 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.

James, I. A., Blackburn, I.-M., Milne, D. L., & Reichelt, F. K. (2001). Moderators of trainee therapists' competence in cognitive therapy. *British Journal of Clinical Psychology*, 40(2), 131–141.

Kalton, G., & Stowell, R. (1979). A study of coder variability. *Applied Statistics*, 276–289.

Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist*, 59(8), 774.

Kazantzis, N. (2003). Therapist competence in cognitive-behavioural therapies: Review of the contemporary empirical evidence. *Behaviour Change*, 20(1), 1–12.

Scale development: BCTS-D

- Keen, A. J., & Freeston, M. H. (2008). Assessing competence in cognitive-behavioural therapy. *The British Journal of Psychiatry*, *193*(1), 60–64.
- Kobak, K. A., Engelhardt, N., & Lipsitz, J. D. (2006). Enriched rater training using Internet based technologies: a comparison to traditional rater training in a multi-site depression trial. *Journal of Psychiatric Research*, *40*(3), 192–199.
- Kobak, K. A., Feiger, A. D., & Lipsitz, J. D. (2005). Interview quality and signal detection in clinical trials. *American Journal of Psychiatry*, *162*(3), 628–628.
- Kobak, K. A., Opler, M. G., & Engelhardt, N. (2007). PANSS rater training using Internet and videoconference: Results from a pilot study. *Schizophrenia Research*, *92*(1), 63–67.
- Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research* (4th ed.). London: Sage Publications.
- Kvale, S. (1996). *InterViews. An introduction to qualitative research writing*. Thousand Oaks, CA: Sage Publications.
- Lin, C.-J., Chang, J. Z.-C., Hsu, T.-C., Liu, Y.-J., Yu, S.-H., Tsai, S. S.-L., ...
Lin, C.-P. (2013). Correlation of rater training and reliability in performance assessment: Experience in a school of dentistry. *Journal of Dental Sciences*, *8*(3), 256–260.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–386.
- Magnusson, D. (1966). *Test theory*. Reading: MA: Addison-Wesley.
- McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments. A review

Scale development: BCTS-D

of current efforts. *The American Psychologist*, 65(2), 73–84.

<https://doi.org/10.1037/a0018121>

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(11), 1123–1132.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(9), S63–7.

Müller, M. J., Rossbach, W., Dannigkeit, P., Müller-Siecheneder, F., Szegedi, A., & Wetzell, H. (1998). Evaluation of standardized rater training for the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Research*, 32(3), 151–160.

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33(3), 484–499.
<https://doi.org/10.1016/j.cpr.2013.01.010>

Muse, K., McManus, F., Rakovshik, S., & Thwaites, R. (2017). Development and psychometric evaluation of the Assessment of Core CBT Skills (ACCS): An observation-based tool for assessing cognitive behavioral therapy competence. *Psychological Assessment*, 29(5), 542.

National Institute for Clinical Excellence. (2009). *Depression in adults: recognition and management (clinical guideline 90)*. London: NICE.

Nichols, D. P. (1999). My Coefficient a is Negative!, (Number 68).

Scale development: BCTS-D

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation, 9*(11), 8.

oxford dictionary competence. (n.d.). Retrieved May 21, 2017, from https://www.google.co.uk/search?client=safari&rls=en&q=oxford+dictionary+competence&ie=UTF-8&oe=UTF-8&gfe_rd=cr&ei=cc0hWbbwF-jW8gfey7GgCg

Perepletchikova, F., Hilt, L. M., Chereji, E., & Kazdin, A. E. (2009). Barriers to Implementing Treatment Integrity Procedures: Survey of Treatment Outcome Researchers. *Journal of Consulting and Clinical Psychology, 77*(2), 212–218. <https://doi.org/10.1037/a0015232>

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*(5), 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*(2), 87–99.

Reid, J. B., & DeMaster, B. (1972). *The efficacy of the spot-check procedure in maintaining the reliability of data collected by observers in quasi-natural settings: two pilot studies*. Oregon Research Institute.

Rosen, J., Mulsant, B. H., Marino, P., Groening, C., Young, R. C., & Fox, D. (2008). Web-based training and interrater reliability testing for scoring the Hamilton Depression Rating Scale. *Psychiatry Research, 161*(1), 126–130.

Scale development: BCTS-D

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/>

Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, 36(2), 129–147.

Shadish, W. R., & Baldwin, S. A. (2003). Meta-analysis of MFT interventions. *Journal of Marital and Family Therapy*, 29(4), 547–570.

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(1), 47–56.

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(6), 837.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420.

Smith, J. A. (2015). *Qualitative psychology: A practical guide to research methods*. Sage.

Stallard, P., Myles, P., & Branson, A. (2014). The cognitive behaviour therapy scale for children and young people (CBTS-CYP): Development and psychometric properties. *Behavioural and Cognitive Psychotherapy*, 42(3), 269–282.

Scale development: BCTS-D

- Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, 5(4), 439–458.
- Strauss, A., & Corbin, J. M. (1997). *Grounded theory in practice*. Sage.
- Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist Competence in Cognitive Therapy for Depression: Predicting Subsequent Symptom Change. *Journal of Consulting and Clinical Psychology*, 78(3), 429–437. <https://doi.org/10.1037/a0019631>
- Taplin, P. S., & Reid, J. B. (1973). Effects of instructional set and experimenter influence on observer reliability. *Child Development*, 547–554.
- Tinsley, H. E., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, 34(4), 414.
- Tracy, K., Adler, L. A., Rotrosen, J., Edson, R., & others. (1997). Interrater reliability issues in multicenter trials, Part I: Theoretical concepts and operational procedures used in Department of Veterans Affairs Cooperative Study# 394. *Psychopharmacology Bulletin*, 33(1), 53.
- 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(2), 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(3), 381.
- Vaughn, S., Schumm, J. S., & Sinagub, J. M. (1996). *Focus group interviews in education and psychology*. Sage.

Scale development: BCTS-D

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy*, 47(2), 119–127.

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(4), 620.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. Retrieved from <http://psycnet.apa.org/journals/ccp/78/2/200/>

Weck, F., Bohn, C., Ginzburg, D. M., & Stangier, U. (2011). Treatment integrity: implementation, assessment, evaluation, and correlations with outcome. *Verhaltenstherapie*, 21(2), 99–107.

Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin*, 114(2), 248–265.
<https://doi.org/10.1037/0033-2909.114.2.248>

Whisman, M. A. (1999). Marital dissatisfaction and psychiatric disorders: results from the National Comorbidity Survey. *Journal of Abnormal Psychology*, 108(4), 701.

Whisman, M. A. (2001). Marital adjustment and outcome following treatments for depression. *Journal of Consulting and Clinical Psychology*, 69(1), 125–129. <https://doi.org/10.1037/0022-006X.69.1.125>

Whisman, M. A. (2007). Marital distress and DSM-IV psychiatric disorders in a population-based national survey. *Journal of Abnormal Psychology*, 116(3), 638.

Scale development: BCTS-D

Whisman, M. A., & Baucom, D. H. (2012). Intimate relationships and psychopathology. *Clinical Child and Family Psychology Review*, 15(1), 4–13.

Young, J., & Beck, A. T. (1980). *Cognitive Therapy Scale*. *Unpublished Assessment Instrument*. Philadelphia, Pennsylvania: University of Pennsylvania.

Appendices

Appendix A: Literature search table

Search database	Search terms used		How many articles found	How many related to couple therapy and competence
PsychInfo and PsycARTICLES ¹⁷	Competency scale	BCT	0	0
		Behavioural couples therapy	0	0
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	0	0
		Couples therapy	0	0
	Therapist competency	BCT	0	0
		Behavioural couples therapy	0	0
		ECBCT	0	0

¹⁷ Search term was applied to abstracts

Scale development: BCTS-D

		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	0	0
		Couples therapy	0	0
	Clinical skill	BCT	5	0
		Behavioural couples therapy	46	0
		ECBCT	1	0
		Enhanced cognitive behavioural therapy for couples	2	0
		CBT for couples	8	0
		cognitive behavioural therapy for couples	25	
		Couples therapy	148	2
	Assessing competence/y	BCT	0	0
		Behavioural couples therapy	0	0
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0

Scale development: BCTS-D

		CBT for couples	0	0
		cognitive behavioural therapy for couples	0	0
		Couples therapy	0	0
	Competency intervention	BCT	0	0
		Behavioural couples therapy	2	1
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	1	0
		Couples therapy	16	1
Science Direct	Competency scale	BCT	55	0
		Behavioural couples therapy	3	0
		ECBCT	0	
		Enhanced cognitive behavioural therapy for couples	0	
		CBT for couples	0	
		cognitive behavioural therapy for couples	0	

Scale development: BCTS-D

		Couples therapy	30	0
	Therapist competency	BCT	26	0
		Behavioural couples therapy	2	0
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	0	0
		Couples therapy	44	0
	Clinical skill	BCT	392	1
		Behavioural couples therapy	8	0
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	1	0
		cognitive behavioural therapy for couples	0	0
		Couples therapy	402	4
	Assessing competence/y	BCT	71	0
		Behavioural couples	0	0

Scale development: BCTS-D

		therapy		
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	0	0
		Couples therapy	25	0
	Competency intervention	BCT	59	0
		Behavioural couples therapy	0	0
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	0	0
		Couples therapy	26	0
PubMed / MEDLINE	Competency scale	BCT	0	0
		Behavioural couples therapy	0	0
		ECBCT	0	0
		Enhanced cognitive	0	0

Scale development: BCTS-D

		behavioural therapy for couples		
		CBT for couples	0	0
		cognitive behavioural therapy for couples	1	0
		Couples therapy	3	0
	Therapist competence/y	BCT	0	0
		Behavioural couples therapy	0	0
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	2	1
		Couples therapy	9	2
	Clinical skill	BCT	10	0
		Behavioural couples therapy	1	0
		ECBCT	26	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural	14	1

Scale development: BCTS-D

		therapy for couples		
		Couples therapy	42	6
	Assessing competence/y	BCT	2	0
		Behavioural couples therapy	0	
		ECBCT	0	
		Enhanced cognitive behavioural therapy for couples	0	
		CBT for couples	0	
		cognitive behavioural therapy for couples	2	0
		Couples therapy	6	1
	Competency intervention	BCT	3	0
		Behavioural couples therapy	1	0
		ECBCT	0	0
		Enhanced cognitive behavioural therapy for couples	0	0
		CBT for couples	0	0
		cognitive behavioural therapy for couples	4	0
		Couples therapy	13	1

Appendix B: Behavioral Couple Therapy Competence Rating Scale

Behavioral Couple Therapy Competence Rating Scale

Therapist _____ Client # _____ Session Date _____ Rating Date _____

Directions: For each item, assess the therapist on a scale of 0 to 6, and record the rating next to the item number. Descriptions are provided for even-number scale points. *If you believe the therapist falls between two of the descriptors, select the intervening odd number (1,3,5).* For example, if the therapist set a very good agenda but did not establish priorities, assign as rating of 5 rather than 4 or 6.

If the descriptions for a given item occasionally do not seem to apply to the session you are rating, feel free to disregard them and use the more general scale below:

0	1	2	3	4	5	6
Poor		Mediocre		Good		Excellent

Part I. General Therapeutic Skills

- _____ 1. AGENDA
 - 0 Therapist did not set agenda.
 - 2 Therapist set agenda that was vague or incomplete.
 - 4 Therapist worked with clients to set a mutually satisfactory agenda that included specific target problems (e.g., learning a problem solving skill).
 - 6 Therapist worked with the client to set appropriate agenda with target problems, suitable for the available time. Established priorities and followed agenda.

- _____ 2. STRUCTURING
 - 0 The clients controlled the pacing of the session.
 - 2 The therapist had some control, but the clients frequently wrestled it away from her/him.
 - 4 The therapist controlled the pacing of the session, and made sure the clients did not get him/her off track.
 - 6 The therapist controlled the pacing of the session, made sure that clients did not get her/him off track, yet was not dismissive or invalidating in the manner the control was maintained.

- _____ 3. NEUTRALITY
 - 0 Therapist blatantly sided with one partner, by being supportive and validating of one while being critical and hostile of the other.
 - 2 For portions of the session, therapeutic neutrality was violated, and the therapist at times seemed to be hostile to one partner **and/or** exclusively validating and supportive of the other.
 - 4 The therapist was neutral, either by avoiding a critical, blaming tone toward either spouse **or** by being supportive and validating of both in an active way.

Scale development: BCTS-D

- 6 The therapist was neutral, both by avoiding a critical, blaming tone toward either spouse **and** by being supportive and validating of both in an active way.

4. UNDERSTANDING

- 0 Therapist repeatedly failed to understand what client(s) said and thus consistently missed the point. Poor empathetic skills.
- 2 Therapist was at times able to reflect or rephrase what the client explicitly said but repeatedly failed to respond to more subtle communication. Limited ability to listen and empathize.
- 4 Therapist generally seemed to grasp the client's "point of view" as reflected in both what the client explicitly said and what the client communicated in more subtle ways. Good ability to listen and empathize.

5. INTERPERSONAL EFFECTIVENESS

- 0 Therapist had poor interpersonal skills. Seemed hostile, demanding, or in some other way destructive to the client.
- 2 Therapist did not seem destructive, but had significant interpersonal problems. At times, therapist appeared unnecessarily impatient, aloof, insincere, or had difficulty conveying confidence.
- 4 Therapist displayed a satisfactory degree of warmth, concern, confidence, genuineness, and professionalism, appropriate for this particular couple in this session.
- 6 Therapist displayed optimal levels of warmth, concern, confidence, genuineness, and professionalism, appropriate for this particular couple in this session.

6. QUALITY OF THERAPEUTIC ALLIANCE

- 0 Therapist did not attempt to work with couple on a problem which seemed relevant, and couple seemed to be fighting with therapist when therapist tried to make things happen in the session.
- 2 At times, the therapist and clients worked well together, but there were frequent ruptures in their working relationship, and the work was either less efficient as a result or there were periods where nothing seemed to be happening.
- 4 The therapist and partners worked well together, even though there were occasional digressions where the work seemed tangential.
- 6 The therapist and client worked optimally during this session, and the work was focused without rupture or diversion.

Part II. Conceptualization, Strategy, and Technique

7. BEHAVING CONSISTENTLY WITH CONCEPTUALIZATION OF CASE

- 0 Therapist behaved in a way that was inconsistent with the conceptualization of the case, as reflected in the assessment sessions and the feedback session.
- 2 Therapist showed some awareness of the treatment plan, but frequently seemed to do things that might have been appropriate for some couples, but were clearly not appropriate for this couple.
- 4 Therapist generally behaved in a way which was consistent with case conceptualization and treatment plan, but occasionally did things which seemed inconsistent with that plan.
- 6 Therapist consistently behaved in accordance with the treatment plan, with no examples of interventions which didn't make sense in light of that plan.

Scale development: BCTS-D

- _____ 8. STRATEGY
- 0 Therapist's overall strategy for the session seemed inappropriate, given the stage of therapy and the immediate context of the previous session.
 - 2 Therapist's overall strategy was plausible, but equally compelling or better strategies were imaginable, given the stage of therapy and the immediate context of the previous session.
 - 4 Therapist's overall strategy made good sense, in light of the stage of therapy and the immediate context of the previous session.
 - 6 Therapist chose the optimal strategy, given the stage of therapy and the immediate context of the previous session.
- _____ 9. IMPLEMENTATION OF TECHNIQUES (this item is weighted higher than others so, we use a scale of 0-12)
- 0 Therapist did a poor job implementing whatever techniques were used during this session.
 - 4 Therapist did a mediocre job of implementing whatever techniques were used during this session.
 - 8 Therapist did a good job implementing whatever techniques were used during this session.
 - 12 Therapist did an excellent job implementing whatever techniques were used during this session.
- _____ 10. HOMEWORK
- 0 Therapist either forgot to debrief last week's homework or failed to give an assignment at all, despite the appropriateness of homework for this couple at this time.
 - 2 Therapist did either perfunctory debriefing of last week's assignment or presented next week's assignment in a vague way or in a way which made compliance questionable.
 - 4 Therapist did a good job debriefing homework and presented the next assignment in a way likely to promote compliance.
 - 6 Therapist did an excellent job debriefing homework from last week, and in presenting next week's assignment.

Part III. Additional Considerations

11 (a). Did any special problems arise during the session (e.g. suicide threats)?

YES

NO

_____ (b). If yes:

- 0 Therapist could not deal adequately with special problems that arose.
- 2 Therapist dealt with special problems adequately, but used strategies or conceptualizations inconsistent with BMT.
- 4 Therapist attempted to deal with special problems using a BMT framework and was moderately skillful in applying techniques.
- 6 Therapist was very skillful at handling special problems using a BMT framework.

Scale development: BCTS-D

12. Were there any significant unusual factors in this session that you feel justified the therapist's departure from the standard approach measured by this scale?

YES (Please explain below) NO

Part IV. Overall ratings and comments

13. How would you rate the therapist overall in this session, as a behavioral marital therapist?

0	1	2	3	4	5	6
Poor	Barely	Mediocre	Satisfactory	Good	Very Good	
	Excellent					
	Adequate					

14. If you were conducting an outcome study in BMT, do you think you would select this therapist to participate at this time (assuming this session is typical)?

0	1	2	3	4
Definitely	Probably	Uncertain	Probably	
Definitely				
Not	Not		Yes	Yes

15. How difficult did you feel this couple was to work with?

0	1	2	3	4	5	6
Not			Moderately			
	Extremely					
Difficult			Difficult			
	Difficult					

16. Comments and suggestions for therapist's improvement:

Appendix C: Overview of data analysis

	Project	Data analysis
<p>Study 1:</p>	<p>Initial small scale psychometric evaluation of the BCTS-D v1 (an adapted version of the CTS-R) to ensure its comparability to the CTS-R. Further, gain feedback on its usability and relevance of the BCTS-D v1 within the context of couple therapy.</p> <ul style="list-style-type: none"> • BCT supervisors (n=14) rating the competency of a training case on the BCTS-D v1 • BCT supervisors (n=14) filling in a feedback questionnaire about the BCTS-D v1 to gain some insight on general content validity, usability and how to improve the scale. 	<ul style="list-style-type: none"> • Description of sample: gender, experience of CBT and BCT, number of cases treated in CBT and BCT, familiar with CTS-R and how often use it / profession • Inter-item reliability: calculate Cronbach's alpha for the overall scale to check if it is comparable to the CTS-R • Inter-rater reliability – calculate the ICC for the overall scale to check if the inter-rater reliability of the BCTS-D v1 is comparable to the CTS-R. • Calculate mean (M) and standard deviation (SD) for validity and usability questions: <ul style="list-style-type: none"> - opportunity for feedback (yes,no) - any aspects of competence missing in the scale (yes/no) - style and appearance of the scale (1-4) - ease of use of the scale (1-4) • Thematic analysis of qualitative feedback to identify themes of how to change and improve the scale was used (by Braun and Clarke, 2006): <ul style="list-style-type: none"> - what missing re competence - what missing re feedback - revisions to improve style - revisions to improve usability - other comments, how to improve scale • Further a respondent validation will be done by the expert group evaluating all feedback given within the scale development.

Scale development: BCTS-D

	Development of the BCTS-D v2	<ul style="list-style-type: none"> Expert group revised the BCTS-D v1 using the psychometric properties and the qualitative feedback gained in phase 1.
Study 2:	<p>This phase is a more robust psychometric evaluation of the BCTS-D v2 (further developed scale) and includes a written and oral feedback study to evaluate face validity, content validity and usability.</p> <ul style="list-style-type: none"> BCT supervisors (n=6) and BCT novices (n=14) rating the competency of a training case on the BCTS-D v2 	<ul style="list-style-type: none"> Description of sample: gender, experience of CBT and BCT, number of cases treated in CBT and BCT, familiar with CTS-R and how often use it / profession Internal consistency: Calculate Cronbach's alpha and Cronbach's alpha if an item was to be deleted (to identify if alpha would increase by deleting an item) Inter-rater reliability: <ul style="list-style-type: none"> Calculate ICC for the overall scale and its three domains (two way, mixed effects, absolute agreement, single measures) Calculating inter-rater reliability for each item triangulating several methods to explore IRR as no clear guidelines on how to calculate IRR when only one rating exists. Explored the ICC, the variance and the p-bar for each item.
	<ul style="list-style-type: none"> BCT supervisors (n=6) and BCT novices (n=14) filling in a questionnaire about the BCTS-D v2 giving written feedback about the scale's content validity, usability and how to improve the scale. 	<ul style="list-style-type: none"> Calculate mean (M) and standard deviation (SD) for validity and usability questions: <ul style="list-style-type: none"> scale provides space to judge competency (1-4) the scale gives opportunity for useful feedback (1-4) the scale has an appropriate scoring system (1-4) any aspects of competence missing in the scale (yes/no) style and appearance of the scale (1-4) ease of use of the scale (1-4) Calculate the CVI (content validity index) to confirm content validity: (i.e. percentage of participants who rated the item as both relevant and clear - a rating of three or four on the four-point scale).

Scale development: BCTS-D

	<ul style="list-style-type: none">• BCT supervisors (n=6) and BCT novices (n=14) joining a focus group to give more in depth feedback about the BCTS-D v2 and gain insight into group consensus about the scale and its psychometric properties.	<ul style="list-style-type: none">• Thematic analysis of both the written qualitative feedback and the feedback given within the focus group (recorded and transcribed) to identify themes of how to change and improve the scale was used (by Braun and Clarke, 2006):<ul style="list-style-type: none">- what missing re competence- what missing re feedback- revisions to improve style- revisions to improve usability- other comments, how to improve scale- feedback about specific items• Further a respondent validation will be done by the expert group evaluating all feedback given within the scale development.
--	--	--

Scale development: BCTS-D

Appendix D: BCTS – D version 1

Behavioural Couple Therapy Scale – Depression Version (BCTS-D)

Adapted from the CTS-R

April 20, 2015 Version

Scale development: BCTS-D

The Behavioural Couple Therapy Scale- Depression Version (BCTS-D) is adapted from the Cognitive Therapy Scale- Revised (CTS-R). It is intended for use in rating behavioural couple therapy (BCT) sessions in which one person has depression. In such instances, the couple often has relationship distress as well as one partner (or both partners) being depressed. Therefore, the CTS-R was modified in order to make the scale appropriate for rating the therapist's degree of competence in working within a couple and depression context. As a result, two types of changes were made to the CTS-R in creating the BCTS-D. *First, existing CTS-R items were altered in wording* to reflect a couple rather than an individual was the focus of treatment, along with salient couple emphases that are important within an existing item on the CTS-R (e.g., pacing of a couple session is a particular challenge in BCT because of disagreements that frequently erupt between partners). *Second, one new item was added* that reflects the importance of the therapist selecting an appropriate focus for the session, emphasizing a direct focus on the patient's depression or the couple's relationship which often is related to the depression. As a result of the above changes, the BCTS-D contains 13 items compared to 12 items in the CTS-R.

Scale development: BCTS-D

Name of Student.....

Date:

BCTS Items	Score	Comments
<p>1. Agenda setting and adherence</p> <p>Did the therapist set a good agenda and adhere to it?</p>		
<p>2. Feedback</p> <p>Were there statements and/or actions concerned with providing and eliciting feedback?</p>		
<p>3. Collaboration</p> <p>Were there statements and/or actions encouraging both members of the couple to participate appropriately, and preventing an unequal power relationship from developing between (a) the couple and the therapist and (b) the two partners?</p>		
<p>4.Pacing and efficient use of time</p> <p>Were there statements and/or actions concerning the pacing of the session, helping to ensure the time was used effectively, including the therapist's ability to control the timing and focus of the session with two partners present?</p>		
<p>5. Interpersonal Effectiveness</p> <p>Was a good therapeutic relationship evident (trust, warmth, etc.) with each partner individually and the couple as a unit? Was a safe environment created for the couple to address difficult issues in session?</p>		
<p>6. Focus on depression in context</p> <p>Were there statements and/or actions designed to focus the session on relevant individual, couple, or environmental factors impacting the patient's depression or relationship distress?</p>		

Scale development: BCTS-D

7. Eliciting appropriate emotional expression Were there questions and/or actions designed to elicit relevant emotions from both partners and promote a good emotional ambience?		
8. Eliciting key cognitions Were there questions and/or actions designed to elicit relevant cognitions (thoughts, beliefs, etc.) from both partners?		
9. Eliciting and planning behaviours Were there questions and/or actions designed to elicit dysfunctional individual behaviours or couple interaction patterns (e.g., negative reciprocity) and engage the couple in planning for change?		
10. Guided discovery Were there questions and/or actions designed to promote self-reflection, helping the couple to make their own connections and discoveries?		
11. Conceptual integration Were there statements and/or actions designed to promote the couple's understanding of the models underpinning BCT?		
12. Application of change methods Did the therapist facilitate in-session learning and change through an appropriate change method (cognitive, emotional, or behavioural)?		
13. Homework setting Did the therapist set an appropriate homework task effectively?		

Scoring: Scoring for the BCT-S involves two scores. First, the 12 items adapted from the CTS-R are summed to provide a score analogous to the CTS-R; this includes summing all items with the exception of item 6. Second a total score is created by adding the score for item 6 to the above sum, thus reflecting the sum of all 13 items.

CTS-R comparable score (sum of all items excluding item 6)=

Total score: (sum of items 1-13)

Student Learning and Action Points

Signed Student:

Signed Supervisor/Assessor:

Date:

Scale development: BCTS-D

Appendix E: Items of the BCTS-D v1 versus CTS-R

BCTS-D version 1	CTS-R
Agenda setting and adherence	Agenda setting and Adherence
Feedback	Feedback
Collaboration	Collaboration
Pacing and efficient use of time	Pacing and Efficient use of Time
Interpersonal Effectiveness	Interpersonal Effectiveness
Focus on Depression in context*	
Eliciting appropriate emotional expression	Eliciting appropriate emotional expression
Eliciting key cognitions	Eliciting key cognitions
Eliciting and planning behaviours	Eliciting behaviours
Guided discovery	Guided discovery
Conceptual integration	Conceptual integration
Application of change method	Application of change method
Homework setting	Homework setting

Scale development: BCTS-D

Appendix F: BCTS-D version 2

BEHAVIOURAL COUPLE THERAPY SCALE – DEPRESSION (BCTS-D)

Sarah Corrie, Melanie S. Fischer, Michael Worrell, & Donald H. Baucom

© The Central London CBT Training Centre, CNWL Foundation Trust

Name of Therapist:

Date of session with couple:

Session No.:

Name of Assessor:

<u>BCTS-D Item</u>	<u>Score</u> 0-6	<u>Comments</u>
<p><i>Item 1. Agenda Setting</i></p> <p>Did the therapist establish a focus for the session? Was a clear agenda set, with priorities identified that provided a clear 'steer' for the session that was to follow?</p>		
<p><i>Item 2. Review of Homework (from the previous session)</i></p> <p>Did the therapist review the homework agreed upon at the previous session? Were new insights or learning elicited? If needed, were any obstacles identified and explored to enable 'follow through' with future homework tasks?</p>		
<p><i>Item 3. Collaboration</i></p> <p>Did the therapist facilitate effective teamwork, promoting an adaptive working relationship between the therapist and couple, in which appropriate levels of transparency and mutual feedback were encouraged? Were statements and/or actions made that encouraged both members of the couple to participate appropriately, and which prevented an unequal power relationship from developing between the couple and the therapist?</p>		

<p><i>Item 4. Facilitating Couple Communication</i></p> <p>Did the therapist facilitate couple communication effectively? Was there evidence of enabling understanding and dialogue between partners through steering the conversation towards reciprocal listening, 'reflecting back' and enhanced empathy? Was the therapist able to provide sufficient structure and management to enable the couple to communicate more effectively without overly-controlling the process?</p>		
<p><i>Item 5. Pacing and Flow</i></p> <p>Did the therapist pace the session well, enabling a smooth transition and sense of flow between different phases of the session, ensuring that priority items on the agenda were addressed? Was the therapist able to control the timing, pacing and flow of the session with two partners present?</p>		
<p><i>Item 6. Therapist's Interpersonal Effectiveness</i></p> <p>Did the therapist create an interpersonal environment that was conducive to the work of the session? Did the therapist form an effective alliance with each partner individually and the couple as a unit? Was a safe environment created for the couple to address difficult issues in session? Did the therapist's interpersonal style convey warmth, interest, and instil hope that change is possible, as well as convey appropriate authority in the way the session was managed?</p>		
<p><i>Item 7. Guided Discovery</i></p> <p>Did the therapist facilitate a process of discovery that conveyed genuine interest, curiosity, and understanding? Were there questions and/or actions designed to promote self-reflection,</p>		

helping the couple to make their own connections and discoveries?		
<p><i>Item 8. Formulation of Depression in Context</i></p> <p>Were there statements and/or actions designed to focus the session on relevant individual, couple, or environmental factors impacting the patient's depression, and any relationship distress experienced in the context of the depression?</p>		
<p><i>Item 9. Selection of Intervention Strategy</i></p> <p>Did the therapist select an intervention strategy that was appropriate to the stage of therapy and the couple's needs (as identified by the case conceptualisation and the way the session unfolded)?</p>		
<p><i>Item 10. Emotion-Focused Interventions</i></p> <p>Did the therapist facilitate the expression and processing of appropriate levels of emotion by the couple? Was there evidence of the therapist's attentiveness to different emotions and guidance given to the couple, either directly or indirectly, on how these emotions could be amplified or contained (depending on their adaptiveness)?</p>		
<p><i>Item 11. Cognitive Interventions</i></p> <p>Did the therapist demonstrate the ability to identify and work effectively with key cognitions and/or cognitive processes? Were any interventions implemented with adequate skill to enable the couple to make a positive shift in perspective?</p>		
<p><i>Item 12. Behavioural Interventions</i></p> <p>Did the therapist work effectively with specific target behaviours (reducing negative behaviours or promoting positive behaviours)? Were specific interventions used effectively to</p>		

promote new ways for the couple to act towards each other?		
<p><i>Item 13. Dyadic Conceptualisation</i></p> <p>Was there evidence that the therapist's statements and actions in the session were informed by an understanding of dyadic patterns of interaction? Did the therapist help the couple conceptualise their difficulties in dyadic terms, and see the dyadic interaction pattern as the problem/target, rather than blaming each other?</p>		
<p><i>Item 14. Setting Homework (for the next session)</i></p> <p>Did the therapist work effectively with the couple to agree upon a new homework task/s? Did the chosen task follow from the focus of the session and make sense to the couple? Were any potential obstacles identified and discussed to increase the likelihood of the couple being able to complete the agreed task/s?</p>		
<p><i>Item 15. Ending the Session</i></p> <p>Did the therapist conclude the session well, seeking feedback on each partner's experience of the session? If the session involved the management of difficult emotions or painful experiences, was the therapist able to ensure that the couple left the consulting room having re-established a degree of equilibrium, or discuss how the couple might do that following the session?</p>		
Total Score (sum of items 1-15)		

Any areas of strength demonstrated in the session:

.....
.....
.....
.....
.....
.....
.....
.....

Suggested areas for student learning and development, and any action points arising:

.....
.....
.....
.....
.....
.....
.....
.....

Signed (Student):

Signed (Supervisor/Assessor):

Date:

Appendix G: Information sheet study one



**Doctorate
in Clinical
Psychology**

Department of Psychology
Egham Hill
Egham
TW20 0EX
TEL: 01784 414012
EMAIL: isabelle.rudolfvonrohr.2013@live.rhul.ac.uk



PARTICIPANT INFORMATION SHEET

A review of a novel BCT competence rating scale

Principal Researcher: Isabelle Rudolf von Rohr, Supervisor: Dr. Helen Pote, Dr Michael Worrell

You are being invited to take part in a review of a Behavioural Couples Therapy competence rating scale. In order to decide whether you would like to take part, please read through the following information explaining why the study is being conducted and what your involvement would be. We are more than happy to answer any questions you may have before agreeing to participate.

Why have I been contacted?

You have been invited to take part in the study because you have been identified as having experience in making judgments about the competence of BCT therapists. We would therefore like to use your feedback previously given during a workshop on the BCTS-D for research purposes.

What is the purpose of the study?

There is a need for methods of assessing BCT competence, particularly those, which assess whether therapists can demonstrate the skills necessary to effectively deliver BCT. As you are aware we have developed a BCT competence rating scale, which can be used to provide formative and summative feedback regarding therapists' performance within observed treatment sessions. The central aim in developing this scale is to provide a tool, which is valid, reliable and usable. Expert review is an essential ingredient of efforts to improve the quality of rating scales during the developmental phase. Hence the current study aims to gain expert feedback on the new rating scale in order to examine (i) usability, (ii) face validity (inclusion of appropriate items that are a credible and plausible measure of competence) and (iii) content validity (adequate sampling of all aspects of competence specified in the scale definition) of the scale.

The study is being conducted by Isabelle Rudolf von Rohr as a part of a DClinPsych at the Department of Psychology, Royal Holloway University of London. If you agree to participate in this project, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be published with open access, meaning available to every internet user.

Scale development: BCTS-D

What does participation involve?

If you decide that you would like to take part, we will ask you (i) to complete a short demographics questionnaire and allow us to use the feedback previously given during a BCT supervisor workshop for research purposes. Filling in the demographics questionnaire will take about 5 minutes. Please note that all questionnaires will be identified by a unique ID number, rather than your name, and that your completed questionnaire will be separated from your demographic form upon receipt to ensure anonymity.

Am I required to take part?

It is entirely up to you if you wish to take part. If you do decide to take part, you are free to change your mind at any time. You can withdraw during any phase of the study, without giving a reason and without any penalty, by letting the researcher know. If this is the case, any data collected from you will no longer be included in subsequent analyses and will be destroyed.

Will my taking part in the study be kept confidential?

All information which is collected from you during the course of the research would be kept strictly confidential within the limits of the law. You will be allocated a unique number, ensuring that all materials related to your participation (e.g. completed questionnaires) will contain a unique number rather than your actual name.

In accordance with British Psychological Society research guidelines, all data for the study will be securely stored for 5 years and will be destroyed after this time. Data will be accessed only by members of the research team. However, individuals from Royal Holloway University of London and other regulatory authorities may require access to relevant data for the purpose of audit and monitoring.

What are the possible advantages of taking part?

Taking part in this study will give you a chance to directly influence the development of this scale. It is important to have a valid and reliable competency scale supporting training, the work as a supervisor and your self-reflection. Thus, the information you provide will be beneficial in helping to highlight areas where the scale requires refinement and thus will help to improve current methods of assessing BCT competence.

What are the possible disadvantages of taking part?

Given the nature of this study, it is highly unlikely that you will suffer harm by taking part. However, if the questionnaire happens to include any questions which, for whatever reason, you do not wish to answer then the question can be omitted.

What if there is a problem?

If you have a concern about any aspect of this project, please speak to the researcher concerned (contact details on page 3) who will do her best to answer your query. If you remain unhappy and wish to make a formal complaint, please contact the Health Research Authority (contact.hra@nhs.net) or call 020 710 48066).

Scale development: BCTS-D

Who has reviewed this study?

The study is in the process of being reviewed for ethics by the Health Research Authority and the Royal Holloway University of London Research Ethics Committee.

Contact Details:

If you require further information or would like to ask any questions, please do not hesitate to contact either the Principal Researcher or Supervisor using the details below.

Principal Researcher:

Isabelle Rudolf von Rohr
Trainee Clinical Psychologist
Department of Psychology
Egham Hill
Egham
TW20 0EX
Tel: 01784 414012
Email:
isabelle.rudolfvonrohr.2013@
live.rhul.ac.uk

Supervisor:

Dr. Helen Pote
Senior Lecturer & Clinical
Psychologist
Department of Psychology
Egham Hill
Egham
TW20 0EX
Tel: +44 1784 414236
Email: h.pote@rhul.ac.uk

Supervisor:

Dr. Michael Worrell
Consultant Clinical Psychologist
Psychology Department
7a Woodfield Road London
W9 2NW
Telephone: 02072669588
Mobile: 07772517750
email: michael.worrell@nhs.net

Appendix H: Informed consent form study one



**Doctorate
in Clinical
Psychology**

Department of Psychology
Egham Hill
Egham
TW20 0EX
TEL: 01784 414012
EMAIL: isabelle.rudolfvonrohr.2013@live.rhul.ac.uk



CONSENT FORM

A review of a novel BCT competence rating scale

Principal Researcher: Isabelle Rudolf von Rohr, Supervisors: Michael Worrell, Dr. Helen Pote

Study Purpose: The study aims to gain expert feedback on a new therapist competence rating scale in order to examine the usability, face validity and content validity of the scale.

Please initial box

I confirm that I have read and understand the information sheet dated 26.05.2016 for the above study. I have had the opportunity to consider the information, ask questions and have had any questions I asked answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without any penalty for doing so.

I understand that personal data will be stored and identified using only a number code, will be accessed only by members of the research team and will be destroyed after a period of 5 years.

I understand that data collected during the study may be looked at by individuals from Royal Holloway University of London and other regulatory authorities, for the purpose of research audit and monitoring, and where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.

I understand that the research will be written up as a thesis which will be deposited both in print and online in the University archives. The thesis will be published with open access, meaning available to every internet user.

I understand that the study is in the process of being reviewed for ethics by HRA and the Royal Holloway University of London Research Ethics Committee.

I understand how to raise a concern or make a complaint.

I agree to take part in the above study.

Name of Participant

Date

Signature

Researcher

Date

Signature

Appendix I: Demographics information sheet

Participant ID _____

Participant Demographics

1. Age: _____

2. Gender: Male Female

3. Profession: _____

4. Current job role: _____

5. Do you have any other BCT training? No Yes

If yes, please specify _____

6. Do you have any formal CBT training? No Yes

If yes, please specify _____

7. If no, please specify any other BCT / CBT training you may have attended (e.g. workshops)?

8. How many years have you been practicing BCT? _____ years

9. How many years have you been practicing CBT? _____ years

10. What would you estimate your total BCT therapy experience as being? (please circle)

< 50 cases 50 – 200 cases 200 + cases

Scale development: BCTS-D

11. What would you estimate your total CBT therapy experience as being? (please circle)

< 50 cases 50 – 200 cases 200 + cases

12. Are you BABCP accredited:

As a CBT therapist? No Yes

As a CBT supervisor? No Yes

As a CBT trainer? No Yes

13. Do / have you provided CBT supervision? No Yes

14. Do / have you provided CBT training? No Yes

15. Have you got any experience using the Cognitive Therapy Scale - Revised (CTS-R)

No Yes

16. How often do you use the CTS-R within your current practice?

> once a Week weekly Monthly less than once a month

Appendix J: Short feedback questionnaire

Participant ID number

BCTS-D

Q1: Are there any important aspects of BCT competence which you feel are missing from the scale (i.e. any key competences which the scale neglects)?

Please Circle **Yes / No**

If you circled yes, which aspects of BCT competence do you feel are missing?

Q2: Do you feel the scale provides ample opportunity to provide both in-depth summative and formative feedback?

Please Circle **Yes / No** (If you circled no, what did you feel was missing?)

Q3: Please rate the overall style, appearance and layout of the scale.

Please Circle

1 poor **2** fair **3** good **4** very good

If you circled a three or below, what revisions do you feel could improve the overall style, appearance and layout of the scale?

Scale development: BCTS-D

Participant ID number

Q4: How easy do you think the rating scale is to use?

Please Circle

1 not easy **2** somewhat easy **3** quite easy **4** very easy

If you circled a three or below, what revisions do you feel could improve the usability of the scale?

Q5: Do you have any other comments or feedback about the rating scale or any other suggestions of ways the rating scale could be improved?

Scale development: BCTS-D

Appendix K: BCTS-D v2 manual

BEHAVIOURAL COUPLE THERAPY SCALE – DEPRESSION (BCTS-D)

THE MANUAL

Sarah Corrie, Melanie S. Fischer, Michael Worrell, & Donald H. Baucom

Scale development: BCTS-D

Acknowledgements

The authors of this scale wish to thank the originators of the Cognitive Therapy Scale-Revised, and accompanying manual, for permission to adapt the CTS-R.

Scale development: BCTS-D

Contents Page

Part 1. An Introduction and Orientation to the BCTS-D

Purpose of the manual

Aims and objectives of the BCTS-D

Background to the scale and its development

Psychometric properties

Domains and items

Scoring and anchor points

Recommendations for how to use the BCTS-D effectively

Part 2. The Items of the BCTS-D

Item 1. Agenda Setting

Item 2. Review of Homework (from the previous session)

Item 3. Collaboration

Item 4. Facilitating Couple Communication

Item 5. Pacing and Flow

Item 6. Therapist's Interpersonal Effectiveness

Item 7. Guided Discovery

Item 8. Formulation of Depression in Context

Item 9. Selecting Interventions

Item 10. Emotion-Focused Interventions

Item 11. Cognitive Interventions

Item 12. Behavioural Interventions

Item 13. Dyadic Conceptualisation

Item 14. Setting Homework

Item 15. Ending the Session

Part 3. Additional information

References

Contact details for the scale developers

Part 1: An Introduction and Orientation to the BCTS-D

Scale development: BCTS-D

Purpose of the Manual

The Behavioural Couple Therapy Scale – Depression (BCTS-D) is a new training and supervision tool developed for the purposes of assessing and facilitating competence in behavioural couple therapy at both formative and summative levels. The purpose of this manual is to provide sufficient background information to enable assessors to understand and use the scale effectively, to deliver more constructive and helpful feedback, and to aid therapists in better understanding the feedback they are given. A further aim is to support behavioural couple therapists in becoming better able to assess their own performance – both within and across sessions.

The manual is intended to be a companion to the BCTS-D. Although the descriptive features included in the BCTS-D scale are designed to guide decision-making about the extent of a therapist's competence on any given item, when using the scale, assessors should *score from the manual*. This is because the manual provides detailed information and guidance on how to rate each item, as well as examples of how the specific competence might present in a session when conducted effectively.

The manual describes the aims and objectives of the scale, the competences and skills that the BCTS-D aims to assess, and offers guidance on how to use the scale. As the scale is in the process of being developed and tested, information on its psychometric properties is currently limited. Further data relating to reliability and validity will be provided in due course and in the meantime, we welcome comments and suggestions on the scale's refinement. (Contact details for the developers are provided at the end of the manual.)

Aims and objectives of the BCTS-D

The BCTS-D aims to assess – both quantitatively and qualitatively – the competences and skills that are deemed to be central to the effective delivery of behavioural couple therapy (BCT). Comprising 15 items, each of which focuses on a specific BCT competence, the scale is designed to assess treatment sessions where these competences might reasonably be expected to be demonstrated if a therapist is practising proficiently. The scale can be used to rate performance on couple sessions that have been either audio- or video-recorded as well as live supervision.

The BCTS-D takes account of the fact BCT is an efficacious intervention for treating relationship distress, depression and a variety of other clinical presentations. Indeed, there is more research evidence demonstrating the efficacy of BCT for alleviating relationship distress than any other approach (see for example, Barbato & D'Avanzo, 2008; Baucom, Whisman & Paprocki, 2012; Epstein & Baucom, 2002; Whisman & Baucom, 2012). The scale can, therefore, be used to assist the evaluation of sessions where depression is a feature, and to assess competence where:

- One partner is experiencing psychopathology and the therapist is using a partner-assisted intervention;
- Both partners are experiencing psychopathology;
- Relationship distress is present in the absence of significant levels of couple psychopathology.

In assessing therapist proficiency, the BCTS-D aims to help assessors identify areas of therapist strength, and areas where further development is desirable through enabling detailed feedback on specific aspects of the therapist's performance in a specific session.

Scale development: BCTS-D

The scale is optimally suited to assessing performance on mid-treatment sessions where the therapist is *demonstrating the use of specific BCT interventions*. It is not intended for summative use during the pre-treatment phase of BCT (i.e., where the therapist conducts joint couple assessments and individual interviews, or during feedback and treatment planning sessions). Equally, the scale is not ideally suited for sessions (1) focusing *almost exclusively* on psychoeducation; (2) a final session where the focus is on ending the therapy or (3) review and follow-up sessions. Although it is possible to use the BCTS-D to provide feedback in these instances, this form of application should be for formative evaluations only. Where a session is being rated summatively, a mid-treatment session is necessary as this enables therapists to demonstrate most easily the delivery of the specific BCT interventions that the BCTS-D has been designed to assess.

It is important to note that the BCTS-D has been designed to capture evidence of specific competences and therapeutic skills as demonstrated *in a specific session only*. The scale does not enable an assessor to form, with any accuracy, a more global picture of the therapist's competence as a BCT therapist.

It is also important to note that the scale does not assess knowledge of BCT theory and technique, or conceptual understanding of the different types of couple presentation. Assessment of this more 'declarative' form of knowledge is best achieved through alternative measures such as multiple choice questionnaires and essays (see Muse & McManus, 2013, for a helpful review of the different methods for assessing competence in CBT). In order to develop a fuller, more global picture of a therapist's knowledge and skill in BCT, a multimethod approach is recommended.

Background to the scale and its development

The items for inclusion in the BCTS-D were identified through a review of the current literature on BCT and current competence frameworks for both CBT and couples therapy for depression. The development of the scale has also been informed by good practice guidelines on the assessment of competence in CBT practice (e.g. BABCP Minimum Training Standards, 2012; Muse & McManus 2013) and existing scales and frameworks including:

- The Cognitive Therapy Scale - Revised (Blackburn et al., 2001);
- The Couple Therapy for Depression Competence Framework (available at: https://www.ucl.ac.uk/pals/research/cehp/research-groups/core/competence-frameworks/Couples_Therapy_For_Depression);
- Original work on enhanced cognitive-behavioural therapy for couples by Epstein and Baucom (e.g., 2002).

The theoretical basis of the scale is derived from two principal sources: the enhanced cognitive-behavioural therapy model developed by Epstein and Baucom (2002) and the Dreyfus and Dreyfus (1986) scale of competence (see section 'Scoring and anchor points', below, for a description of how the Dreyfus competence scale has been applied to the BCTS-D).

Psychometric properties

The scale is currently undergoing development and data on the reliability and validity of the scale will be made available in due course. For current information, please contact the scale developers (contact details are provided at the end of this manual).

Scale development: BCTS-D

Domains and items

The BCTS-D comprises 15 items, each of which represents a specific, identified competence in BCT. These 15 items are grouped into three broad domains as shown in Table 1:

Table 1. The Three Domains Underpinning the BCTS-D

Domain 1. Structure of the session	Domain 2. Interaction with the couple and management of the therapeutic process	Domain 3. Interventions selected and employed
1. Agenda Setting	3. Collaboration	8. Formulation of Depression in Context
2. Review of Homework (from the previous session)	4. Facilitating Couple Communication	9. Selection of Intervention Strategy
14. Setting Homework (for the next session)	5. Pacing and Flow	10. Emotion-Focused Interventions (the competent conduct of)
15. Ending the Session	6. Therapist's Interpersonal Effectiveness	11. Cognitive Interventions (the competent conduct of)
	7. Guided Discovery	12. Behavioural Interventions (the competent conduct of)
	13. Dyadic Conceptualisation	

Scale development: BCTS-D

Each of the individual items, along with a brief description, is provided in Table 2, below:

Table 2. The 15 Items of the BCTS-D and their Ordering

Item 1	Agenda Setting	Did the therapist establish a focus for the session? Was a clear agenda set, with priorities identified that provided a clear 'steer' for the session that was to follow?
Item 2	Review of Homework (from the previous session)	Did the therapist review the homework agreed upon at the previous session? Were new insights or learning elicited? If needed, were any obstacles identified and explored to enable 'follow through' with future homework tasks?
Item 3	Collaboration	Did the therapist facilitate effective teamwork, promoting an adaptive working relationship between the therapist and couple, in which appropriate levels of transparency and mutual feedback were encouraged? Were statements and/or actions made that encouraged both members of the couple to participate appropriately, and which prevented an unequal power relationship from developing between the couple and the therapist?
Item 4	Facilitating Couple Communication	Did the therapist facilitate couple communication effectively? Was there evidence of enabling understanding and dialogue between partners through steering the conversation towards reciprocal listening, 'reflecting back' and enhanced empathy? Was the therapist able to provide sufficient structure and management to enable the couple to communicate more effectively without overly-controlling the process?
Item 5	Pacing and Flow	Did the therapist pace the session well, enabling a smooth transition and sense of flow between different phases of the session, ensuring that priority items on the agenda were addressed? Was the therapist able to control the timing, pacing and flow of the session with two partners present?
Item 6	Therapist's Interpersonal Effectiveness	Did the therapist create an interpersonal environment that was conducive to the work of the session? Did the therapist form an effective alliance with each partner individually and the couple as a unit? Was a safe environment created for the couple to address difficult issues in session? Did the therapist's interpersonal style convey warmth, interest, and instil hope that change is possible, as well as convey appropriate authority in the way the session was managed?

Scale development: BCTS-D

Item 7	Guided Discovery	Did the therapist facilitate a process of discovery that conveyed genuine interest, curiosity, and understanding? Were there questions and/or actions designed to promote self-reflection, helping the couple to make their own connections and discoveries?
Item 8	Formulation of depression in Context	Were there statements and/or actions designed to focus the session on relevant individual, couple, or environmental factors impacting the patient's depression, and any relationship distress experienced in the context of the depression?
Item 9	Selection of Intervention Strategy	Did the therapist select an intervention strategy that was appropriate to the stage of therapy and the couple's needs (as identified by the case conceptualisation and the way the session unfolded)?
Item 10	Emotion-Focused Interventions	Did the therapist facilitate the expression and processing of appropriate levels of emotion by the couple? Was there evidence of the therapist's attentiveness to different emotions and guidance given to the couple, either directly or indirectly, on how these emotions could be amplified or contained (depending on their adaptiveness)?
Item 11	Cognitive Interventions	Did the therapist demonstrate the ability to identify and work effectively with key cognitions and/or cognitive processes? Were any interventions implemented with adequate skill to enable the couple to make a positive shift in perspective?
Item 12	Behavioural Interventions	Did the therapist work effectively with specific target behaviours (reducing negative behaviours or promoting positive behaviours)? Were specific interventions used effectively to promote new ways for the couple to act towards each other?
Item 13	Dyadic Conceptualisation	Was there evidence that the therapist's statements and actions in the session were informed by an understanding of dyadic patterns of interaction? Did the therapist help the couple conceptualise their difficulties in dyadic terms, and see the dyadic interaction pattern as the problem/target, rather than blaming each other?
Item 14	Setting Homework (for the next session)	Did the therapist work effectively with the couple to agree upon a new homework task/s? Did the chosen task follow from the focus of the session and make sense to the couple? Were any potential obstacles identified and discussed to increase the likelihood of the couple being able to complete the agreed task/s?

Scale development: BCTS-D

Item 15	Ending the Session	Did the therapist conclude the session well, seeking feedback on each partner's experience of the session? If the session involved the management of difficult emotions or painful experiences, was the therapist able to ensure that the couple left the consulting room having re-established a degree of equilibrium, or discuss how the couple might do that following the session?
---------	--------------------	---

The ordering of the 15 items follows, where possible, the order in which they are likely to occur in the session thus providing a logical sequence to the scale items. For example, the early items (Agenda Setting, Review of Homework) reflect the fact that these items are addressed at the start of the session. Conversely, Homework Setting and Ending the Session are placed at the end of the scale to reflect that they occur at the end of the session. Other items (grouped under Domains 2 and 3) may appear at different and multiple points during the session, and their position in the scale does not denote a sequential significance.

A fuller description of each of these individual items, along with features of competent performance and what the assessor might anticipate seeing and hearing if this competence is being demonstrated, is provided in Part 2 of this manual.

Scoring and anchor points

The scale produces both numerical ratings and qualitative data. Every item on the BCTS-D should be given a numerical score, and any item scored 2 or below (i.e., that has failed to reach the passing grade) should also be accompanied by qualitative feedback to aid therapist learning and development.

In order to score the BCTS-D, the 15 items are summed. Each individual item is rated from 0-6 and all the items are weighted equally. Thus, possible scores range from 0 (i.e., 15 x 0) to 90 (15 x 6). The 'pass mark' is a score of 45 (which equates to 50%).

Whole numbers only should be used to reflect the extent to which the therapist has fulfilled the key features on each item. Scores of 5 and 6 would be deemed to reflect an 'expert' level of skill and are reserved for those therapists demonstrating exceptional levels of proficiency. As such, it is anticipated that these scores will be relatively rare.

Scores are expected to follow a normal distribution with the majority of scores being at the mid-point of 3 which is a passing 'grade', with relatively few therapists scoring at the extremes across most items. It is important that therapists have an awareness of these scoring norms, so that therapists who are doing a good job expect scores generally in the 3 range rather than mostly 4-6.

Where a therapist's performance on a particular item falls mid-way between two scores without clear evidence that would suggest one score over the other, the general guideline is that the rating which is closest to the mid-point of the scale should be assigned. For example, if a therapist's performance falls between 2 and 3, a score of 3 should be awarded. If a therapist's performance falls between 4 and 5, a score of 4 should be assigned, and so on.

Scale development: BCTS-D

For summative assessments, a pass mark is determined based on the required total score only. Therapists do not need to achieve a minimum score on every item in order to achieve an overall pass.

Where a therapist appropriately omits content in one domain (e.g., where the therapist does not, for good reason, use any behavioural interventions in a session but prioritises instead the use of cognitive interventions), the therapist should not be penalised and would be awarded a default rating of 3 in the omitted area. It is acceptable to omit a particular domain providing that the assessor is confident that this is based on sound decision-making on the part of the therapist.

Numerical scores should be accompanied by qualitative feedback that informs the therapist of areas of strength and development detected in the recording. Although qualitative feedback is not essential on every single item, it is important that it is provided on any items where a therapist scored 2 or below (i.e. where they failed to show competence). Where possible, this should be supported by specific examples from the recording as this detailed 'pinpointing' is greatly appreciated by therapists.

The BCTS-D has been informed by the Dreyfus and Dreyfus (1986) five level framework of competence, adding in a sixth level to denote the absence of competence:

Table 3: Adapted Dreyfus Competence Scale

Incompetent	The therapist commits errors and displays poor and unacceptable behaviour, leading to negative therapeutic consequences.
Novice	At this level the therapist displays a rigid adherence to taught rules and is unable to take account of situational factors. He/she is not yet showing any discretionary judgement.
Advanced beginner	The therapist treats all aspects of the task separately and gives equal importance to them. There is emerging evidence of being able to adopt a situational perspective and discretionary judgement.
Competent	The therapist is able to see the tasks linked within a conceptual framework. There is an ability to make plans within this framework and use standardised and routinised procedures.
Proficient	The therapist sees the patient's problems holistically, prioritises tasks and is able to make quick decisions. The therapist is clearly skilled and able.
Expert	The therapist no longer uses rules, guidelines or maxims. There is a deep, sophisticated and tacit understanding of the key issues, and the therapist is able to use novel techniques as needed. These skills are demonstrated even in the face of difficulties (e.g., excessive avoidance).

The Dreyfus and Dreyfus competence scale has been adapted for the purposes of the BCTS-D as indicated below, in Table 4 (NB: in Part 2 of this Manual, more detailed information is provided on the anchor points specific to that item. However, all items employ the following generic guidelines):

Table 4. The Dreyfus Competence Scale adapted for the BCTS-D

0	There were significant concerns about the therapist’s performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision. This would be viewed as extremely inadequate therapy on this item as if the therapist had clearly lost track of what was to be accomplished in this regard.
1	The therapist did not display competence on this item and needs to attend to key areas as identified in the assessor’s feedback. Something relative to the item was present, but notably lacking. Significant improvement is needed.
2	The therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor’s feedback. A score of 2 usually signifies that the session was generally well conducted on this item, but something notable was missing, or it went well for most of the session with a deviation at some point, such that it was not fully acceptable, just below passing.
3	The therapist demonstrated competence on this item although refinement (as identified by the assessor) would be beneficial. This is the score that will be given to most therapists on most items if they are doing a “good job.” This score is passing.
4	The therapist’s performance was competent on this item with only minor areas of refinement needed (as specified in the qualitative feedback). There were no concerns about the therapist’s performance. A score of 4 is a “good job plus.” Generally, the session was well conducted and some aspect stood out throughout the session or at some point during the session as above the norm.
5	The therapist demonstrated an obvious, high degree of skill on this item. A score of 5 reflects that this item was very well conducted. Nothing inappropriate was done, and nothing important was omitted. The therapist was simply not at the exemplary level of mastery which is required for a score of 6.
6	The therapist demonstrated a superior level of skill on this item, and would be deemed to be working at ‘master level’. A score of 6 is rare and signifies exceptional performance on this item. One might select this for a training session to demonstrate highest levels of performance.

Recommendations for how to use the BCTS-D effectively

In order to assess a BCT session effectively, it is important to have some background information on the couple and the nature of their concerns, as well as the therapist’s aims for the session. This helps contextualise the work. When evaluating a session, for example, it is important to take into consideration the appropriateness of the intervention for the stage of

Scale development: BCTS-D

therapy and the couple's difficulties. For summative assessments in particular, BCT therapists are expected, therefore, to provide accompanying documentation which will usually include some demographic information about the couple (e.g., brief background information, current circumstances), a brief case conceptualisation of any relevant background individual, couple or environmental factors that were informing their work with the couple, and a statement about the aims of the session and therapist's self-assessment of what went well and less well in the session.

Although it may seem unnecessarily prescriptive to recommend a specific procedure for rating sessions, it is important to develop a systematic approach that can enable a consistency of attention to the quality of a therapist's performance. The literature on clinical decision-making highlights a number of cognitive biases which can inadvertently influence assessors' evaluations (see Gambrill, 2005, for a review of this literature).

To avoid bias, it is prudent to avoid misguided decision-making tendencies such as scoring on the basis of a global impression of the session, scoring an item based on the ratings already given to other items, the overall impression formed of a particular therapist, the perceived 'likeability' of the therapist, or the outcome of a session (e.g., a therapist might have conducted a very good session with a difficult couple who leaves the session upset with each other). It is important to remain aware of these confounding factors during the course of rating sessions and for even experienced BCTS-D assessors to audit their approach to ensure consistency and accuracy.

We recommend, therefore, that when rating sessions assessors adopt the following procedure:

- Step 1: Listen to the entire recording, noting down specific examples of strength and areas for development as the session progresses;
- Step 2: Compile qualitative feedback;
- Step 3: Assign numerical ratings to individual items based on a review of the qualitative feedback, and the specific examples identified. A useful way to approach this is to establish whether some of the features of the item under consideration are present. Then consider whether the therapist met competence criteria (i.e., a score of 3 or more). If the therapist includes most of the key features and uses them appropriately (i.e., misses few relevant opportunities to use them), the therapist should be rated as competent on the particular item being considered.
- Step 4: Sum the scores for all 15 items to arrive at a total score.

As noted previously, the BCTS-D can be used formatively and summatively. As a formative tool, it is recommended that therapists use the measure to self-assess their work, as an aid to reflection and self-supervision, and for discussion with their BCT supervisor. The measure can be used by those undertaking BCT training, as well as experienced BCT therapists who seek to refine their approach and ensure on-going model fidelity.

We strongly recommend that the measure should only be used for summative assessments by those who have themselves completed a substantive training in BCT and who have been trained in the use of BCTS-D.

Scale development: BCTS-D

In Part 2, the manual examines each of the individual items in more detail. This is achieved through providing an introduction to the item, describing the typical features of competent performance, providing scoring and anchor points, and offering some 'primers for the assessor' (including questions, statements or interventions on the part of the therapist that might be considered examples of the competence or skill 'in action').

Part 2. The 15 Items of the BCTS-D

Item 1. Agenda Setting

Introduction

Beginning the session effectively requires that the therapist works with the couple to establish a focus for their time together. It is part of the therapist's role to ensure that the couple's main concerns and/or priorities are identified at the outset and built into a plan for the session. Additionally, the therapist may have particular items that he or she believes needs to be prioritised (e.g., teaching the couple a particular skill) and where this is the case, this also needs to be identified at the beginning of the session. Therefore, the agenda should take into account both immediate concerns and issues, as well as the overall treatment plan.

The way in which the session plan is established should be personalised to take account of the couple's interpersonal style and their particular needs, given the stage of therapy. At times, this might require greater direction on the part of the therapist, at times less so. The agenda can also be agreed very swiftly or may take longer, if the couple needs helping in identifying priorities say, for example, in the context of having had a bad week. What matters is that a focus is identified for the session in a way that facilitates engagement of the couple and provides clarity for what is to follow.

At the start of the session, the therapist might also elicit feedback on the previous therapy session to identify any themes, insights, or misunderstandings that occurred and which might also need addressing in the current session. Seeking feedback on the previous session is rated under Item 1 as the couple's reflections may need to be included on the agenda.

(NB: It is neither necessary nor desirable for the therapist to have a rigid structure, and there is no expectation that specific time slots are allocated to specific agenda items.)

Features of competent performance

Competent performance on agenda setting is evidenced through the following features:

1. The presence of a session plan

The therapist began the session by establishing priority areas (i.e., time and attention was given to actually establishing a session plan).

2. The identification of specific content areas or patterns of interaction

The therapist worked with the couple to identify particular areas that needed to be prioritised. These areas might have revolved around specific content domains in the relationship (e.g., handling finances) or emphasized an interaction pattern (e.g., demand-withdraw pattern). Where the couple needed assistance with identifying priorities (for example, if the couple's thinking was dominated by high levels of negative affect following a challenging week), the therapist worked with the couple to identify specific areas that can be addressed. Alternatively, the therapist might have proposed a focus based on the overall treatment plan.

3. Appropriateness of the content and interaction areas

The aims of the session plan and content areas and interaction patterns specified were appropriate to the couple's concerns and stage in therapy, and were feasible in the time available.

4. Working in partnership with the couple

Scale development: BCTS-D

The therapist worked to ensure that the couple had an opportunity to share their hopes for, or needs of, the session and took their priorities into account in the session plan. The therapist conveyed genuine interest in the couple's priorities and approached the session with sufficient flexibility to allow these to be appropriately accommodated (that is, the therapist was not so attached to their own agenda that the couple's concerns were overlooked).

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist launched into the session without any attempt to establish a plan for the session and without any attempt to identify any priorities for the session.</p> <p>OR: The therapist demonstrated a total lack of interest in the couple's main concerns and 'told' the couple what the focus of the session was to be, creating an alienating, over-controlling climate at the outset.</p> <p>In summary: There were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to establish a focus for the session but did so ineffectively and as such, their performance was highly limited. For example:</p> <p>The therapist may have 'checked in' with the couple about their well-being but did so in a perfunctory manner.</p> <p>OR: The therapist unilaterally imposed a session plan, even if done politely.</p> <p>OR: The therapist did not attempt to identify core areas of concern for the couple that could have usefully been a focus of the session.</p> <p>OR: the areas identified seemed highly inappropriate given the nature of the couple's difficulties, needs, treatment goals or stage of therapy.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback.</p>
2	<p>The therapist began by establishing a plan for the session, but this was done in an inconsistent or clumsy fashion. The therapist's actions showed evidence of emerging competence on this item, and an awareness of the importance of developing a shared agenda, but the delivery indicated a lack of basic competence that could have, or actually did, negatively impact the flow of the session. For example:</p> <p>The therapist was overly rigid in their approach to beginning the session (e.g., attempting to allocate rigid time slots to particular items).</p> <p>OR: The therapist seemed so preoccupied with establishing a plan for the session</p>

Scale development: BCTS-D

	<p>that there was insufficient attentiveness to the couple's comments.</p> <p>OR: Too many items were identified for the time available, and the therapist struggled to know how to prioritise amongst these.</p> <p>OR: the session plan was too unclear or insufficiently specified to provide an appropriate steer for use of the time.</p> <p>In summary, the therapist demonstrated an emerging competence on this item, but further work is needed to reach competence, with key areas for development identified in the assessor's feedback.</p>
3	<p>Overall the therapist began the session effectively and set a good, clear and realistic agenda that had the potential to provide a helpful 'steer' for the session. The therapist worked with the couple to create a suitable session plan and appropriate collaboration was evident.</p> <p>The therapist's performance was broadly competent despite some minor inconsistencies or areas of refinement needed.. At times, the process may have lacked a degree of fluidity but the essential components were present.</p> <p>In summary, the therapist demonstrated competence on this item although refinement (as identified by the assessor) would be beneficial.</p>
4	<p>The therapist began the session effectively, obtained feedback on the previous session if appropriate to do so (including adding any feedback to the agenda if necessary) and worked to create a suitable plan that accommodated both the couple's and the therapist's priorities.</p> <p>The therapist's performance was proficient on this item with only minor areas of refinement needed. The way the session began appeared to be flexible and fluid whilst also being underpinned by a clear structure that guided the therapist's thinking and actions.</p> <p>In summary, the therapist demonstrated an effective use of the agenda setting process. There were no concerns about the therapist's performance on agenda setting and the therapist demonstrated competence.</p>
5	<p>The therapist began the session very effectively. Whilst providing a welcoming atmosphere, a clear session plan was established and specific content areas were identified. These were appropriate. There was effective collaboration with the couple to arrive at a session plan. At the same time, the therapist demonstrated their own approach and style to beginning the session that contributed a positive interpersonal aspect to the process.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. The therapist's actions were entirely appropriate and nothing important was omitted. The performance was simply not at the exemplary level of master which is required for a score of 6.</p>

Scale development: BCTS-D

6	<p>The therapist demonstrated an exceptional level of proficiency in how they began the session and helped the couple to identify session priorities. In consequence, an appropriate session plan was created easily and efficiently in partnership with the couple.</p> <p>The therapist established a focus for the session with confidence, ease and skill, displaying a personal approach that seemed congruent with their own style of practice whilst also accommodating the couple's interpersonal style, any relevant individual characteristics (specified in the case conceptualisation), and the challenges confronting the couple (for example, adjusting how the session began in the context of the couple reporting a very difficult week, bad news or where the therapist is confronted with negative reciprocity or high levels of hostility from the outset of the session).</p> <p>In summary, the therapist demonstrated an obvious, exceptionally high degree of skill on this item, and would be deemed to be working at 'master level'. The therapist's actions were entirely appropriate and nothing important was omitted. A score of 6 on this item is likely to be rare and indicates a level of skill that might be used to demonstrate the highest levels of performance in a training session.</p>
---	--

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of effective performance:

1. Did the therapist convey an air of confidence that establishing a plan for the session would be useful in using the time to best advantage?
2. Did the therapist seek the couple's feedback on their experience of the previous session (and if necessary were any matters arising put on the agenda)?
3. Was the therapist able to sift through the couple's brief initial updates on well-being and couple functioning to identify the most important themes to discuss?
4. Was each partner encouraged to participate in the process of establishing a focus for the session? (For example, did the therapist invite the couple to share their priorities for the session?)
5. Was the therapist appropriately transparent in identifying items that he/she thought would be important to discuss in the session?
6. Where necessary, did the therapist convey skill in negotiating priorities?
7. Was there a sense of effective team work and genuine collaboration in how the therapist worked with the couple to ensure a good basis for proceeding?

Examples of statements that reflect these indicators include the following:

- Before we start, let me get your feedback on our previous session (What was most useful? Anything that was confusing or unhelpful?).
- Let's think about how we want to use our time together today.
- Let's agree a plan for the session.
- What do you think would be most useful for us to talk about?

Scale development: BCTS-D

- There's a lot in what you are saying. If we could only cover one of those issues today, which one would you choose? What's most important to you?
- If we think about the goals of our work together, which one or two of those points would be most useful to discuss?
- Where should we start?
- (And if the therapist is unsure of the rationale for the couple or one partner wishing to prioritise a specific item:) How would discussing this topic help us make progress here today/support you in achieving your goals for therapy?

(NB: the above are offered as suggestions of potential illustrations of competence only.)

Item 2. Review of Homework (from the previous session)

Introduction

BCT is an active therapy that brings with it the expectation that couples will engage in homework. These homework tasks help couples transfer the learning that takes place in therapy to daily life and, as such, bridge the gap between the sessions and their everyday lives.

The ability and willingness of the couple to engage in homework is supported by a therapist who can work with them to identify an appropriate task/s, enable the couple to appreciate the potential benefits of engaging in this task/s, and anticipate and problem-solve any potential obstacles. (The therapist's proficiency in homework setting is scored under Item 14.)

A further aspect of therapist skill is ensuring that agreed homework tasks are reviewed in the following session so that insights can be identified, new learning made explicit, and any obstacles that prevented completion addressed. It is anticipated, therefore, that early on in the session, the therapist will review the couple's attempts to complete the home practice tasks agreed during the previous meeting.

NB: In some cases, reviewing the homework may be completed relatively swiftly (for example, where the agreed task is part of an intervention plan that is well-underway and where the couple is increasingly taking ownership of their therapy and achieving consistent progress). At other times, a more detailed review may be required (for example, if completion was derailed by conflict or if additional, therapist-assisted problem-solving is needed to ensure successful completion). What is being rated here is the therapist's ability to review the homework in a way that is consistent with the stage of therapy and the couple's idiosyncratic needs.

Features of competent performance

Competent performance for Item 2 comprises three main aspects evidenced through the following features:

1. *The presence/absence of a homework task in which it was clear that a specific home practice task/s was agreed with the couple at the previous session*

The therapist took the time to review with the couple the task that had been agreed, clarifying what it was that the couple attempted (however tightly or loosely this was defined) and their experience of doing so.

2. *The therapist worked with the couple to elicit any insights or new learning obtained from engaging in the homework*

The therapist helped the couple reflect upon their experiences of the homework, any learning obtained and how completing the homework linked to the couple's aims for therapy.

3. *The therapist enquired about any obstacles or challenges to completion of the homework*

The therapist was attentive to, and probed for, any challenges encountered and either problem-solved these at the start of the session (if appropriate to do so) or agreed with the

Scale development: BCTS-D

couple to put these on the agenda for further discussion, to enable 'follow through' with future homework tasks.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist made no reference to the homework from the previous session.</p> <p>OR: The therapist demonstrated a total lack of interest in the couple's efforts to complete homework to the extent that his/her actions came across as undermining of the couple's efforts to follow through on home practice tasks.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to review the homework but did so ineffectively such that their performance was highly limited. For example:</p> <p>The therapist may have 'checked in' with the couple about their homework but did so in a perfunctory manner.</p> <p>OR: The therapist made no attempt to elicit any learning that had emerged from the couple's engagement in the task.</p> <p>OR: The therapist seemed irritated by any difficulties that the couple encountered with following through on the homework.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback.</p>
2	<p>The therapist reviewed the homework and attempted to elicit new learning but this was done in an inconsistent or clumsy fashion. The therapist's actions showed evidence of emerging competence in this item, and an awareness of the importance of reviewing homework, but the delivery indicated a lack of basic competence that could potentially (or actually did) negatively impact the flow of the session. For example:</p> <p>The therapist seemed overly concerned with the couple engaging in the task 'correctly' at the expense of helping them consider what had been gained through engagement with the task.</p> <p>OR: The therapist seemed confused by, and unable to decide how to respond to, any difficulties that the couple encountered with task completion.</p>

Scale development: BCTS-D

	<p>OR: The therapist made no attempt to reinforce the efforts made by the couple to follow through on the agreed tasks.</p> <p>OR: the therapist failed to give adequate attention to the couple's difficulties in following through on homework (e.g. colluding with rather than probing and gently challenging the couple's stated reasons for not following through on agreed tasks).</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor's feedback.</p>
3	<p>Overall the therapist reviewed the homework effectively. At times, the process may have lacked a degree of fluidity, but the essential components were present, with the three features of competent performance (listed above) evident. For example:</p> <p>The therapist took time to review the specific home practice task/s that were agreed at the previous session and to explore what was, and was not attempted.</p> <p>AND: The therapist attempted to elicit any insights or new learning that occurred for the couple (although these may not necessarily have been explored fully or linked back to the tasks of therapy).</p> <p>AND: The therapist enquired about any obstacles or challenges encountered (although these may have been addressed in sufficient depth).</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist reviewed the homework effectively. For example:</p> <p>The couple's efforts to follow through on agreed tasks were reinforced and gains were 'celebrated' as instances of new learning.</p> <p>AND/OR: The therapist worked to elicit new learning and encouraged the couple to consider how insights and learning could be generalized to other situations.</p> <p>AND: Any obstacles were greeted by the therapist with curiosity, and identified explicitly as opportunities for the couple and the therapist to learn more.</p> <p>In summary, the therapist's performance was competent on this item with only minor areas of refinement needed. There were no concerns about the therapist's performance on this item with the therapist demonstrating a good, solid level of proficiency. Some aspect of homework review stood out throughout the session or at some point during the session as above the norm.</p>
5	<p>The therapist reviewed the homework very effectively. The couple's efforts were reinforced and gains were 'celebrated' as instances of new learning. The therapist conveyed interest and curiosity in what was achieved and worked highly effectively to elicit new learning, and to help the couple link this back to their therapeutic goals and</p>

Scale development: BCTS-D

	<p>the aims of therapy. 'Next steps' to consolidate their learning and behaviour change may also have been considered.</p> <p>The therapist conveyed curiosity and confidence in identifying, formulating and addressing obstacles to completion.</p> <p>In summary, the therapist demonstrated an obvious, very high level of skill on this item. Nothing inappropriate was done, and nothing important was omitted. The proficiency demonstrated was simply not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist's approach to reviewing the homework was highly skilled, demonstrating excellence at the level of reviewing what was completed, eliciting new learning, guiding the couple towards insights they may have overlooked and instilling enthusiasm in the couple for what they had achieved, if appropriate.</p> <p>With considerable expertise, the therapist not only worked to elicit new learning, but also helped the couple link this back to their therapeutic goals and the aims of therapy. The therapist also guided the couple towards considering appropriate 'next steps' to consolidate their learning and behaviour change.</p> <p>Where the couple encountered obstacles, the therapist conveyed appropriate empathy combined with problem-solving. The therapist was sensitive to, and able to elicit, information relating to tacit concerns or process issues that prevented completion, enabling new insights as to the couple's needs.</p> <p>In summary, the therapist demonstrated a superior level of skill on this item, and would be deemed to be working at 'master level'.</p>

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of effective performance:

1. Did the therapist ask the couple to briefly summarise the homework that had been agreed at the previous session?
2. Did the therapist ask the couple to briefly outline/remind each other and the therapist of the rationale for completing the homework ?
3. From their response, could it be concluded that the couple appreciated the relevance of the task that had been agreed?
4. Did the therapist ask about the couple's efforts to complete the task?
5. Did the therapist work with the couple to identify any new learning that emerged through engagement with the homework?
6. Were any obstacles that prevented completion identified and reflected upon in a thoughtful, non-judgemental manner that enabled new insights to emerge?
7. Was the therapist attentive to any concerns or objections that the couple might have had about the task?

Scale development: BCTS-D

Examples of statements that reflect these indicators include the following:

- Would one of you like to remind me of what it was we agreed you'd work on this week?
- To recap, what was the rationale for completing this task – why did we think this would be useful?
- How did you get on with the task?
- What do you think you learned/discovered from following through on this? (Did you learn anything new?)
- How does this task relate to the bigger picture of what you want to achieve through our work together, do you think?
- Did anything get in the way of your being able to follow through on this? (If so, what happened?)
- This sounds important and we'd want to make sure that you don't have similar problems in the future. How about we put this on the agenda for today's session so we can do some more detailed problem-solving?

(NB: the above are offered as suggestions of potential illustrations of competence only.)

Item 3. Collaboration

Introduction

Effective BCT relies on productive teamwork, in which all parties are expected to be active participants. Item 3 is concerned with the working alliance that is created between the therapist and the couple, but only in terms of the task-focused aspects of the session (NB: the quality of the therapeutic relationship that is developed with the couple in terms of the 'core conditions' (through the therapist's qualities in the session such as warmth, validation) is scored under Item 6. Therapist's Interpersonal Effectiveness).

The therapist should adopt a style that promotes an adaptive relationship with the couple with regard to providing input, being engaged in the therapy and working proactively towards agreed goals. This is achieved through a therapeutic approach that promotes 'teamwork.' Collaboration should be consistent throughout the session, and the therapist should avoid being unnecessarily controlling or passive. At the same time as promoting a 'teamwork' approach, the therapist attempts to avoid an inappropriate power relationship from developing between the couple and the therapist.

The therapist creates an atmosphere in which both partners experience respect and that each person's input is valued and taken into account. The therapist is respectful of variations among couples and cultural influences regarding gender, age, and other factors that might impact how the couple interacts with the therapist. The therapist works to create a collaborative atmosphere in which any such factors are clarified and their implications for the working alliance are considered.

Good collaboration will also involve achieving a balance between verbal and non-verbal features: for example, deciding when to talk and when to listen; when to intervene and when not to intervene, and when to offer suggestions and when to wait for the couple to devise their own. At times effective teamwork may necessitate quite a didactic approach. In contrast to individual CBT, the style of collaboration in BCT is likely to be more active and directive and in order to be optimally 'collaborative' with the couple and manage the session effectively, the therapist may need to provide recommendations, give instructions, or interrupt an unproductive interpersonal exchange. At other times, the therapist may take a 'step back' and work to elicit the couple's own understanding and suggestions in order to enable the couple to make choices and take responsibility. Regardless of the level of directiveness employed, the therapist should demonstrate skill in encouraging the couple to participate fully (e.g., through use of appropriate questions, guidance, shared problem-solving and decision-making), with a positive evident impact on the couple and/or the way the session unfolds.

Features of competent performance

Competent performance on Collaboration is evidenced through the following three features:

- 1. The therapist's verbal skills*

The therapist's verbal statements provided evidence of attempting to create a working environment that fostered collaboration and increased the likelihood of a working alliance that could support the task-focused elements of therapy.

Scale development: BCTS-D

- 2. The extent to which the therapist sought, or attempted to seek, the couple's thoughts, insights, suggestions, and ideas.*

The therapist was explicit in encouraging the couple's ownership of the session through inviting suggestions, ideas about and responses to the content of the session. The therapist was attentive to, and encouraging of, the couple's active participation.

- 3. The contributions of each party present in the session were balanced, such that it was not the therapist talking all the time, or the therapist allowing the couple/one member of the couple to dominate the session or talk in a rambling, unstructured way.*

All parties had an opportunity to contribute to the session. Where more time was spent focusing on one partner, there was a clear rationale for this, with the therapist ensuring that both partners understood and were comfortable with the therapist's approach. The therapist neither dominated the session (i.e., lecturing the couple) nor became passive (allowing the couple or one partner in the couple to dominate the conversation).

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>No attempt at collaboration was made and the concept of teamwork appeared to be entirely absent from the session.</p> <p>The therapist was over-controlling of the session to the extent that the couple was unable to participate in the session, and the impact was damaging to the therapy in some way.</p> <p>OR: The therapist talked over the couple most of the time so that the couple was silenced when attempting to make a contribution and share their ideas and feedback.</p> <p>OR: The therapist inappropriately allowed the couple to 'run' the session such that the therapist's contributions were almost entirely absent from how the session unfolded, resulting in an unproductive or unhelpful session.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to work collaboratively but did so ineffectively such their performance was highly limited.</p> <p>For example, the therapist was clearly struggling (if attempting) to enable a teamwork approach and it seemed as though therapist and couple were constantly 'competing' for the chance to be heard.</p>

Scale development: BCTS-D

	<p>OR: The therapist attempted to convey the importance of an egalitarian relationship but used an overly authoritarian approach that was likely to have been off-putting for the couple.</p> <p>OR: The therapist was overly passive, even when the session clearly required a much more assertive and directive stance in order to remain 'on track'.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback.</p>
2	<p>The therapist struggled to make effective decisions concerning how to strike a balance between verbal and non-verbal features and as a result, collaboration was inconsistent. There were, however, times when the therapist appeared to achieve a good degree of teamwork.</p> <p>Overall, there was a sense of collaborative working, where the therapist aimed for an adaptive, respectful relationship between partners and with the therapist. However, there might have been occasions where the therapist was over/under-controlling in a way that did not appear connected to the needs of the couple in that particular session.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor's feedback.</p>
3	<p>Overall the therapist achieved a good degree of collaboration in the session. The couple was clearly 'socialised' to the idea of team working, and all parties contributed to the session in a productive way. Where there were difficulties (for example, in the context of escalating negative reciprocity, or one partner being verbally hostile or critical to the other) the therapist exercised appropriate judgement in how directive or non-directive to be.</p> <p>At times, the therapist's ability to ensure an appropriate level and type of collaboration may have been inconsistent, but overall the essential components were present, with the three features of competent performance (listed above) evident.</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist established a collaborative working relationship that was maintained effectively throughout the session. The therapist conveyed a good understanding of when to be more directive and when to wait for the couple to offer their own ideas/feedback. The 'tone' of the session conveyed effective teamwork such that the couple was enabled to achieve an appropriate level of ownership of the ideas discussed. The therapist was appropriately transparent in sharing their hypotheses, recommendations, and summaries.</p> <p>In summary, the therapist's performance was competent on this item with only minor</p>

Scale development: BCTS-D

	<p>areas of refinement needed. There were no concerns about the therapist's performance on this item with the therapist demonstrating a good, solid level of proficiency.</p>
5	<p>The therapist actively worked to promote a collaborative working relationship and was very effective in achieving an egalitarian approach. The couple's ideas, experiences, efforts and feedback were encouraged and welcomed and the therapist was transparent in their thinking and actions in ways that enhanced the effectiveness of the session.</p> <p>In promoting an effective team working style, the therapist achieved a high level of skill in balancing more and less direction as a function of the moment-to-moment interactions between the couple, and between the couple and the therapist, and did so with ease.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. Nothing inappropriate was done, and nothing important was omitted. The proficiency demonstrated was simply not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist's approach to developing and maintaining an effective, collaborative relationship in the session was extremely skilful. The therapist demonstrated excellence in creating effective team working such that the couple felt empowered to participate fully, sharing their ideas, experiences and offering feedback, and were enabled to take a high degree of ownership of the session. This is likely to have been experienced by the couple as empowering and instilling hope – even where difficult emotions were present.</p> <p>Where this was a difficult session, the therapist was still able to ensure effective teamwork and a collaborative stance that enabled forward movement.</p> <p>The therapist was clearly able to become more and less directive as a function of the needs of the couple and demonstrated an extremely high level of skill in exercising judgement about the style of collaboration adopted.</p> <p>In summary, the therapist demonstrated a superior level of skill on this item, and would be deemed to be working at 'master level'. A score of 6 is rare and signifies exceptional performance. The skill demonstrated is such that one might select this for a training session to demonstrate the highest levels of performance on collaboration.</p>

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of good performance:

1. Were both partners in the couple encouraged to participate fully in the session?

Scale development: BCTS-D

2. Was the therapist clearly welcoming of the participation of both partners' contributions?
3. Did the therapist convey genuine interest in what the couple had to share?
4. Was the therapist able to establish a positive working alliance with the couple, and with each partner?
5. Did the therapist give the couple sufficient space to think and respond to the content of the session without losing control of the session?
6. Was the therapist overly directive or too controlling, such that the couple's full participation was prevented or constrained in some way?
7. Was the therapist too passive or inactive, such that the session was 'couple led' rather than collaborative?

Examples of statements or actions that reflect these indicators include the following:

- Do you have any thoughts about how we could test out that idea?
- Let me get your thoughts on what we have just discussed. How might it be relevant to what you want to achieve from coming here?
- Perhaps we could figure out, together, an alternative way of looking at/approaching this issue.
- Before agreeing to this (homework) task, let's identify some of the obstacles that might prevent us learning anything from it.
- That's a difficult one, so let's put our heads together and try and think it through.
- Could you help me make sense of this?
- Let's look at this together.
- You have a lot of experience living with/attempting to manage this problem, so could you help me understand a bit more about.... from your perspective?

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 4. Facilitating Couple Communication

Introduction

Facilitating couple communication is a central feature of BCT and, as such, will be present in many BCT sessions. However, the form that this takes may vary as a function of the difficulties for which a couple is seeking help, the stage of therapy, and any process issues arising in the session that require therapist intervention.

In order to facilitate couple communication, the therapist will need a prior awareness of the patterns of interaction that are typical for the couple (this will be aided by a well thought through dyadic conceptualisation, scored under Item 13). The therapist will also need to demonstrate an ability to manage the session, steering a couple towards effective communication in situations where there is a high degree of conflict and/or where couple communication skills have not yet been consolidated.

As effective BCT requires the facilitation of a dialogue between partners, in most instances, it is preferable to avoid (unless a high degree of conflict necessitates this) a situation whereby the therapist directs attention to one partner and then the other, conveying an impression of conducting individual therapy with an extra person (i.e. the person's partner) in attendance.

Effective performance on Item 4 will require moment-to-moment choices about the communication needs of the couple. For example, at times, the therapist may allow one partner, or the couple, considerable latitude in finding the words to express themselves (for example, where a couple is working productively towards mastery of interventions such as 'sharing thoughts and feelings' and one partner is working actively to extend his/her emotional vocabulary). However, at other times, in order to manage the session optimally, the therapist may be highly directive, interrupting one or both partners to enable a more effective communication strategy.

The therapist seeks opportunities to encourage relationally schematic processing (understanding how each person's actions influence the others' and how patterns evolve over time) through priming, summarising, and drawing each partner's attention to salient information that may have been overlooked and which is essential to enabling more effective couple communication.

(NB: In contrast to Item 6 where the emphasis is on the therapist's own interpersonal effectiveness, Item 4 is concerned with the extent to which the therapist demonstrates an ability to facilitate communication between the couple.)

Features of competence performance

Competent performance on Facilitating Couple Communication is evidenced through the following features:

1. *The therapist directed the couple towards interacting with each other*

The therapist capitalised on opportunities to encourage the couple to communicate with one another (as opposed to each partner communicating primarily with, or through, the therapist).

2. *The therapist directed attention to what was missed*

Scale development: BCTS-D

The therapist listened to the essence of what each partner sought to communicate and directed the couple's attention to information on thoughts, feelings, behaviours, preferences, motivations and needs that the partner may have overlooked.

3. *Management of the process of interaction*

The therapist used an appropriate amount of structure and directiveness to support effective communication, curtailing hostile, critical or other forms of dyadic interactional difficulties swiftly and effectively.

4. *Attention to both positive and negative communication*

Positive and negative communication operate somewhat independently of each other. The therapist intervened to decrease negative communication and also made efforts to increase positive, facilitative communication, including both the content of communication and the verbal and nonverbal ways in which it was communicated.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist made no attempt to facilitate interaction between the couple. The session was spent effectively conducting individual therapy with the other partner present.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to facilitate couple communication but did so ineffectively such their performance was highly limited. For example, most of the time, the therapist appeared to be conducting individual therapy but periodically, there was a recognition that efforts needed to be more 'couple-focused', although this seemed to be an afterthought and was implemented in a clumsy fashion.</p> <p>OR: the therapist attempted to facilitate couple communication but seemed lost as to how to approach this, and the couple continued with maladaptive communication.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback.</p>
2	<p>The therapist clearly attempted to facilitate effective couple communication and at times, appeared to be close to doing so quite effectively or did so but only for limited portions of the session. However, there were evident difficulties with a tendency to:</p> <p>Conduct individual therapy with the partner in the room (when it was clear that this was not an intentional strategy on the therapist's part).</p>

Scale development: BCTS-D

	<p>OR: The therapist struggled to promote effective couple engagement, losing control of the interactional process at times.</p> <p>OR: The therapist was not sufficiently clear and structured in the approach taken such that the couple is likely to have felt confused by what was expected of them.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor's feedback.</p>
3	<p>Overall, the therapist demonstrated an ability to facilitate couple communication in the session, although at times this was inconsistent. Nonetheless, the essential components were present, with the four features of competent performance (listed above) evident.</p> <p>The therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist was able to facilitate couple communication effectively throughout the session, capitalising on opportunities arising to promote this skill between partners. The four features of competent performance (listed above) were evident, and the outcome was that the couple was enabled to communicate more effectively about issues central to their goals, the case conceptualisation, or treatment plan.</p> <p>In summary, the therapist's performance was competent on this item with only minor areas of refinement needed. The therapist demonstrated a good, solid level of proficiency.</p>
5	<p>The therapist actively worked actively throughout the session (to the degree needed) to promote effective couple communication and demonstrated considerable skill in doing so. For example:</p> <p>The therapist was able to capitalise on opportunities arising in the session to promote this skill, provided appropriate direction to the couple to keep them 'on track' with their efforts to communicate effectively.</p> <p>AND: The therapist was able to highlight information that one partner may have overlooked and that may have been critical to enabling each partner to feel heard and understood.</p> <p>AND: The therapist demonstrated skill in knowing when to more actively 'direct' the process and when to take a step back, guiding the moment-to-moment interactions between the couple more indirectly.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. Nothing inappropriate was done, and nothing important was omitted. It simply was not at the exemplary level of master which is required for a score of 6.</p>

Scale development: BCTS-D

6	<p>The therapist demonstrated exceptional proficiency in promoting effective couple communication. For example:</p> <p>The therapist consistently capitalised on opportunities to promote this skill and provided appropriate direction to the couple to keep them 'on track' with their efforts to communicate effectively.</p> <p>AND: The therapist was attentive to, and was able to highlight, information that one partner may have overlooked and that may have been critical to enabling each partner to feel heard and understood.</p> <p>AND: The therapist demonstrated superior skill in knowing when to more actively 'direct' the process and when to take a step back, providing a seamless approach to guiding the moment-to-moment interactions between the couple.</p> <p>AND: Where strong, negative emotions were present for one or both partners, the therapist navigated these extremely effectively, enabling the couple to find new ways to communicate in the face of difficult internal experiences.</p> <p>In summary, the therapist demonstrated a superior level of skill on this item, and would be deemed to be working at 'master level' such that this might be selected for a training session to demonstrate highest levels of performance.</p>
---	---

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of good performance:

1. Did the therapist seek opportunities to facilitate couple communication?
2. Did the therapist's approach to facilitating couple communication appear to be informed by the dyadic conceptualisation, guiding the process in a way that reflected their apparent understanding of the couple's strengths and limitations?
3. Where there appeared to be the potential for escalating negative reciprocity, did the therapist intervene and put the interaction 'back on track'?
4. Did the therapist resort to conducting individual therapy with the partner in the room?
5. Was the therapist too passive or inactive, such that opportunities to facilitate couple communication in a way that is consistent with a BCT approach were lost? Did the therapist seem 'run over' by the couple?
6. Was the therapist too directive or controlling (for the needs of the couple as they presented in the session), such that opportunities to facilitate couple communication in a way that is consistent with a BCT approach were lost?

Examples of statements or actions that reflect these indicators include the following:

Scale development: BCTS-D

- I think (your partner) was saying something else just then, something more tender that you overlooked. Can you remember what that was?
- I think (your partner) may have not fully understood what you were saying just then. Could you find another way to express that, so he/she can really appreciate your perspective on this?
- You say some very loving things towards one another, but you have a tendency to miss that...
- Let me interrupt you there - I don't think it's going to be helpful to 'hash over' what happened – it will just end up dominating the session. Let's see if we can find another, more helpful way, to put your partner in your shoes to understand what that situation was like for you.
- I'm going to stop you there. One of the things I notice happening is that, as you start to describe the emotion, you get into it and it starts to escalate. I want you to see if you can find another, more controlled way to express the emotion (to partner). Let's see if we can try that here and now.
- Try that again and this time see if you can find a more neutral, or more positive way to say that.
- Do you both feel that you now understand each other on that issue?
- Alternatively, where the therapist senses the emergence of escalating conflict, there is evidence of the therapist intervening through (a) halting the conversation; (b) talking directly to one of the partner, conveying empathy understanding and a recommendation of how to approach the conversation and then (c) putting the couple 'back on track' to continue the conversation from a more productive starting point.

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 5. Pacing and Flow

Introduction

Therapy is likely to proceed in an optimal fashion when adequate attention is given to issues of pacing and flow. The session should be well-paced in relation to the items identified on the agenda, with sufficient time allocated to address key issues. The session should flow smoothly through discrete phases, with a clear beginning, middle and concluding/ending phase.

The session should be paced in a way that is responsive to the couple's needs. For example, the therapist may slow the pace where a couple is struggling to grasp key concepts or interventions, and move more swiftly where there the couple is working effectively with specific ideas or interventions that the therapist has introduced previously. Where the therapist manages the pacing and flow skilfully, the overall impression is of a session that is well time-managed, neither too slow nor too quick, and progressing at a speed that accommodates the couple's needs and speed of learning. In pacing the session, the therapist needs to maintain sufficient control to guide the couple's use of the time, limit discussion of peripheral issues and curtail unproductive interactions, whilst also avoiding rushing through the priority agenda items. At times the therapist needs to slow the session to make certain that both partners are listening to each other, to lower emotion, etc. At other times, the therapist might attempt to increase the pace of the session when it is dragging, one or both partners seem lifeless or less engaged, etc. Appropriate pacing and flow also involves bringing the session to a close so that it does not merely end as time elapses.

Features of competent performance

Competent performance on Pacing and Flow is evidenced through the following features:

1. Movement between discrete phases of the session

The therapist enabled a smooth transition between the start, middle, and end of the session.

2. Adherence to the agenda items and session priorities

The session was paced in such a way that the agenda items and session priorities were addressed without the session seeming rushed. Adequate time was available for reflection, discussion, and any necessary problem-solving.

3. Degree of congruence between the pace of therapy and the learning speed of the couple

The therapist was able to quicken and slow down the pace of the session to accommodate the speed and ease with which the couple was able to absorb and digest the information discussed.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	The session was extremely badly paced to the extent that was clearly no evidence of
---	---

Scale development: BCTS-D

	<p>the therapist attending to the pacing and flow of the session.</p> <p>OR: the therapist seemed so rigidly attached to his/her own agenda for the session (explicit or not) that the couple was rushed through or otherwise 'managed' through the session in a way that was likely to have caused damage to the working relationship.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to attend to matters of pacing and flow but did so ineffectively such performance was highly limited. For example, most of the time, the therapist seemed oblivious to the need to move discrete phases of the session (beginning, middle and end) but periodically, there was a recognition that the session needed to gather momentum, or the pace needed to be reduced. Efforts were made but almost as an afterthought and the pacing and flow of the session was approached in a clumsy fashion.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback.</p>
2	<p>The therapist clearly attempted to attend to matters of pacing and flow and at times, appeared to be close to doing so quite effectively. However, there were evident difficulties with a tendency to:</p> <p>Move between the beginning, middle and end phases of the session in an awkward or 'clunky' fashion.</p> <p>OR: The therapist spent too long on one particular phase of the therapy (e.g., too much time was devoted to setting the agenda as opposed to addressing the items on the agenda), and then other phases of the session were rushed.</p> <p>OR: The time appeared to 'run away' with the therapist such that he/she became aware towards the end of the session that important tasks had been neglected until it was too late in the session to address them productively.</p> <p>OR: The therapist appeared to be pacing the session in a way that did not suit well the speed of learning and information-processing of the couple.</p> <p>In summary, the therapist demonstrated an emerging competence on this item, but further work is needed to reach competence, with key areas for development identified in the assessor's feedback.</p>
3	<p>Overall, the therapist demonstrated an ability to manage the pacing and flow of the session effectively, although at times this was inconsistent. Nonetheless, the essential components were present, with the three features of competent performance (listed above) evident.</p>

Scale development: BCTS-D

	<p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist was proficient in attending to issues of pacing and flow throughout the session, creating a smooth transition between the different stages of the session (beginning, middle and end). The session did not appear rushed, nor did it drag, and there was time for processing of key ideas. The three features of competent performance (listed above) were evident and the outcome is a session that was well-paced.</p> <p>In summary, the therapist’s performance was competent on this item with only minor areas of refinement needed. There were no concerns about the therapist’s performance on this item with the therapist demonstrating competence and good performance.</p>
5	<p>The therapist demonstrated a high degree of skill in managing pacing and flow throughout most of the session. There was an excellent, largely seamless transition between the start, middle, and end of the session; the agenda items were addressed effectively and efficiently, without the session feeling rushed. The therapist also appeared to be aware of, and attentive to, the learning speed of the couple and altered the pace accordingly.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. Nothing inappropriate was done, and nothing important was omitted. It simply was not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist demonstrated an excellent approach to managing the pacing and flow of the session and this was consistent across the session. As for 5 (above), the therapist was able to move smoothly and seamlessly between the discrete phases of the session, with all phases – start, middle and end – afforded sufficient time for the couple to be able to consolidate their understanding and learning. Priorities for the session were address effectively without the session ever appearing rushed or dragging.</p> <p>Adequate time was available for reflection, discussion and problem-solving where this was needed.</p> <p>The therapist was clearly attentive to the speed of learning of the couple and of the individual partners and appeared to be making moment-to-moment decisions about how to pace the sessions accordingly.</p> <p>In summary, the therapist demonstrated a superior level of skill on how the session was paced and flowed from one area of discussion to the next. The therapist would be deemed to be working at ‘master level’.</p>

Scale development: BCTS-D

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of good performance:

1. Did the therapist manage the pace and flow of the session in a way that seemed responsive to the couple's needs and abilities in that session?
2. Was there any time during the session when the session moved too slowly/quickly?
3. Was the therapist able to curtail unproductive digressions and/or intervene during moments of escalating conflict to ensure that disruptions to the smooth pacing and flow of the session were prevented?
4. Was sufficient time allocated to reviewing the key points of the session and any homework assignments agreed?
5. Did the couple appear 'rushed' by the therapist at any point during the session?
6. Did the session come to a close rather than just ending because time expired?

Examples of statements or actions that reflect these indicators include the following:

- How much time should we spend on this topic?
- Let's pause for a moment. You've given me lots of information so I want to make sure I have understood all the key points you are making. So if I summarise....
- We may have gone off track a little. Shall we get back to the key topic that you wanted to discuss today?
- I wonder if now would be a good time to move on to the next item on our agenda.
- We have around 15 minutes left before the end of the session. Is there anything you think we must cover before the end?
- Have we covered this topic in sufficient depth for today, so that we can move on to the next item we wanted to discuss?

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 6. Therapist's Interpersonal Effectiveness

Introduction

In order for therapy to progress effectively, it is essential for the therapist to be able to put the couple at ease whilst simultaneously conveying an appropriate air of confidence and authority. The therapist needs to be competent in using both verbal and non-verbal strategies to communicate effectively with the couple, to instil hope and to provide an interpersonal climate that is likely to foster the couple's confidence in disclosing information that is critical for the work of therapy to proceed. This needs to be balanced with an ability to respond constructively to any emerging patterns of maladaptive communication (for example, where a couple persists in high levels of criticality or where one partner attempts to encourage the therapist to 'take sides'). (NB: The therapist's effectiveness in facilitating communication between the couple is rated separately under Item 4. Here, the focus is the therapist's own interpersonal skills to engage the couple and to create an emotional climate that increases the likelihood of an effective session).

In addition to conveying warmth, optimism, and interest in the couple, the therapist also needs to convey credibility as an expert on relationship functioning and in their knowledge of how to intervene to bring about positive change. Couples can assign considerable importance to a variety of visible therapist demographic characteristics, such as age, cultural background, gender, etc. and the therapist needs to demonstrate an ability to respond constructively and non-defensively to any questions concerning the therapist's professional credentials and ability to provide adequate support.

Features of competent performance

1. Ease of relating

The therapist conveyed an appropriate level of ease in his/her style of relating with the couple whilst maintaining a professional manner. Self-disclosure was used appropriately and thoughtfully where this was prudent for therapeutic reasons (for example, to reassure a couple that the therapist has sufficient life experience and knowledge of therapy to be able to help them).

2. Understanding and empathy

The therapist conveyed an interest in, and acceptance of, the couple that was likely to have been experienced by the couple as evidence of the core conditions of warmth, genuineness, empathy and understanding. At the same time, the therapist was willing to constructively challenge aspects of the couple's functioning where this proved necessary. Understanding and empathy is balanced with providing structure as needed to promote an optimal environment for therapy. The therapist did this in a comfortable manner that conveyed concern for the couple and assurance that a safe environment existed for the couple to work.

3. Appropriate authority

While adjusting their interpersonal style to the needs of the couple (e.g., to accommodate preferences, backgrounds and cultural issues), the therapist maintained professional

Scale development: BCTS-D

boundaries and conveyed credibility in their knowledge of couple relationships. This credibility involves both knowledge along with the ability to process what is happening in the session in real time such that the therapist maintained the big picture of what was happening along with attending to the momentary interactions between the partners.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist's style of communication was highly inappropriate and/or ineffective in ways that are likely to have undermined the therapist's credibility in the eyes of the couple. The therapist made comments that indicated that the therapist did not understand the couple or came across in a critical, evaluative manner toward one or both partners.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist periodically attempted to come across as interpersonally effective, but performance in this area was highly limited. For example, the attempts to achieve an ease of relating may have resulted in an inappropriate use of self-disclosure, humour, or authority.</p> <p>OR: The therapist could not respond in a reassuring and appropriately authoritative manner to questions concerning the therapist's experience and credibility.</p> <p>OR: It was evident that the therapist 'sided' with one partner in a way that was potentially damaging to the therapeutic relationship.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback.</p>
2	<p>The therapist was clearly aiming for interpersonal effectiveness and at times, appeared to be close to achieving this. However, there were evident difficulties with a tendency to:</p> <p>Appear overly formal or informal in ways that may have made the couple feel anxious or uncertain about the therapist.</p> <p>OR: The therapist was unhelpfully reticent or clearly under-confident in responding to questions about the therapist's knowledge and experience.</p> <p>OR: the expression of the 'core conditions' was inconsistent.</p> <p>OR: The therapist could not adjust their interpersonal style to the needs of the couple.</p>

Scale development: BCTS-D

	<p>In summary, the therapist may have demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor's feedback.</p>
3	<p>Overall, the therapist demonstrated a sound level of interpersonal effectiveness. For example:</p> <p>Overall, the therapist demonstrated credibility and appropriate authority.</p> <p>AND: Understanding and empathy were conveyed.</p> <p>AND: The therapist appeared to relate easily to the couple.</p> <p>In summary, although interpersonal effectiveness may not have been entirely consistent throughout the session, overall the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist was proficient in this area. For example:</p> <p>The therapist conveyed both confidence and authority in appropriate and helpful ways.</p> <p>AND: Accurate understanding and empathy were conveyed to the couple frequently throughout the session.</p> <p>AND: The therapist provided the core conditions with relative ease.</p> <p>In summary, the therapist's performance was effective on this item with only minor areas of refinement needed. The therapist demonstrated a good, solid level of proficiency overall, with some aspect standing out throughout the session or at some point during the session as above the norm.</p>
5	<p>The therapist demonstrated a high degree of skill in this area. Interpersonal effectiveness was consistent throughout the session, with the therapist conveying both their authority as a credible relationship expert and an understanding of how to alter their style of interpersonal approach in order to convey the 'core conditions' optimally for the couple.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. Nothing inappropriate was done, and nothing important was omitted. It simply was not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist demonstrated excellence on this item. As for 5 (above), interpersonal effectiveness was consistent throughout the session. The therapist conveyed an ease of relating that enabled the couple to engage with the session. The 'core conditions' were offered in a way that reflected both the therapist's authentic style and seemed to take account of the preferences and needs of the couple, including any relevant</p>

Scale development: BCTS-D

	<p>cultural issues (e.g. cultural differences in how health care professionals are perceived and related to). Throughout the session, the therapist came across as an expert in couple relationships in a way that the couple appeared to find reassuring.</p> <p>In summary, the therapist demonstrated a superior level of skill on this item, and would be deemed to be working at 'master level'.</p>
--	---

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of good performance:

1. From what you observed/heard, did you consider the therapist to be interpersonally effective?
2. Did the therapeutic relationship with the couple seem positive and productive?
3. Did the therapist display appropriate empathy, understanding, warmth, and genuineness?
4. Did the therapist's style encourage trust and disclosure by the couple?
5. Did you conclude that the therapist showed appropriate respect and valuing of the couple and each individual partner, while retaining professional boundaries?
6. Did the therapist convey confidence and appropriate authority?
7. Did the therapist appear able to adapt his/her interpersonal style (e.g., degrees of formality/informality, appropriate use of humour, awareness of cultural issues in informing style of interaction between doctor and patient, etc.) to the interpersonal and communication style preferences of the couple?
8. Were any cultural issues particularly relevant and, if so, did the therapist appear to convey an awareness of these?
9. Were any indirect (or direct) challenges of the therapist's authority/credibility handled appropriately (e.g., questioning whether the therapist had sufficient life experience to be qualified to provide relationship guidance)?
10. Did the therapist use their interpersonal style not only to decrease negatives but also increase positives? Did you hear any evidence of this in the session?

Examples of statements or actions that reflect these indicators include the following:

- (Summarising what each partner said to convey understanding and empathy) Have I understood you correctly?
- Shared laughter
- In my opinion/experience what I think is likely to prove helpful here is...
- One of the things the research evidence tells us is...
- Let's look at how we might be able to draw on these general principles and apply them to your current concerns
- Statements designed to convey empathy and understanding, such as, 'That must have felt incredibly difficult for you'

Scale development: BCTS-D

- Statements designed to convey appreciation, recognition, and validation of the couple such as, 'You've made a great effort here. Thank you'; 'Despite the huge difficulties, you did really well'
- Many people would feel that way, but you have decided to do something about it.

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 7. Guided Discovery

Introduction

Guided discovery has been interpreted in the literature in different ways. In BCT the term is often used to denote a specific, planned and proactive cognitive intervention. However, for the purposes of rating therapist performance on the BCTS-D, guided discovery is used to denote the therapist's style of engaging the couple, as described by Padesky (1993; see below). This style enables an on-going process of exploration, questioning and summarising to assist the couple in gaining new perspectives, knowledge and understanding *for themselves*, without the use of unhelpful confrontation, debate, or lecturing. As a result, evidence of guided discovery should be present throughout the session and is used to pave the way for therapeutic change.

Underpinning the use of guided discovery are a number of principles concerning the ways in which therapists can elicit new perspectives that have implications for change. Specifically, there is an assumption that people are most likely to adopt new perspectives on their circumstances and needs if they believe that they have arrived at these perspectives themselves (as opposed to being 'told' or challenged by the therapist). In consequence, the therapist uses a style of questioning and exploration that can facilitate the couple's ownership of any new perspectives or solutions identified.

The therapist facilitates a process of exploration that relies upon genuine interest, curiosity, and understanding. Skilful questioning is used as well as empathic listening, use of capsule summaries, and so-called synthesising questions (Padesky, 1993). This style of enquiry helps the couple develop, examine and synthesise novel hypotheses regarding their difficulties and experiences. It is important to note, however, that there are differences in the style of guided discovery used in BCT as opposed to individual CBT. Specifically, effective management of 'couple phenomena' tends to require greater 'directiveness' on the part of therapist and may even, at times, convey more of an 'instructional' tone.

BCT therapists are also less likely to use the more traditional, Socratic-style questioning typical of individual CBT, and it is *unlikely* that the therapist will seek to 'expose' inaccuracies in one partner's interpretation of an event (for example, working with one individual to identify examples of where he or she is processing information in a cognitively biased way). This is because such an approach has the potential to evoke criticism from the partner ("I always said that your thinking is messed up") or leave the individual feeling embarrassed or shamed for having been 'singled out' for their 'flawed' attributions. To avoid a situation where one partner may feel vindicated and the other defensive, guided discovery in BCT is used to set up to enable conversations or encourage interactions outside of the session (e.g., through homework that might provide a new set of experiences for the couple and result in different cognitions) that promote thinking about topics (and each other) in new ways. The aim is to help each individual create a balanced, reasonable perspective on their partner, the relationship and the environment, to reduce distorted or extreme thinking (in an either positive or negative direction) and to generate a style of conversation in which it is possible to change problematic attributions without directly challenging them.

Features of competent performance

Scale development: BCTS-D

The couple should be supported in developing and reassessing their beliefs relevant to their current situation, and in generating potential solutions, with varying degrees of guidance and direction from the therapist as needed. Competence performance on this item is evidenced through the following three features:

1. Conveying interest and curiosity

The therapist's style of guiding discovery was open, curious and non-judgemental.

2. Appropriate level of 'directiveness'

The therapist provided a level of direction in guiding the couple and offering suggestions, recommendations and instruction, that was optimal for the stage of therapy, the couple's strengths and limitations (i.e., what they can, and cannot, problem-solve unaided) and having taken account of the couple's pattern of interaction (e.g., a couple who are experiencing high levels of conflict are likely to require a greater level of instruction than those who do not).

3. Facilitating new learning and understanding

The process of guided discovery enabled by the therapist resulted in the potential for new learning for the couple, helping them uncover novel perspectives that created better self-understanding and understanding of each other which could be used to facilitate change. The therapist avoided premature problem-solving in order to work with the couple to explore and uncover deeper, richer and fuller possibilities for themselves and their relationship. The emphasis was placed on whether the therapist helped the couple design a reasonable new experience with the potential to alter their cognitions, rather than the outcome of these efforts. That is, if the therapist provided appropriate guided discovery yet one or both partners maintained their distorted beliefs, the therapist would receive a high score on this item; the focus is on the therapeutic process rather than the outcome.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	There was no evidence of guided discovery. The couple was 'lectured' as to what is correct and incorrect, and/or what they should do. OR: The therapist did not provide any guidance, offer any new perspectives, or provide appropriate direction, creating evident confusion or distress for the couple. In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.
1	The therapist attempted to engage the couple in a process of guided discovery but did so ineffectively and as such, performance was highly limited. This may have taken the form of any of the following:

Scale development: BCTS-D

	<p>The style of questioning came across as patronising or otherwise inauthentic.</p> <p>OR: The therapist highlighted the ‘flaws’ in one partner’s cognitive processing or specific cognitions that may have left them feeling belittled or shamed in front of their partner.</p> <p>OR: The therapist asked a series of questions but failed to convey empathic understanding and/or use any capsule summaries to help the couple consolidate any new, emerging insights.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor’s feedback.</p>
2	<p>The therapist attempted to engage in guided discovery but there were difficulties in how this was conducted. The therapist’s actions showed evidence of emerging competence on this item, but the delivery was problematic. For example:</p> <p>The therapist asked questions but was clearly leading the couple to a pre-determined ‘right answer’ that resulted in the couple temporarily disengaging or leaving one or both partners feeling unheard and misunderstood (i.e., ‘forced discovery’ rather than ‘guided discovery’).</p> <p>OR: The therapist was insufficiently directive in their approach when this was clearly needed in order to facilitate new learning.</p> <p>OR: The style of questioning came across as slightly patronising.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor’s feedback.</p>
3	<p>Overall the therapist made good use of guided discovery. The therapist conveyed interest and curiosity, and provided an appropriate level of direction and (if necessary) instruction to the couple to facilitate new learning, although there was a degree of inconsistency in performance across the session. Or if the guided discovery focused on activities the couple was to conduct outside the session, the experience was structured appropriately to facilitate the likelihood of new learning.</p> <p>In summary, the therapist’s performance was broadly competent on this item. The essential features were present, and only minor inconsistencies or areas of refinement needed. Refinement would be beneficial.</p>
4	<p>The therapist provided clear evidence of skill in guided discovery. The essential features of conveying genuine interest and curiosity were present and the therapist provided the right level of direction in guiding the couple and offering suggestions, recommendations, and instruction that were optimal for the stage of</p>

Scale development: BCTS-D

	<p>therapy, the couple’s strengths and limitations and having taken account of the couple’s communication style. The result was that the couple was enabled to discover new ideas and make new connections between aspects of their understanding and experience that were likely to support further learning and change.</p> <p>In summary, the therapist’s performance was clearly competent on this item, with only minor areas of refinement needed. There were no concerns about the therapist’s performance on this item.</p>
5	<p>The therapist used guided discovery highly effectively. All of the key features were present. The therapist conveyed interest and curiosity, and whilst remaining non-judgemental of both partners, was able to guide and challenge effectively where this was needed. The level of directiveness provided was appropriate and conveyed confidence and authority combined with warmth and a genuine desire to learn about the couple’s experience. Both partners were, through the therapist’s use of guided discovery across the session, enabled to arrive at new understandings that could pave the way for potential change (or the therapist created optimal conditions for new learning even if it did not occur for one or both partners). Performance was consistent across the session.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. Nothing inappropriate was done, and nothing important was omitted. It simply was not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist’s use of guided discovery was at ‘master’/expert level. All of the above features (listed under 5 above) were present. In addition, the therapist demonstrated an ability to personalise guided discovery and to adapt the elements to process issues encountered within the session. There was excellent guided discovery leading to a deep level of understanding for the couple. Where difficulties occurred in the session, the therapist remained highly effective, with evidence of a deeper understanding of information relevant to the case conceptualisation or treatment plan having been attained. The therapist avoided premature problem-solving in order to allow time for a fuller, richer and deeper level of exploration.</p> <p>In summary, the therapist demonstrated a superior level of skill on this item, and would be deemed to be working at ‘master level’.</p>

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of performance:

1. Did the therapist listen carefully, making use of ‘capsule summaries’ to check with the couple that he/she was understanding their intended meaning/s?

Scale development: BCTS-D

2. Did the therapist convey empathic understanding of each partner's perspective?
3. Did the therapist use appropriate questions?
4. Did the manner in which the questions were asked appear to facilitate the couple's understanding?
5. Did the questions result in or facilitate change, however small?
6. Did the therapist avoid the use of questions that might leave one partner feeling 'exposed' for the 'inaccuracies' of their thinking?
7. Did the therapist help construct an experience that could, either during the session or outside the session, provide optimal opportunity for new cognitions to develop or distorted ones to be challenged, regardless of the outcome of these efforts?

Examples of statements that reflect these indicators include the following:

- What's important about this topic for each of you?
- You each have different ideas about this situation. I wonder if we could look at the pros and cons of each of these...
- Can you see any potential advantages to how your partner approaches this? And can you see any potential disadvantages to your own approach?
- Can you elaborate on that – why is that important for you?
- There may be recent or past events that have contributed to this issue/what you are struggling with now. Let's see if we can identify what some of these might be.
- Give me an idea about how you might bring two perspectives together.
- I wonder if we might try giving you a new experience outside of our session before we get together again. How about if we...
- The therapist discusses various issues that are likely to change the attribution without directly challenging it.

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 8. Formulation of Depression in Context

Introduction

The presence of psychopathology diminishes an individual's world and in the context of depression, it can be anticipated that a couple's relationship will change in a negative direction. At the same time, a dysfunctional relationship represents a chronic stressor that is likely to be implicated in both the onset and development of the patient's depression, as well as a vulnerability factor for relapse following recovery. The therapist needs, therefore, to be able to harness the relationship as a resource that can assist recovery from depression, whilst also working to eliminate relationship distress if present. An understanding of the impact of the depression on the partner is also important, with the therapist able to raise and address coping and self-care for the non-depressed partner if this proves necessary.

Item 8 assesses the extent to which the therapist is able to identify and focus the session on the relevant individual, couple, and environmental factors that appear implicated in the patient's depression. The therapist needs to be clear as to whether their primary objective is to (1) target the disorder or (2) improve the relationship during a given session or portion of a session, recognizing that some interventions might impact both. Competence on this item is demonstrated through the therapist's statements and actions reflecting a clear objective in making this differentiation and focusing accordingly.¹⁸

Features of competent performance

Competent performance on this item is evidenced through the following three features:

1. *An evident knowledge of depression (at the level of phenomenology, diagnostic criteria, and cognitive and behavioural concepts that seek to explain the presence and maintenance of the disorder)*

The therapist demonstrated, through statements and/or actions an understanding of depression and how it manifests for individuals in the form of difficulties, symptoms, and impaired functioning.

2. *An evident knowledge of how the disorder 'plays out' in an interpersonal context*

The therapist demonstrated, through statements and/or actions an understanding of the ways in which depression manifests in a couple's relationship.

3. *An evident knowledge of how the patient's depression is affecting the partner and the relationship*

¹⁸ It is not necessary that the therapist states explicitly that one, or both partners is depressed in order to obtain a pass on this item. Working on the depression itself may not be an explicit focus of the session and so the therapist should not be penalised if good work focuses the work in an alternative direction. Therapists should, therefore, be awarded a pass on this item unless something significant is missing or a significant error occurred.

Scale development: BCTS-D

Whether or not an explicit focus of the session, the therapist conveyed a sensitivity to the ways in which the patient's depression might be impacting the partner and, if appropriate, provided guidance on ways to address this.

4. *An understanding (conveyed through statements and/or actions) of what the individual needs to change and maintain in order to respond effectively to the depression*

The therapist was able to use disorder-specific concepts and models to identify and work with relevant individual factors.

5. *An understanding (conveyed through statements and/or actions) of how the partner and couple can assist the depressed individual and contribute to healthy relationship functioning.*

The therapist was able to use disorder-specific concepts and models to guide interventions to assist the individual and contribute to healthy relationship functioning.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist appeared to have no knowledge of depression. The therapist made fundamental errors in the information about depression that was communicated to the couple.</p> <p>OR: The therapist appeared to have no appreciation that an interpersonal context was relevant to the depression.</p> <p>OR: There was a lack of understanding that the patient's depression might be impacting the partner.</p> <p>OR: There was no evidence that the therapist had any appreciation of relevant disorder-specific concepts and interventions.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to consider the patient's depression in a relational context but did so ineffectively and as such, performance was highly limited. This may have taken the form of:</p> <p>Possible misunderstandings of depression at the level of phenomenology, diagnostic criteria or core cognitive and behavioural concepts and interventions.</p> <p>OR: A tendency to focus on the patient such that therapy gave the impression of taking the form of individual therapy with the partner as bystander.</p>

Scale development: BCTS-D

	<p>OR: Limited attempts to consider the partner’s needs.</p> <p>OR: Limited attempts to facilitate the partner’s contribution to the session.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor’s feedback.</p>
2	<p>The therapist attempted to examine features that were relevant to formulating the depression in context, but there were difficulties in how this was conducted. The therapist’s actions showed evidence of emerging competence in this item, but the delivery was problematic.</p> <p>The therapist may have delivered aspects of psychoeducation in a clumsy fashion that led to the couple becoming confused, or tended to lecture the couple.</p> <p>OR: The therapist may have unhelpfully ‘challenged’ the depressed patient’s cognitions, resulting in their feeling potentially disempowered or embarrassed by their ‘faulty thinking’.</p> <p>OR: The therapist inadvertently put the partner in the role of being the depressed patient’s therapist, encouraging degrees of monitoring and care-taking that were inappropriate or that would keep the person in the ‘depressed role.’</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor’s feedback.</p>
3	<p>Overall the therapist demonstrated competence in the ability to identify and focus the session on individual, couple and environmental factors that appeared implicated in the patient’s depression.</p> <p>would be beneficial.</p> <p>In summary, the therapist’s performance was broadly competent on this item. The essential features were present, and only minor inconsistencies or areas of refinement needed.</p>
4	<p>The therapist provided clear evidence of skill in identifying and working with individual, couple, and environmental factors that appeared implicated in the patient’s depression. The five key features identified above were evident.</p> <p>The therapist was clear whether the primary objective was to (1) target the disorder or (2) improve the relationship, with statements and actions reflecting a clear objective in this regard.</p> <p>In summary, the therapist’s performance was competent on this item with only minor areas of refinement needed (as specified in the qualitative feedback).</p>
	<p>The therapist was highly effective on this item, demonstrating an above average level</p>

Scale development: BCTS-D

5	<p>of skilfulness in their understanding of how depression manifested, and was maintained in the context of the couple's relationship. The therapist appeared clear about their primary objective (targeting the disorder or improving the relationship), and their actions were consistent with this objective. The therapist created an environment in which the couple was empowered to unite against the depression. Performance was consistent across the session.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. Nothing inappropriate was done, and nothing important was omitted. It simply was not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist's competence on this item was at 'master'/expert level. As for anchor point 5 (above), the therapist demonstrated an above average level of skilfulness in their understanding of how depression manifested, and was maintained in the context of the couple's relationship. The therapist appeared clear about their primary objective (targeting the disorder or improving the relationship), and their actions are consistent with this objective.</p> <p>The therapist created an environment in which the couple was empowered to unite against the depression with the likely outcome that the couple was well-engaged and able to work effectively towards identifying areas for change. Performance was consistent across the session.</p> <p>Disorder-specific expertise was evident, as was expertise in understanding depression in an interpersonal context.</p> <p>In summary, the therapist demonstrated a superior level of skill on this item, and would be deemed to be working at 'master level'.</p>

Primers for the Assessor

Examples of statements or actions that reflect these indicators include the following:

1. Were there statements and/or actions designed to focus the session on relevant individual, couple, or environmental factors impacting the patient's depression and relationship distress?
2. Did the therapist convey an understanding of the factors commonly implicated in predisposing, triggering, and maintaining depression? Were they able to think in a 'disorder-specific way'?
3. Were misinterpretations (by the patient or the partner) of the depressed person's actions identified and challenged in an appropriate way?
4. Did the therapist provide psychoeducation or 'mini lectures' on the nature of depression to help the couple make sense of the situation? Were handouts used as an aid to psychoeducation – and if so, were these presented appropriately with the couple encouraged to personalise them to their circumstances?
5. Did the therapist make statements and/or deliver interventions aimed at teaching the partner how to interpret the depressed person's behaviour?

Scale development: BCTS-D

6. Did the therapist appear aware of, and able to respond effectively to, unhelpful if well-intentioned ways that the partner may have been 'recruited' into the depression or inappropriately accommodated the depression (e.g., by encouraging the depressed person to rest, or taking over the partner's responsibilities)?
7. Were appropriate interventions used to increase empathic responding to negative statements by the depressed partner?
8. Was suicidal ideation and/or behaviour identified and discussed from both an individual and couple perspective?
9. Did the therapist aim to increase positives or initiate behavioural activation to address depression?
10. If appropriate to the stage of therapy and the needs of the couple, was the couple's physical relationship considered?
11. Were attempts made to consider appropriate social support?

Examples of statements or actions that reflect these indicators include the following:

- We've just looked at some of the common ways in which depression can 'play out' for people – individually and in their relationships. How does this apply to you? (Have you seen depressed partner do this? What's it like for you when this happens – how does it impact?)
- When she says, "I can't take it anymore", I understand that it's really difficult for you to know how to make sense of this - whether this is your partner expressing how she feels in that moment, and she just needs you to listen, or whether this is a decision that she has made to act on her despair, and one which you need to be worried about. Let's spend some time figuring out what (depressed patient) means so we can find a way forward.
- It's really hard for you when (depressed patient) says "No one cares." Let's see if we can find some different ways to respond to those kinds of statements that would work better for both of you.
- For you, depression is very different from other illnesses such as cancer or heart disease, and so it doesn't seem like a real 'problem' at all. It makes sense then, why you would want (partner) to just get over it. Let's look at depression in a bit more detail, so we can understand what's happening for (depressed partner) and what might be more helpful ways to respond.

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 9. Selecting Interventions

Introduction

This item is concerned with the therapist's choice of intervention strategy and its appropriateness for the focus of the session and the couple's presenting difficulties and needs. The selection of an intervention should flow logically from the case conceptualisation, so it will be useful for the Assessor to cross-reference with Item 13. Dyadic Conceptualisation.

It is important to note that this item is not concerned with the effectiveness with which the intervention was *delivered* which should be scored under Items 10-12. Rather, this item seeks to assess the extent to which the therapist's thinking and decision-making concerning the selection of an intervention strategy was informed by appropriate criteria and a sound rationale.

(NB: If the session is a partner-assisted intervention to address the psychopathology experienced by one partner, it would be expected that the therapist conveys an awareness of the 'content specificity hypothesis' and that this guides the therapist's thinking around the use of a specific intervention. For example, if working with a couple where one partner is experiencing depression, it would be expected that the therapist (a) selects interventions that are based on an understanding of the likely benefits of behavioural activation, (b) selects cognitive interventions that are consistent with modifying cognitions and cognitive biases commonly associated with depression (e.g. negative beliefs about the self, world and future), or (c) addresses emotional factors related to the depression such as despair from either partner. Similar rationales are present if the session focuses on the couple's relationship.)

Features of competent performance

Competence performance on this item is evidenced through the following three features:

1. The therapist selects a clearly recognisable BCT intervention

The therapist's actions were consistent with a clearly identifiable intervention that is commonly used in BCT.

2. Congruence between the choice of intervention, dyadic conceptualisation, and stage of therapy

The therapist's choice of intervention was appropriate to the stage of therapy and related clearly to the case conceptualisation of relevant individual, couple, and environmental factors, including sources of primary and secondary distress that may have been identified as relevant to the couple's presentation.

3. The selected strategy was introduced to the couple in a clear and appropriate way

The therapist introduced the selected intervention in a way that was both apparent (that is, the listener could clearly identify that this is what the therapist was doing) and accessible to the couple.

Scale development: BCTS-D

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist did not select any strategy that would be deemed to be a clearly recognisable BCT intervention.</p> <p>OR: The therapist's choice of intervention was entirely incongruous with the dyadic conceptualisation.</p> <p>OR: The therapist made no attempt to introduce the strategy to the couple and the result was actively detrimental to the session.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to select an appropriate intervention but did so ineffectively and as such, performance was highly limited. For example:</p> <p>The therapist seemed indecisive about which intervention to select.</p> <p>OR: The therapist selected one intervention but changed intervention part way through, without any obvious rationale, and this had significant negative effects on the effectiveness of the session.</p> <p>OR: The introduction and overview of the intervention that was necessary for the couple to engage with it was not forthcoming, leaving the couple confused or uncertain as to what was required.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback. Something relative to the item was present, but notably lacking. Significant improvement is needed.</p>
2	<p>The therapist clearly attempted to select an appropriate intervention but there were some limitations in how this was approached. For example:</p> <p>The therapist did not choose the most obviously relevant intervention for the couple's needs, and the selected intervention was significantly lacking in addressing the couple's needs compared to a more obvious intervention.</p> <p>OR: The intervention was potentially useful but not entirely consistent with the stage of therapy.</p> <p>OR: The therapist had difficulty in introducing the intervention to the couple in a way</p>

Scale development: BCTS-D

	<p>that was accessible and meaningful to both partners.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor's feedback. The session was generally well conducted on this item, but something notable was missing, or it went well for most of the session with a deviation at some point, such that it was not fully acceptable, just below passing.</p>
3	<p>The therapist selected an appropriate and relevant intervention, and the intervention selected was a clearly recognisable BCT intervention. For example:</p> <p>The intervention selected was broadly consistent with a dyadic conceptualisation.</p> <p>AND/OR: The intervention selected was broadly appropriate to the stage of therapy.</p> <p>AND: The therapist did a reasonable job of introducing the intervention to the couple.</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist demonstrated evidence of skill in selecting a clearly recognisable and appropriate BCT intervention.</p> <p>AND: The intervention selected was consistent with a dyadic conceptualisation, with the potential to help the couple make positive changes in their relationship.</p> <p>AND: The intervention selected was optimum for the stage of therapy, with the potential to help the couple make positive changes in their relationship.</p> <p>AND: The therapist did a good job of introducing the intervention to the couple.</p> <p>In summary, the therapist's performance was effective on this item with only minor areas of refinement needed. The therapist demonstrated a good, solid level of proficiency overall, with some aspect standing out throughout the session or at some point during the session as above the norm.</p>
5	<p>The therapist demonstrated a very high level of proficiency in selecting a clearly recognisable and appropriate BCT intervention.</p> <p>AND: The intervention selected was entirely consistent with, and flowed logically from a dyadic conceptualisation, with the potential to help the couple make positive changes in their relationship.</p> <p>AND: The intervention selected was optimum for the stage of therapy, with the potential to help the couple make positive changes in their relationship.</p> <p>AND: The therapist introduced the intervention to the couple with considerable skill, personalising it to their circumstances and needs in a way that engaged the couple.</p>

Scale development: BCTS-D

	<p>In summary, the therapist demonstrated an obvious, high degree of skill on this item. Nothing inappropriate was done, and nothing important was omitted. However, performance was not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist demonstrated exceptional proficiency on this item, selecting a clearly recognisable BCT intervention that was likely to be optimal for the couple's needs, given both the dyadic conceptualisation and the stage of therapy.</p> <p>The intervention, and the rationale for its selection, was presented with considerable skill to the couple which was personalised to their circumstances and needs in a way that engaged the couple.</p> <p>All of the features of competent performance were present to an advanced degree.</p> <p>In summary, the therapist demonstrated a superior level of skill on this item, and would be deemed to be working at 'master level'. One might select this for a training session to demonstrate highest levels of performance.</p>

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of performance:

1. Did the session incorporate an identifiable intervention/s?
2. Was the selected intervention one that would be widely recognised as a core BCT intervention (e.g., sharing thoughts and feelings; decision-making guidelines) or standard CBT principles incorporated into a couple context (e.g., exposure)?
3. Was the choice of intervention consistent with the therapist's dyadic conceptualisation? (To ascertain this, cross-reference with Item 13. Dyadic Conceptualisation.)
4. Did the therapist introduce the intervention effectively, such that the couple was appropriately 'primed' for what was to follow?

Examples of statements or actions that reflect these indicators include the following:

- I'd like to talk with you about how we might tackle this issue (as a lead in to talking with the couple about a specific intervention).
- The therapist explicitly introduces what the assessor recognises as a specific, standard CBT intervention.
- Last time we met we talked about (name of intervention), and you agreed that you would practice that for homework. I'd like to hear how you got on with this and to give you the chance to practice it some more today.
- I'd like us to pause a moment. This seems to be a sticking point for you both. Can you think of anything we've been talking about recently, any specific skills you have learnt, that might help you address this issue more productively? (And if the couple struggle to identify the relevant intervention, the therapist makes a suggestion.)

Scale development: BCTS-D

- I think that a more productive way of addressing this question/dilemma/issue would be to use the (name of intervention) we looked at before. How about you try to do that now?

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 10. Emotion-Focused Interventions

Introduction

The therapist needs to facilitate the expression and processing of appropriate levels of emotion by the couple. The ability to deal with negative emotions is a strong predictor of whether couple relationships will last. It is important, therefore, to help couples learn how to understand the emotional realm and manage emotions effectively.

Item 10 addresses the competent use of the emotion-focused intervention/s selected, and the therapist's ability to implement this intervention/s in a way that is appropriate to the overall aims and objectives of the session and the agreed aims of therapy. (NB: The appropriateness of the intervention should be scored under Item 9. Selecting Interventions. Thus, if the assessor questions the appropriateness of the emotion-focused intervention chosen but concludes that this intervention was implemented with a high degree of skill, the therapist would be scored lower on Item 9 but attain a high score on this item.)

Negative emotions are normal and can be adaptive. However, negative emotions that are too frequent or extreme, or problems with affect regulation are likely to interfere with couple functioning outside of the session. Equally, restricted or minimized emotions are likely to negatively impact both the therapy and the couple's relationship.

There are two principal ways in which emotion-focused interventions might be used in a session (both of which are legitimate for the purposes of rating the session). The first is where the therapist introduces a specific emotion-focused intervention based on a prior formulation of individual, couple or environmental factors relevant to the couple's difficulties. An example might be where the therapist seeks to equip one partner with greater skills in emotion regulation, or broaden an individual's range of emotional experience based on a formulation of relevant individual factors. The second is where the therapist works with the couple to attend to the role of affect, and helps the couple connect their emotion to relevant cognitions and behaviours. Competent performance on this item will enable the couple to access and express their emotions in a way that facilitates change.

Even where the therapist is focusing primarily on cognitive or behavioural interventions, it is anticipated that the emotional domain will feature and be positively impacted. *In consequence, if there is evidence of the therapist remaining attentive to emotional factors, this item can still be scored.* What is primary in rating items 10-12 is that the therapist has decided upon a 'point of entry' into the couple's distress (emotion, cognition or behaviour) and whilst working at this point of entry, remains attentive to the other domains. Where a therapist appropriately omits content in one domain (e.g., cognitive, behavioural or emotional interventions), the therapist should not be penalised and would be awarded a default rating of 3 in the omitted area. It is acceptable to omit a particular domain providing that the assessor is confident that this is based on sound decision-making on the part of the therapist.

(NB: The therapist must also be able to deal effectively with affect that arises during the session. However, the therapist's skill in managing problematic emotional states and reactions that arise in the session is scored under Item 4. Facilitating Couple Communication.)

Scale development: BCTS-D

Features of competent performance

Competence performance on this item is evidenced through the following two features:

1. *Facilitation of access to, and expression of, a range of emotions*

The therapist created an emotional climate that enabled each partner to become more aware of, and better able to express, the variety of emotions that were relevant to their individual and relationship needs.

2. *Implementing emotion-focused interventions to guide the couple in responding effectively to one another's emotional worlds and experience*

The therapist drew upon typical, well-recognised emotion-focused interventions (such as sharing thoughts and feelings, practicing healthy compartmentalisation of difficult emotional states, teaching arousal reduction strategies) according to the couple's needs.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist did not in any way attempt to identify, acknowledge or work with relevant emotions in the session.</p> <p>OR: The therapist's actions were discouraging of, or actively prevented, the couple from expressing emotion when it would have been central to the therapeutic process.</p> <p>OR: The therapist appeared entirely oblivious to the emotions that the couple were expressing, and which they were highlighting as relevant to their concerns.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to work with emotion but did so ineffectively and as such, their performance was highly limited. For example:</p> <p>The therapist seemed aware of the need to work with emotions but was unable to do so, and appeared to be at a loss to know where to start.</p> <p>OR: The therapist attempted to work with emotions that were not relevant to the couple at that time.</p> <p>OR: The therapist was unable to help the couple make links between emotions and relevant cognitions and behaviour (despite the couple being well-placed to make such links had the task been conducted competently).</p>

Scale development: BCTS-D

	<p>OR: The therapist inappropriately heightened emotions that then interfered with the effectiveness of the session.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback. Something relevant to the item was present, but notably lacking. Significant improvement is needed.</p>
2	<p>The therapist was attempting to facilitate the expression and processing of appropriate levels of emotion but problems were present. For example:</p> <p>The therapist may have begun well in working with relevant emotions, but at some point there was a significant deviation such that the session went off track.</p> <p>OR: The therapist was not able to convey a good explanation of the role of emotions such that the couple could not understand what the therapist was aiming to achieve.</p> <p>OR: The therapist missed a significant number of opportunities to introduce emotion-focused interventions or other work with emotions in the session.</p> <p>OR: The therapist attempted to either heighten or contain emotional expression at an appropriate time but the attempts lacked in some way (e.g., gave up too easily, therapist's tone was ineffective, etc.).</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence. Something notable was missing, or there was a deviation from good performance at some point such that it was not fully acceptable, just below passing. Key areas for development are identified in the assessor's feedback.</p>
3	<p>The therapist was clearly aware of the need to target emotions and attentive to opportunities to do so. Overall, this was performed in a competent fashion. For example:</p> <p>The therapist was clearly attentive to emotions and the emotional realm during and the session.</p> <p>AND/OR: Overall, the therapist selected and worked with the emotions that were most relevant to the couple at that stage in therapy.</p> <p>AND/OR: The therapist was using emotion-focused interventions that were consistent, in some way, with the agreed aims of therapy.</p> <p>AND/OR: The therapist drew upon well-recognised BCT emotion-focused interventions according to the couple's needs.</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>

Scale development: BCTS-D

<p>4</p>	<p>The therapist was clearly aware of the need to target relevant emotions and did so effectively. The therapist's questions, statements and actions enabled the couple to attain a new understanding of their needs (or, in the context of a challenging session, can be judged as likely to have done so if the session had been more straightforward). For example:</p> <p>The therapist was able to help the couple identify and name relevant emotions and did so in an effective manner.</p> <p>AND/OR: The therapist was consistent in seeking opportunities to work with relevant emotions throughout the session.</p> <p>AND/OR: The therapist's statements and actions were effective in facilitating shifts for the couple in terms of their understanding of emotional factors and emotional expression and their role in couple functioning.</p> <p>AND/OR: The therapist's statement and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to express, the variety of emotions that were relevant to their individual and relationship needs.</p> <p>In summary, the therapist's performance was competent on this item with only minor areas of refinement needed. There were no concerns about the therapist's performance.</p>
<p>5</p>	<p>The therapist's statements and actions were clearly guided by an attentiveness to the emotional realm and an intention to support the couple in managing emotions more effectively. The therapist's performance was highly skilled and likely to have been demonstrably helpful (assuming the absence of factors that made the session a very challenging one), enabling the couple to better understand the relevance of emotions to their relationship functioning, and helping them acquire new emotion-focused skills. For example:</p> <p>The therapist was highly attentive to opportunities for working with emotions in the session.</p> <p>AND/OR: Emotion-focused interventions were delivered with a very high level of skill.</p> <p>AND/OR: The therapist enabled the couple to make progress in identifying and working with their emotional experience (even where difficult and emotive issues were discussed).</p> <p>AND/OR: The therapist's statements and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to express, the variety of emotions that were relevant to their individual and relationship needs. A sense of safety around the expression of emotions was established.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. The therapist's actions were entirely appropriate and nothing important was</p>

Scale development: BCTS-D

	<p>omitted. The performance was simply not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist demonstrated a highly advanced level of ability to work with emotions and deliver emotion-focused interventions. The therapist's statements, questions and actions were consistently facilitative for the couple in making new links between emotions, cognitions and behaviours, and expressing emotions in adaptive ways. For example:</p> <p>The therapist was highly, and consistently, attentive to opportunities for working with emotions in the session, helping the couple make connections between emotions and their relationship functioning of which they had not formerly been aware.</p> <p>AND/OR: Emotion-focused interventions were delivered with an exceptionally high level of skill.</p> <p>AND/OR: The therapist enabled the couple to make progress in identifying and working with their emotional experience (even where difficult and emotive issues were discussed).</p> <p>AND/OR: The therapist's statements and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to express, the variety of emotions that were relevant to their individual and relationship needs. A sense of safety around the expression of emotions was established.</p> <p>In summary, the therapist demonstrated an obvious, exceptionally high degree of skill on this item, and would be deemed to be working at 'master level'. The therapist's actions were entirely appropriate and nothing important was omitted. A score of 6 on this item is likely to be rare and indicates a level of skill that might be used to demonstrate the highest levels of performance in a training session.</p>

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of performance:

1. Did the therapist attempt to offer guidelines relating to how partners can listen to one another?
2. Did the therapist offer guidance on how to respond when a partner finishes speaking (use of summarising, etc.)?
3. Did the therapist provide appropriate modelling of listening skills?
4. Did the therapist attempt to amplify tender emotions?
5. Did the therapist ask each partner to reflect on their emotional experience?
6. Did the therapist provide psychoeducation and normalisation around emotional experience, as appropriate?

Scale development: BCTS-D

7. Did the therapist guide the couple through arousal reduction strategies if the conceptualisation suggested that one partner needed help with containing negative emotions?
8. Did the therapist work with the couple to put in place methods that could create a sense of safety around challenging emotions, such as practicing healthy compartmentalisation, seeking alternative means to communicate feelings and eliciting support or scheduling times to discuss emotions with one another?
9. Did the therapist work with the couple to help create meaningful links between emotions, and relevant cognitions and behaviours?
10. If necessary (for example, in order to reduce the level of affect and engage with the task in hand) did the therapist connect the couple's experience to the couple's session process and/or broader formulation?
11. Did the therapist describe emotions through metaphors or images?

Examples of statements or actions that reflect these indicators include the following:

- I'd like you to state your view subjectively, not as a statement of fact.
- I'd like you to express your emotions, not just your ideas.
- I would like you to show, through your facial expressions and your body language, that you accept your partner's thoughts and feelings (even if you don't agree with his/her perspective).
- Can you summarise what your partner said just then, making sure you include what you understood to be to them the most important feelings, desires, conflicts and thoughts.
- What are you hearing from your partner – what is he/she telling you?
- How would you feel towards your partner if you could achieve that?
- Let's look at some alternative ways of getting some emotional support (leading to a discussion about this).
- Let's look at ways you can practice healthy compartmentalisation (leading to a discussion about this).

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 11. Cognitive Interventions

Introduction

BCT accords a central role to cognitions and cognitive factors in the origins and maintenance of psychological difficulty, and in the development and perpetuation of couple distress. BCT does, therefore, typically draw upon a range of cognitive interventions to enable change in this domain. It is important that the therapist is able to work effectively with specific cognitions and/or patterns of information processing that are problematic, introduce change methods to modify them, and promote more adaptive alternatives. Item 11 addresses the competent use of the cognitive intervention/s selected, and the therapist's ability to implement this intervention/s in a way that is appropriate to the overall aims and objectives of the session and the agreed aims of therapy.

There are a variety of cognitive factors to which the therapist may seek to draw a couple's attention and change efforts including:

- Selective Attention (that is, what each notices about their partner, the relationship, and the environment);
- Attributions (that is, causal and responsibility explanations for couple-related events that have occurred);
- Expectancies (that is, predictions of what will occur in the relationship in the future);
- Assumptions (that is, what each partner believes about the fundamental nature of people and relationships);
- Relationship standards (that is, what each partner believes that people and relationships should be like).

Accordingly, there are a variety of interventions that a therapist may use such as:

- Evaluating the experiences and logic supporting a cognition;
- Weighing advantages and disadvantages of a cognition;
- Considering worst and best possible outcomes of situations;
- Providing educational mini-lectures, readings, and tapes.

If the therapist is using a partner-assisted intervention to address psychopathology experienced by one partner, it would be expected that the therapist conveys an awareness of any well-documented cognitive and behavioural 'profiles' specific to that particular disorder, and that this guides the therapist's thinking around the use of a specific cognitive intervention. For example, if working with a couple where one partner is experiencing depression, it would be expected that the therapist implements interventions that are consistent with modifying cognitions and cognitive biases commonly associated with depression (e.g., negative beliefs about the self, world, and future).

It is important to note that Item 11 is concerned exclusively with the degree of skill with which the therapist implemented their chosen cognitive intervention. The appropriateness of the intervention should be scored under Item 9. Selecting Interventions. Thus, if the assessor questions the appropriateness of the cognitive intervention chosen but concludes that this intervention was implemented with a high degree of skill, the therapist would be scored lower on Item 9 but attain a high score on this item.

Scale development: BCTS-D

NB: Even where the therapist is focusing primarily on emotion-focused or behavioural interventions, it is anticipated that the behavioural domain will be positively impacted. *In consequence, if there is evidence of the therapist remaining attentive to behavioural factors, this item can still be scored.* What is primary in rating items 10-12 is that the therapist has decided upon a 'point of entry' into the couple's distress (emotion, cognition or behaviour) and whilst working at this point of entry, remains attentive to the other domains. Where a therapist appropriately omits content in one domain (e.g., cognitive, behavioural or emotional interventions), the therapist should not be penalised and would be awarded a default rating of 3 in the omitted area. It is acceptable to omit a particular domain providing that the assessor is confident that this is based on sound decision-making on the part of the therapist.

Features of competent performance

Competence performance on this item is evidenced through the following three features:

1. *Identification of a key cognition/s and/or facet of cognitive processing requiring modification*

The therapist identified and targeted a specific cognition or information processing bias that was appropriate to the couple's needs and the aims of therapy. Given the presence of both partners, the therapist might focus on differences in understanding or experience between the two partners rather than focusing on one person's cognitions per se which would be more typical in individual CBT.

2. *Introducing the cognitive intervention to the couple*

The therapist drew upon a well-recognised cognitive intervention/s to address the couple's needs and introduced this in a way that was clear and accessible. Whereas this could involve a direct focus on the cognitions, the therapist might also ask the couple to engage in some experience that leads to greater understanding or changes in cognition (e.g., asking them to share their thoughts and feelings so they can better understand how each person experienced a situation which could lead to altered attributions).

3. *Guided practice in the session*

The therapist guided the couple effectively in the application of this intervention.

Scoring and anchor points

0	<p>The therapist did not in any way attempt to identify, acknowledge or work with relevant cognitions and/or cognitive factors in the session.</p> <p>OR: The therapist's actions were discouraging of, or actively prevented, the couple from sharing potentially relevant cognitions and/or cognitive factors.</p>
---	--

Scale development: BCTS-D

	<p>OR: The therapist appeared entirely oblivious to the cognitions and cognitive factors that the couple were sharing, and which they were highlighting as relevant to their concerns.</p> <p>In summary, there were significant concerns about the therapist’s performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to work with cognitions and/or cognitive factors but did so ineffectively and as such, their performance was highly limited. For example:</p> <p>The therapist seemed aware of the need to work with cognitions and cognitive factors but was unable to do so, and appeared to be at a loss to know where to start.</p> <p>OR: The therapist attempted to work with cognitions and cognitive factors that were not relevant to the couple at that time.</p> <p>OR: The therapist attempted to work with relevant cognitions and cognitive factors but did so in an ineffective way (e.g., confrontational, debating with the couple).</p> <p>OR: The therapist was unable to help the couple make links between cognitions and relevant emotions and behaviours (despite the couple being well-placed to make such links had the task been conducted competently).</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor’s feedback. Something relevant to the item was present, but notably lacking. Significant improvement is needed.</p>
2	<p>The therapist was attempting to work relevant cognitions and cognitive factors but problems were present. For example:</p> <p>The therapist may have begun well in working with relevant cognitions and cognitive factors, but at some point there was a significant deviation such that the session went off track.</p> <p>OR: The therapist was not able to convey a good explanation of the role of cognitions and cognitive factors such that the couple could not understand what the therapist was aiming to achieve.</p> <p>OR: The therapist missed a significant number of opportunities to introduce cognitively-focused interventions in the session.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence. Something notable was missing, or there was a deviation from good performance at some point such that it was not fully acceptable, just below passing. Key areas for development are identified in the assessor’s feedback.</p>
	<p>The therapist was clearly aware of the need to target cognitions and other cognitive</p>

Scale development: BCTS-D

3	<p>factors and was attentive to opportunities to do so. Overall, this was performed in a competent fashion. For example:</p> <p>The therapist was clearly attentive to cognitions and other relevant cognitive factors during and the session.</p> <p>AND/OR: Overall, the therapist selected and worked with the cognitions and cognitive factors that were most relevant to the couple at that stage in therapy.</p> <p>AND/OR: The therapist was using cognitive interventions that were consistent, in some way, with the agreed aims of therapy.</p> <p>AND/OR: The therapist drew upon well-recognised BCT cognitively-focused interventions according to the couple's needs.</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist was clearly aware of the need to target relevant cognitions and cognitive factors and did so effectively. The therapist's questions, statements and actions enabled the couple to attain a new understanding of their needs (or, in the context of a challenging session, can be judged as likely to have done so if the session had been more straightforward). For example:</p> <p>The therapist was able to help the couple identify and name relevant cognitions and cognitive factors and did so in an effective manner.</p> <p>AND/OR: The therapist was consistent in seeking opportunities to work with relevant cognitions and cognitive factors throughout the session.</p> <p>AND/OR: The therapist's statements and actions were effective in facilitating shifts for the couple in terms of their understanding of cognitions and other cognitive factors and their role in couple functioning.</p> <p>AND/OR: The therapist's statement and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to express, the variety of cognitions and cognitive factors that were relevant to their individual and relationship needs.</p> <p>In summary, the therapist's performance was competent on this item with only minor areas of refinement needed. There were no concerns about the therapist's performance.</p>
5	<p>The therapist's statements and actions were clearly guided by an attentiveness to the cognitive realm and an intention to support the couple in modifying relevant cognitive factors to create meaningful change in their relationship. The therapist's performance</p>

Scale development: BCTS-D

	<p>was highly skilled and likely to have been demonstrably helpful (assuming the absence of factors that made the session a very challenging one), enabling the couple to better understand the relevance of cognitions and cognitive factors to their relationship functioning, and helping them acquire new cognitive skills. For example:</p> <p>The therapist was highly attentive to opportunities for working with cognitions and cognitive factors in the session.</p> <p>AND/OR: Cognitive interventions were delivered with a very high level of skill.</p> <p>AND/OR: The therapist enabled the couple to make progress in identifying and working with their cognitions and other relevant cognitive factors (even where difficult and emotive issues were discussed).</p> <p>AND/OR: The therapist's statements and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to modify, the variety of cognitions and cognitive factors that were relevant to their individual and relationship needs. A sense of safety around the revealing of cognitions and cognitive factors was established.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. The therapist's actions were entirely appropriate and nothing important was omitted. The performance was simply not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist demonstrated a highly advanced level of ability to work with cognitions and cognitive factors, and to deliver cognitive interventions. The therapist's statements, questions and actions were consistently facilitative for the couple in making new links between cognitions, emotions and behaviours, and helping each partner modify potentially problematic cognitions. For example:</p> <p>The therapist was highly, and consistently, attentive to opportunities for working with cognitions and cognitive factors in the session, helping the couple make connections between cognitions and their relationship functioning of which they had not formerly been aware.</p> <p>AND/OR: Cognitive interventions were delivered with an exceptionally high level of skill.</p> <p>AND/OR: The therapist enabled the couple to make progress in identifying and working with relevant cognitions and/or cognitive factors (even where difficult and emotive issues were discussed).</p> <p>AND/OR: The therapist's statements and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to express, the variety of cognitions and cognitive factors that were relevant to their individual and relationship needs. A strong sense of safety around the revealing of cognitions and cognitive factors was established.</p>

Scale development: BCTS-D

	<p>In summary, the therapist demonstrated an obvious, exceptionally high degree of skill on this item, and would be deemed to be working at 'master level'. The therapist's actions were entirely appropriate and nothing important was omitted. A score of 6 on this item is likely to be rare and indicates a level of skill that might be used to demonstrate the highest levels of performance in a training session.</p>
--	---

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of effective performance:

1. Was the therapist able to identify and elicit those cognitions and information-processing biases that seemed implicated in the couple's difficulties?
2. Were the cognitions and any associated biases elicited well?
3. Was the therapist able to access and work with those cognitions and biases that seemed implicated in the couple's distress and/or the nature of the presenting problems?
4. Was any specific cognitive intervention used during the session introduced and implemented delivered effectively?
5. Did the therapist help the couple problem-solve how any new standards would be taken into account behaviourally?

Examples of statements and/or actions that reflect these indicators include the following:

- When you get distressed what you will notice is... (reference to changes in the way information is processed).
- When one person notices another behave in a certain way, it's natural to try to figure out why he or she is acting that way. For example... (then the therapist identifies an example that is specific to the couple).
- What thoughts went through your mind when...?
- It's possible that your understanding of (partner's) behaviour was absolutely correct, but it's also possible that (partner) reacted this way for different reasons. Understanding what was driving (partner's) response is important because it will affect how you react to (partner).
- People have beliefs about what good relationships are like. Let's look at this in terms of your own relationship.
- Let's look at each of your beliefs about what relationships should be like.
- Let's look at the pros and cons of each of your standards in relation to this issue.... I'd like each of you to think about the possible advantages of your partner's standard/rule...What might be the disadvantages of your own standard/rule?
- Give me some ideas as to how we might bring those two perspectives together.
- Is there a way we can moderate that rule/standard, etc.?
- Let's see if there are enough areas of overlap to make this work.
- What might be some of the worst and best possible outcomes here?
- Let's step back and look at this more broadly.

Scale development: BCTS-D

- Is there any other possible interpretation of/explanation for what happened?
- 'Mini lectures' on some aspect of information processing such as perceptual or interpretive biases, attributions, assumptions or standards.

Item 12. Behavioural Interventions

Introduction

The extent of relationship satisfaction experienced by each partner will be heavily influenced by how the couple is functioning: that is, their behavior. Couples often begin their 'stories' with complaints or concerns at the behavioural level, and in BCT, behavior is an important focus of change in its own right, not solely employed for the purposes of bringing about a cognitive change (e.g., to test out and revise a catastrophic cognition).

Factors in a couple's relationship that warrant behavioural intervention can take numerous forms such as withdrawal, avoidance, compulsions, and various types of safety-seeking behaviours, in addition to punishing behaviours (e.g., acts that convey criticality, hostility, or undermining). Additionally, the couple may engage in acts of support, affection, and caring which represent sources of strength, resilience and adaptive coping upon which the therapist may wish to build, or there may be other types of positive behaviour that the therapist wishes to introduce.

It is important, therefore, that the therapist is able to work effectively with specific behaviours that are problematic, introduce change methods to modify them, and seek opportunities to increase rewarding behaviours that might enrich the relationship and promote longer-term relationship satisfaction.

Item 12 addresses the competent use of any behavioural intervention/s selected, and the therapist's ability to implement this intervention/s in a way that is appropriate to the overall aims and objectives of the session and the agreed aims of therapy. The appropriateness of the intervention should be scored under Item 9. Selecting Interventions. Thus, if the assessor questions the appropriateness of the behavioural intervention chosen but concludes that this intervention was implemented with a high degree of skill, the therapist would be scored lower on Item 9 but attain a high score on this item.

Interventions scored under Item 12 may include those relating to:

- Positive and negative behaviours
- Instrumental and expressive behaviours
- Behaviours directed towards the self, the partner, the relationship or the environment
- Communication and other behavioural interactions

The behavioural interventions used are also likely to stem from one of two broad categories:

- Guided behaviour change: that is, those interventions that do not involve a skill component (e.g., caring days, behaviour exchange);
- Skills-based interventions: that is, interventions that involve teaching skills to couples or providing them with tools (e.g., communication skills).

NB: Even where the therapist is focusing primarily on emotion-focused or cognitive interventions, it is anticipated that the behavioural domain will be positively impacted. *In consequence, if there is evidence of the therapist remaining attentive to behavioural factors, this item can still be scored.* What is primary in rating items 10-12 is that the therapist has decided upon a 'point of entry' into the couple's distress (emotion, cognition or behaviour) and whilst working at this point of entry, remains attentive to the other domains. Where a therapist appropriately omits content in one domain (e.g., cognitive, behavioural or emotional

Scale development: BCTS-D

interventions), the therapist should not be penalised and would be awarded a default rating of 3 in the omitted area. It is acceptable to omit a particular domain providing that the assessor is confident that this is based on sound decision-making on the part of the therapist.

Features of competent performance

Competence performance on this item is evidenced through the following three features:

1. *Identification of a problematic behaviour requiring modification OR a positive behaviour that is to be strengthened*

The therapist identified with the couple a specific target behaviour and outlined the rationale for working on this behaviour to promote change.

2. *Introducing the behavioural intervention to the couple*

The therapist drew upon a well-recognised behavioural intervention/s (such as guided behaviour change, or a skills-based intervention) and introduced this to the couple in a way that was clear and accessible.

3. *Guided practice in the session*

The therapist guided the couple in the application of this intervention. For example, in the case of a skills-based intervention, the therapist facilitated practice in the session and provided appropriate guidance. In the case of guided behaviour change, the therapist worked with the couple in a thorough, thoughtful way, to generate ideas and clarify a task that was optimum for the couple given their stage of therapy.

Scoring and anchor points

0	<p>The therapist did not in any way attempt to identify, acknowledge or work with relevant behavioural factors in the session.</p> <p>OR: The therapist's actions were discouraging of, or actively prevented, the couple from sharing information about relevant behavioural factors.</p> <p>OR: The therapist appeared entirely oblivious to the behavioural factors that the couple were sharing, and which they were highlighting as relevant to their concerns.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to work with relevant behaviours but did so ineffectively and as such, their performance was highly limited. For example:</p> <p>The therapist seemed aware of the need to work with behavioural factors but was</p>

Scale development: BCTS-D

	<p>unable to do so, and appeared to be at a loss to know where to start.</p> <p>OR: The therapist attempted to work with behaviours that were not relevant to the couple at that time.</p> <p>OR: The therapist was unable to help the couple make links between behaviours and relevant cognitions and emotions (despite the couple being well-placed to make such links had the task been conducted competently).</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback. Something relevant to the item was present, but notably lacking. Significant improvement is needed.</p>
2	<p>The therapist was attempting to work relevant behaviours but problems were present. For example:</p> <p>The therapist may have begun well in working with relevant behaviours, but at some point there was a significant deviation such that the session went off track.</p> <p>OR: The therapist was not able to convey a good explanation of the role of behaviours such that the couple could not understand what the therapist was aiming to achieve.</p> <p>OR: The therapist missed a significant number of opportunities to introduce behavioural interventions in the session.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence. Something notable was missing, or there was a deviation from good performance at some point such that it was not fully acceptable, just below passing. Key areas for development are identified in the assessor's feedback.</p>
3	<p>The therapist was clearly aware of the need to target behavioural factors and was attentive to opportunities to do so. Overall, this was performed in a competent fashion. For example:</p> <p>The therapist was clearly attentive to behaviours during the session.</p> <p>AND/OR: Overall, the therapist selected and worked with the behaviours that were most relevant to the couple at that stage in therapy.</p> <p>AND/OR: The therapist was using behavioural interventions that were consistent, in some way, with the agreed aims of therapy.</p> <p>AND/OR: The therapist drew upon well-recognised BCT behavioural interventions according to the couple's needs.</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>

Scale development: BCTS-D

4	<p>The therapist was clearly aware of the need to target relevant behaviours and did so effectively. The therapist’s questions, statements and actions enabled the couple to attain a new understanding of their needs (or, in the context of a challenging session, can be judged as likely to have done so if the session had been more straightforward). For example:</p> <p>The therapist was able to help the couple identify relevant behaviours and did so in an effective manner.</p> <p>AND/OR: The therapist was consistent in seeking opportunities to work with relevant behaviours throughout the session.</p> <p>AND/OR: The therapist’s statements and actions were effective in facilitating shifts for the couple in terms of their understanding of behavioural factors and their role in couple functioning.</p> <p>AND/OR: The therapist’s statement and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to reveal behaviours that were relevant to their individual and relationship needs.</p> <p>In summary, the therapist’s performance was competent on this item with only minor areas of refinement needed. There were no concerns about the therapist’s performance.</p>
5	<p>The therapist’s statements and actions were clearly guided by an attentiveness to the behavioural domain and an intention to support the couple in making behavioural changes that were relevant to their relationship. The therapist’s performance was highly skilled and likely to have been demonstrably helpful (assuming the absence of factors that made the session a very challenging one), enabling the couple to better understand the relevance of behavioural factors to their relationship functioning, and helping them acquire new skills. For example:</p> <p>The therapist was highly attentive to opportunities for working with behaviours in the session.</p> <p>AND/OR: Behavioural interventions were delivered with a very high level of skill.</p> <p>AND/OR: The therapist enabled the couple to make progress in identifying and working with their behaviours (even where difficult and emotive issues were discussed).</p> <p>AND/OR: The therapist’s statements and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to modify, the behavioural repertoires that were relevant to their individual and relationship needs. A sense of safety around the revealing of behaviours was established.</p>

Scale development: BCTS-D

	<p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. The therapist's actions were entirely appropriate and nothing important was omitted. The performance was simply not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist demonstrated a highly advanced level of ability to work with behaviours and to deliver behavioural interventions. The therapist's statements, questions and actions were consistently facilitative for the couple in making new links between behaviours, cognitions and emotions, and helping each partner modify potentially problematic behaviours. For example:</p> <p>The therapist was highly, and consistently, attentive to opportunities for working with behavioural factors in the session, helping the couple make connections between specific behaviours and their relationship functioning of which they had not formerly been aware.</p> <p>AND/OR: Behavioural interventions were delivered with an exceptionally high level of skill.</p> <p>AND/OR: The therapist enabled the couple to make progress in identifying and working with relevant behaviours (even where difficult and emotive issues were discussed).</p> <p>AND/OR: The therapist's statements and actions fostered an emotional climate that enabled each partner to become more aware of, and better able to describe and modify, those behaviours that were relevant to their individual and relationship needs. A strong sense of safety around the revealing of potentially problematic behaviours was established.</p> <p>In summary, the therapist demonstrated an obvious, exceptionally high degree of skill on this item, and would be deemed to be working at 'master level'. The therapist's actions were entirely appropriate and nothing important was omitted. A score of 6 on this item is likely to be rare and indicates a level of skill that might be used to demonstrate the highest levels of performance in a training session.</p>

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of effective performance:

1. Did the therapist describe the importance of behavioural factors in promoting change and use the conversation to introduce a particular intervention to address a target behaviour?
2. Did the therapist work with the couple using a well-recognised behavioural intervention (such as guided behaviour change or a skills-based intervention)?

Scale development: BCTS-D

3. Was it obvious to you what the therapist was aiming to achieve through use of a specific behavioural intervention?
4. Was the intervention described clearly so that the couple was enabled to understand and begin practicing it?
5. Did the therapist provide appropriate guidance to the couple on how to implement the intervention?
6. Did the therapist provide an opportunity for the couple to practice the intervention in the session, where appropriate and feasible to do so?
7. Did the therapist steer the couple's efforts at implementing the behavioural intervention appropriately, interrupting if necessary in order to achieve a productive experience for the couple?

Examples of statements or conversations that reflect these indicators include the following:

- If (...) were as you wanted it to be, how would that look? Can you describe this to your partner? Let's think of a way that we might be able to bring more of this (desired outcome) about.
- You said you wanted to feel loved, appreciated and wanted – how would that look with your partner?
- What would she/he be like if she/he were being a good partner/spouse?
- You don't necessarily have to think of grand gestures; sometimes just the small, easy to attain things are what make someone feel cared for/loved. Let's see if we can generate some ideas about how this might look for the two of you.
- I'd like to teach you a way of approaching this differently... (with the therapist then leading into the behavioural intervention).
- Can you tell each other more about what you need in this situation?
- The therapist provides examples of psychoeducation, such as information about the differences between instrumental and expressive behaviours, and uses these as a basis for introducing a behavioural intervention (such as guided behaviour change).
- The therapist provides examples of 'mini-lectures' on components that are central to the couple's needs.

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 13. Dyadic Conceptualisation

Introduction

BCT requires that therapists' interventions are supported by an underpinning dyadic case conceptualization. Item 13 is concerned with the extent to which the therapist demonstrated skill in helping the couple develop a greater understanding of their difficulties in dyadic terms.

Many couples enter therapy exhibiting specific, and well-documented patterns of negative interaction, and it is essential, therefore, that the therapist knows how to identify and conceptualise their impact as a precursor to intervening. Additionally, partners will typically view the other member of the couple as the source of the problem, and the therapist needs to work to help the couple see maladaptive patterns of interaction as the target problem they can work together to alter.

There are two main forms of dyadic problems which are likely to feature in the therapist's conceptualisations. These are: (1) primary distress: where the couple is unable to resolve issues that stem from personal differences and/or similarities and (2) secondary distress: the distress resulting from the interaction patterns that a couple has developed to cope with unresolved issues in their relationship and which have become problematic in their own right.

Primary distress refers to the subjective distress experienced by one or both partners when existing patterns of behavior fail to meet partners' needs, desires and preferences. These can reflect differences (or similarity) in individual factors – for example, in affective and cognitive styles, the capacity for relationally schematic processing and differences in relationally-oriented and individually-oriented motives (e.g. where one individual seeks a high degree of autonomy in a relationship and the other seeks a high degree of intimacy).

Secondary distress develops from the specific ways in which members of a couple interact behaviourally in response to issue that are sources of their primary distress. Here, four main interaction patterns are likely to be of relevance of which the therapist needs to be aware:

- *Mutual attack*: where partners engage in reciprocal and often escalating aggressive acts such as criticism, threats and other forms of hostility.
- *Demand-withdrawal*: where one partner pursues the other in an attempt to secure attention or compliance (often in an aggressive way) while the other partner withdraws.
- *Mutual withdrawal*: Both partners engage in distancing themselves from one another to avoid interactions that are experienced as aversive.
- *Unilateral or mutual disengagement*: One or both partners have ceased to be invested in the relationship such that any responses from the other partner have little impact (conveying for example, a sense of no longer caring).

Additionally, the therapist's conceptualisation is likely to include hypotheses relating to micro-level patterns (that is, specific interactions occurring in a particular context) and macro-level patterns (that is, broad themes and patterns of interaction that occur across a variety of situations) as well as the extent to which individuals can engage relationally schematic processing.

Item 13 addresses the therapist's competence in helping the couple view and address their problems with a dyadic perspective. This is likely facilitated by the therapist's underlying BCT-

Scale development: BCTS-D

consistent case conceptualization, but this more comprehensive conceptualization may not be explicitly apparent in each session. Thus, this item is more focused on the therapist's competence in facilitating the couple's dyadic view on their problems rather than inferring what the therapist's overall conceptualization of the case may be.

Features of competent performance

Competent performance on dyadic conceptualisation is evidenced through three primary features:

1. *The therapist helped the couple to think about their difficulties and needs in dyadic terms*

Through their statements, it was evident that the therapist was attempting to help the couple develop a dyadic view of their problems (i.e., seeing the interaction pattern as the target of change as opposed to each partner engaging in blame and arguing for their own perspective). This may, at times, include explicit psychoeducation but this is not necessary at all times. For example, therapists may also help the couple to develop this view by asking questions and using reflections that highlight the dyadic nature of the issue.

2. *The appropriateness of the pattern of maladaptive communication selected for discussion in the session*

The therapist selected a pattern of maladaptive communication for discussion in the session that was clearly relevant to the couple's difficulties and so represented a logical and appropriate target for discussion and change.

3. *Use of the dyadic conceptualisation to guide the selection of further intervention strategies*

Where appropriate, the therapist used the dyadic conceptualisation as a platform for selecting further interventions, or assigning specific homework tasks.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist did not attempt, in any way, to support the couple in thinking about their difficulties in dyadic terms.</p> <p>OR: The therapist appeared intent on helping the couple decide which partner was right and which partner was wrong.</p> <p>OR: The therapist explicitly sided with one partner, passing judgement as to whose view was right and whose view was wrong.</p> <p>OR: The therapist appeared oblivious to marked patterns of maladaptive interpersonal communication taking place in the session (e.g., letting negative</p>
---	---

Scale development: BCTS-D

	<p>reciprocity continue unchecked).</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to work with a dyadic conceptualisation but did so ineffectively and as such, their performance was highly limited. For example:</p> <p>The therapist seemed aware of the need to work with patterns of interaction but was unable to do so, and appeared to be at a loss to know where to start.</p> <p>OR: The therapist may have tried to draw the couple's attention to patterns of dyadic communication but quickly got side-tracked or 'railroaded' by the couple's style of interaction in the session.</p> <p>OR: The therapist was not able to identify questions or offer suggestions that could help the couple start to understand their difficulties in dyadic terms.</p> <p>OR: The therapist misunderstood an obvious dyadic interaction pattern, possibly because the therapist did not elicit sufficient information to understand the pattern and rather jumped to premature conclusions about what the pattern is.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback. Although something relative to the item was present, there were notable deficiencies in performance with significant improvement needed.</p>
2	<p>The therapist was conceptualising, and attempting to help the couple conceptualise, the key difficulties in dyadic terms, but problems were present. For example:</p> <p>The therapist may have begun well, offering psychoeducation on patterns of interaction relevant to the couple, but at some point there was a significant deviation such that the session went off track.</p> <p>OR: The therapist was not able to convey a good explanation of dyadic conceptualisation such that the couple could not understand what the therapist was aiming to achieve.</p> <p>OR: The therapist introduced the notion of dyadic conceptualisation but missed a significant number of opportunities in the session for helping the couple understand and consider its implications for improving their relationship.</p> <p>OR: The therapist attempted to provide psychoeducation on relevant patterns of interaction but this was left at too abstract a level for the couple to understand how it applied to their particular circumstances.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development</p>

Scale development: BCTS-D

	identified in the assessor's feedback.
3	<p>The therapist was clearly conceptualising, and attempting to help the couple conceptualise, the key difficulties in dyadic terms, and was able to do so in a competent fashion. For example:</p> <p>The therapist found a way to introduce the idea of dyadic conceptualisation, using language and terminology that was accessible and relevant to the couple.</p> <p>AND/OR: Without formal psychoeducation, the therapist was able to help the couple develop a dyadic perspective on an issue (e.g., through a series of targeted questions and reflections).</p> <p>AND/OR: The therapist used psychoeducation appropriately to introduce patterns of maladaptive interaction relevant to the couple.</p> <p>AND/OR: The therapist was able to initiate discussion about the implications of the relevant patterns identified for making positive change.</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist was clearly conceptualising, and attempting to help the couple conceptualise, the key difficulties in dyadic terms, and was able to do so in a competent and effective fashion, supporting the couple in attaining a new understanding of their needs. For example:</p> <p>The therapist was able to help the couple relate well-documented patterns of maladaptive interaction to their personal circumstances, facilitating shifts in understanding for the couple.</p> <p>AND/OR: Psychoeducation was provided in a way that clearly engaged and felt meaningful for the couple, facilitating further discussion in the session.</p> <p>AND/OR: The therapist was able to work effectively with examples of maladaptive patterns of interaction occurring in the session, drawing attention to this being "...an example of what we were just talking about".</p> <p>In summary, the therapist's performance was competent and effective on this item with only minor areas of refinement needed. There were no concerns about the therapist's performance and feedback is in the service of refining the therapist's performance.</p>
5	<p>The therapist was clearly guided by a dyadic conceptualisation which evidently formed the basis for their statements and actions in the session. The therapist's performance was demonstrably helpful and enabled the couple to make progress in changing patterns of maladaptive interaction (even where difficult and emotive issues were discussed).</p>

Scale development: BCTS-D

	<p>AND/OR: The therapist was able to work very effectively with the couple, using psychoeducation and examples, to help them reframe their difficulties in dyadic terms.</p> <p>AND/OR: The therapist was able to use this conceptualisation to help the couple 'take a step back' from their usual maladaptive patterns of interacting to consider options that might create more effective interpersonal relating.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item which was very well conducted. There were no inappropriate steps taken and nothing important was omitted. It was not, however, at the exemplary master level which is required for a score of 6.</p>
6	<p>The therapist was clearly guided by a dyadic conceptualisation which evidently formed the basis for their statements and actions in the session. The therapist's performance was demonstrably helpful and enabled the couple to make progress in changing patterns of maladaptive interaction (even where difficult and emotive issues were discussed).</p> <p>AND/OR: The therapist was able to work very effectively with the couple, using psychoeducation and examples, to help them reframe their difficulties in dyadic terms.</p> <p>AND/OR: The therapist was able to use this conceptualisation to help the couple 'take a step back' from their usual maladaptive patterns of interacting to consider options that might create more effective interpersonal relating.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item which was conducted exceptionally well. A score of 6 on this item is likely to be rare and signifies practice at 'master level'. It indicates a level of skill that might be used to demonstrate the highest levels of performance in a training session.</p>

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following performance indicators:

8. Could you identify, from the content of the session, specific hypotheses relating to patterns of maladaptive interaction that seemed to be informing the therapist's statements, decisions and actions?
9. Did the therapist appear to possess a good understanding of patterns of interaction typically found amongst distressed couples? If so, did they draw on this understanding to inform their dyadic conceptualisation?
10. Did the therapist find a way – using psychoeducation, examples or metaphors - to help the couple grasp the concept of dyadic conceptualisation?

Scale development: BCTS-D

11. Was the therapist able to use and 'rework' any patterns of maladaptive interaction (such as mutual blame) occurring in the session?
12. Was the dyadic conceptualisation used to guide any interventions used in the session, or assigned for homework?

Examples of statements, methods or approaches that reflect these indicators include the following:

- The therapist made statements such as, "Let's stop and look at what happened just then" (i.e., to highlight specific patterns of maladaptive interaction occurring in the session at a particular moment).
- "You both have legitimate needs here, and so are working really hard to convince each other that your individual perspective on the situation is correct. The problem is that this leaves each of you feeling unheard and misunderstood which makes the problem escalate. Let's see if we can find another way to approach this."
- "You are both very caring and considerable people, so you have a tendency to keep things to yourself, rather than risk upsetting your partner by raising difficult topics. The problem is that over time, nothing gets dealt with and you can start to feel resentful. I would like you to feel that you have freedom to raise difficult topics and know that it's safe to do so. Shall we talk about how it might be possible to raise difficult topics in a safe and caring way?"
- "I think there's a pattern here in what you are telling me/what I am observing, and I think it's important in helping us understand what causes the arguments/problem to escalate. Let me tell you what I think I am seeing and you can tell me if it fits with your experience. If I'm right, then we can think about what we can do to help you change that pattern."
- 'Mini-lectures'/psychoeducation relating to specific patterns of maladaptive interaction to which the therapist seeks to draw the couple's attention.

(NB: the above are offered as suggestions of potential illustrations of competence only)

Item 14. Setting Homework (for the next session)

Introduction

Unlike Item 2, which is concerned with reviewing the homework from the previous session, Item 14 addresses the therapist's skill in helping the couple identify an appropriate task/s, ensuring that the couple understands the potential benefits of undertaking the agreed homework, and anticipating any potential obstacles.

The homework should aim to help couples generalise in-session learning and skills development to their relationship outside of therapy; thus, homework should flow logically from the material covered in the session, be consistent with the case conceptualisation, and link back to the agreed goals of therapy.

In addition, it is important to obtain the couple's feedback regarding a proposed assignment ("Does it sound useful?" "Does it seem manageable?" "Is the assignment clear?" "What will be learned from the accomplishment/non-accomplishment of the task?"). These questions can help to determine whether the couple is clear about the task, and understands its rationale.

Because the setting of homework tends to occur towards the end of the session, there is sometimes a tendency to rush the process. This tendency should be avoided, as it can lead to ill-prepared and unclear tasks being set. Hence it is good practice to leave sufficient time to set the homework appropriately.

Features of competent performance

Competent performance for Item 14 comprises five main aspects evidenced through the following features:

1. *The presence/absence of a clear task*

The therapist took the time to establish with the couple a clear and appropriate task.

2. *The task should be derived from material discussed in the session, such that there was a clear understanding of what would be learnt from performing the task*

The therapist explored the couple's understanding of how engagement in the agreed task would support learning, growth, and change, and how completing the homework linked to the couple's aims for therapy.

3. *The homework task should be set in a collaborative fashion*

In addition to offering ideas, the therapist sought the couple's own views and ideas on what might represent a suitable homework task.

4. *Sufficient time should be allowed for the homework to be agreed upon, discussed, and evaluated for its practicalities and potential impact*

The homework was set in a thoughtful manner allowing the couple to reflect on both the potential benefits of following through on the homework task, as well as any potentially negative consequences (e.g., exposing oneself to uncomfortable feelings).

Scale development: BCTS-D

5. *Did the therapist enquire about any obstacles or challenges to completion of the homework?*

The therapist was attentive to, and probed for, any potential challenges or obstacles to completion and helped the couple problem-solve these potential impediments to constructive practice.

Scoring and anchor points

Examples of how therapist performance would be scored on this item using the anchor points are as follows:

0	<p>The therapist made no reference to homework and the couple left the session without any homework tasks.</p> <p>OR: The therapist set a highly inappropriate task that appeared to have no potential benefits to the couple in the context of their aims for therapy.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
1	<p>The therapist attempted to set homework but did so ineffectively such that their performance was highly limited. For example:</p> <p>The therapist may have 'set' homework but did so in a perfunctory manner.</p> <p>OR: The therapist set homework in a highly directive and unilateral fashion with little apparent regard for the couple's understanding of the task, or their response as to whether it would be helpful.</p> <p>OR: The therapist seemed irritated by any reservations that the couple expressed about engaging in a particular homework task.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback.</p>
2	<p>The therapist clearly attempted to set homework, but this was done in an inconsistent or slightly clumsy fashion. The therapist's actions showed evidence of emerging competence on this item, and an awareness of the importance of homework, but their delivery indicated a basic competence that could potentially (or actually did) negatively impact the flow of the session. For example:</p> <p>The therapist was overly didactic in manner, and more overtly 'instructional' than was necessary.</p> <p>OR: The therapist did not seem to have a clear understanding of what needed to be prioritised, so the homework set was vague, lacking in specificity.</p>

Scale development: BCTS-D

	<p>OR: The therapist made no attempt to help the couple consider what they might gain from engaging in the task.</p> <p>OR: the therapist did not attempt to elicit the couple's views on attempted obstacles or challenges that could undermine effective follow through.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but further work is needed to reach competence, with key areas for development identified in the assessor's feedback.</p>
3	<p>Overall the therapist did an effective job of setting homework. At times, the process may have lacked a degree of fluidity, but the essential components were present, with the five features of competent performance (listed above) evident. For example:</p> <p>The therapist allocated sufficient time to discuss and agree upon a specific homework task/s that was appropriate to the aims of therapy and the content of the session.</p> <p>AND: The therapist attempted to help the couple consider what they might gain through engaging with the task (although this may not necessarily have been explored fully or linked back to the tasks of therapy).</p> <p>AND: The therapist worked in an appropriately collaborative fashion with the couple to agree upon the task (as opposed to being overly didactic in approach).</p> <p>AND: The therapist made some attempt to enquire about any obstacles or challenges that could be encountered (although these may not have been addressed in in sufficient depth)</p> <p>In summary, the therapist demonstrated competence on this domain although refinement would be beneficial.</p>
4	<p>The therapist worked effectively to identify an appropriate homework task/s. The task was clear and flowed logically from the content of the session. Adequate time was given to identifying and planning the homework.</p> <p>The couple's ideas were actively sought to ensure that homework felt like a meaningful contribution to the aims of therapy. The therapist helped the couple appreciate the potential benefits of completing the task and problem-solved any potential obstacles.</p> <p>The therapist's performance was competent on this item with only minor areas of refinement needed.</p>
5	<p>The therapist demonstrated an obvious, high degree of skill on this domain. A clear homework task was identified and agreed in a collaborative fashion, such that the couple appeared to have a strong sense of ownership of the agreed activity/ies. The homework task was given sufficient time and attention to enable all parties to be confident that the task would be completed and any potential obstacles were discussed, with problem-solving.</p>

Scale development: BCTS-D

	<p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. The therapist's actions were entirely appropriate and nothing important was omitted. The performance was simply not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist's competence on this item was at 'master'/expert level. As for anchor point 5 (above), all five features of competent performance were present. Here, the therapist demonstrated a very high level of proficiency in setting homework. For example:</p> <p>A clear homework task was identified and agreed in a collaborative fashion, such that the couple appeared to have a strong sense of ownership of the agreed activity/ies.</p> <p>AND: The task identified was entirely appropriate for the stage of therapy and had the potential to help the couple make significant progress.</p> <p>AND: The homework task was given sufficient time and attention to enable all parties to be confident that the task would be completed and any potential obstacles were discussed, with problem-solving.</p> <p>In summary, the therapist demonstrated an obvious, exceptionally high degree of skill on this item, and would be deemed to be working at 'master level'. The therapist's actions were entirely appropriate and nothing important was omitted. A score of 6 on this item is likely to be rare and indicates a level of skill that might be used to demonstrate the highest levels of performance in a training session.</p>

Primers for the Assessor

To help you score this domain, it may be useful to ask yourself if you observed any of the following indicators of effective performance:

1. Did the therapist work with the couple to identify a homework task that was appropriate to the stage of therapy and the nature of the couple's difficulties or concerns?
2. Was the task clear, specific and achievable?
3. Did the task relate to what had been discussed in the session? Was there a logical connection?
4. If 'therapist led', did the therapist adequately explain the rationale underpinning the assignment?
5. Where appropriate, did the therapist encourage the couple to offer their own suggestions as to what might make an effective homework task?
6. Did the therapist ensure that the homework task made sense to the couple? Did they understand the relevance of the task to the 'bigger picture'?
7. Was the couple supported to make links between the task agreed and their goals for therapy?

Scale development: BCTS-D

8. Did the therapist check that the couple was confident about conducting the task correctly?
9. Were any potential obstacles identified and discussed to increase the likelihood of the couple being able to complete the agreed task/s?
10. Was the therapist attentive to any (potentially subtle) concerns or objections to the task that the couple might have had, and were objections welcomed in order to seek new understanding and make any modifications to the homework?

Examples of statements that reflect these indicators include the following:

- Based on what we've discussed today, what do you think would be useful to practice over the next week?
- What do you think you might learn/gain if you are able/ unable to carry out the task?
- How would practicing this task/engaging in this homework help you move a step closer to achieving your goals?
- How confident are you that you'll be able to follow through on what we've agreed?
- What could get in the way of your following through on this week's homework? What might prevent you carrying out the homework task successfully? (And what could you do about these?)
- Can you explain back to me what you'll be working on for the next week, just so we're all clear.

(NB: the above are offered as suggestions of potential illustrations of competence only.)

15. Ending the Session

Introduction

It is important for the therapist to create a good ending to the session. Even where the content of the session has proved emotive, or where high levels of distress have been present, skilful BCT will enable the couple to have an experience of 'closure' and feel enabled to end the session in a safe and contained way. This item is concerned with how the therapist manages to bring the session to a close and the skilfulness with which he/she enables the couple to leave with a consolidated understanding of what has been discussed. The ending should be a distinct phase of therapy that the therapist works towards, rather than an abrupt occurrence signalling that the therapist has run over on time.

(NB: Because the BCTS-D is designed to capture competence in a mid-treatment session, it is not usually used to determine the effectiveness with which the ending of therapy is conducted. As such, item 15 is not concerned with the termination of therapy, rather the degree of skilfulness with which the therapist concluded a given session.)

Features of competent performance

Competent performance for Item 15 comprises three main aspects evidenced through the following features:

1. Sufficient time for ending

The therapist ensured that there was sufficient time for the ending of the session so that key aspects of the session could be summarised.

2. Consolidation and feedback

The therapist provided an opportunity for the couple to consolidate and reflect upon their experience of the session, so that any important issues could be addressed before the couple left the consulting room. Not all sessions end on a positive note, and this is not a criterion for a high rating on this item. This item addresses whether the session was appropriately brought to an end. If the session ended with negative emotions or was somewhat unproductive, appropriate termination might involve a discussion of what made the session difficult, how the couple might interact differently during upcoming sessions or between sessions so as not to replicate the current experience, or a discussion of what the couple might do outside of session to get 'back on track' given that they currently are upset with each other. If the session went well, termination might involve reflecting upon what each partner did to contribute to that process, how they might continue this pattern in the future, and reinforcement from the therapist for a job well done.

3. Attentiveness to any process issues

The therapist conveyed an awareness of and sensitivity to any process issues associated with the ending of the session and responded to these effectively (for example, if a session had particularly focused on the behaviour change needs of one partner, the therapist acknowledged an awareness of this so that the partner did not feel 'picked on').

Scoring and anchor points

Scale development: BCTS-D

<p>0</p>	<p>The therapist failed to demonstrate any awareness of the need to end the session in a timely way.</p> <p>OR: The therapist's actions implied a wish to end the session as soon as possible.</p> <p>OR: The therapist ended the session with inappropriate comments relating to needing to be elsewhere.</p> <p>OR: The couple were 'dismissed' by the therapist without any attempt to consider the impact of the ending of the session for the couple.</p> <p>OR: The therapist was blatantly unaware that one or both members of the couple had been significantly impacted by the session and needed time to 'debrief'.</p> <p>In summary, there were significant concerns about the therapist's performance on this item which should be described more fully in the qualitative feedback and possibly addressed with the therapist in the context of supervision.</p>
<p>1</p>	<p>The therapist evidenced some awareness of the need to end the session with care, but did so ineffectively and as such, their performance was highly limited. For example:</p> <p>The therapist rushed through the ending, having run out of time.</p> <p>OR: Insufficient time was allocated to reviewing the session such that the couple could not consolidate and reflect upon their experience of the session.</p> <p>OR: The therapist acknowledged any process issues arising in a perfunctory manner, such that the couple was left without adequate 'closure'.</p> <p>In summary, the therapist did not display competence on this item and needs to attend to key areas as identified in the assessor's feedback. Something relevant to the item was present, but notably lacking. Significant improvement is needed.</p>
<p>2</p>	<p>The therapist was attempting to end the session well but problems were present. For example:</p> <p>The therapist may have begun well in concluding the session, but at some point there was a significant deviation such that the conclusion of the session went off track.</p> <p>OR: The therapist gave some, but insufficient, time to the end of the session so that the key themes of the session were not adequately summarised.</p> <p>OR: The therapist did not effectively support the couple in consolidating their understanding of the key points raised in the session.</p> <p>OR: The therapist omitted to seek feedback on the impact of the session.</p> <p>In summary, the therapist demonstrated an emerging competence on this item but</p>

Scale development: BCTS-D

	<p>further work is needed to reach competence. Something notable was missing, or there was a deviation from good performance at some point such that it was not fully acceptable, just below passing. Key areas for development are identified in the assessor's feedback.</p>
3	<p>The therapist was clearly attempting to end the session well, and overall managed the ending of the session effectively. For example:</p> <p>The therapist sought to provide sufficient time for the ending of the session.</p> <p>AND/OR: Generally, the key themes of the session were adequately summarised.</p> <p>AND/OR: The therapist, overall, managed to support the couple in consolidating their understanding of the key points raised in the session.</p> <p>OR: An attempt was made to seek feedback on the impact of the session.</p> <p>In summary, the therapist demonstrated competence on this item although refinement would be beneficial.</p>
4	<p>The therapist was effective in ending the session. For example:</p> <p>Sufficient time was allocated to ending the session.</p> <p>AND/OR: The key themes of the session were effectively summarised.</p> <p>AND/OR: The therapist was skilled in supporting the couple in consolidating their understanding of the key points raised in the session, probing for any points of confusion or misunderstanding.</p> <p>AND/OR: The therapist sought feedback on the impact of the session and any issues arising were identified and addressed.</p> <p>AND/OR: Any process issues arising were identified and reviewed.</p> <p>The therapist's performance was competent on this item with only minor areas of refinement needed.</p>
5	<p>The therapist demonstrated a high degree of proficiency on managing the ending of the session. For example:</p> <p>Sufficient time was allocated to ending the session such that the session appeared to flow logically towards its conclusion.</p> <p>AND: The key themes of the session were effectively summarised, with the therapist ensuring that any summaries were a collaborative effort.</p>

Scale development: BCTS-D

	<p>AND: The therapist was highly skilled in supporting the couple in consolidating their understanding of the key points raised in the session, probing for any points of confusion or misunderstanding.</p> <p>AND: The therapist sought feedback on the impact of the session and any issues arising were identified and addressed.</p> <p>AND: The therapist conveyed an awareness of, and sensitivity towards, any process issues that had arisen in the session.</p> <p>In summary, the therapist demonstrated an obvious, very high degree of skill on this item. The therapist's actions were entirely appropriate and nothing important was omitted. The performance was simply not at the exemplary level of master which is required for a score of 6.</p>
6	<p>The therapist's competence on this item was at 'master'/expert level. As for anchor point 5 (above), all three features of competent performance were present with the therapist demonstrating an exceptional level of proficiency in how the ending of the session was managed. For example:</p> <p>Sufficient time was allocated to ending the session such that the session appeared to flow logically and smoothly towards its conclusion.</p> <p>AND: The key themes of the session were effectively summarised, with the therapist ensuring that any summaries were a collaborative effort.</p> <p>AND: The therapist was highly skilled in supporting the couple in consolidating their understanding of the key points raised in the session, probing for any points of confusion or misunderstanding as well as key points of learning.</p> <p>AND: If appropriate, the therapist commended the couple on the efforts they had made in the session.</p> <p>AND: The therapist sought feedback on the impact of the session and any issues arising were identified and addressed.</p> <p>AND: The therapist conveyed an awareness of, and sensitivity towards, any process issues that had arisen in the session, ensuring that the session did not end until adequate 'closure' had been achieved.</p> <p>In summary, the therapist demonstrated an obvious, exceptionally high degree of skill on this item, and would be deemed to be working at 'master level'. The therapist's actions were entirely appropriate and nothing important was omitted. A score of 6 on this item is likely to be rare and indicates a level of skill that might be used to demonstrate the highest levels of performance in a training session.</p>

Scale development: BCTS-D

Primers for the Assessor

To help you score this item, it may be useful to ask yourself if you observed any of the following indicators of effective performance:

1. Did the therapist provide the couple with the opportunity to consolidate their recollection and understanding of the main points of the session?
2. Did the therapist enlist the couple's active involvement in summarising the session, seeking feedback on what had seemed most important about the session?
3. Was the couple given the opportunity to provide negative feedback on their experience of the session (and if this was forthcoming, did the therapist react in a helpful and non-defensive manner)?
4. Were any particularly salient features of the session identified? For example:
 - a. if the session was a particularly emotionally challenging one for the couple, did the therapist acknowledge this?
 - b. if the session led to the discovery of important new information or positive changes were reported, did the therapist acknowledge these in the concluding phase of the session?
5. Was the couple commended for any efforts that they made, including any attempts to work with the therapist to address emotionally sensitive topics?
6. Was adequate time devoted to drawing the session to a close?
7. Was the couple encouraged to summarise what, for them, had been any key learnings or important points of discussion?
8. Did the therapist seek feedback from the couple as their experience of the session?
9. Was the couple given sufficient opportunity to reflect on their experience of the session?

Examples of statements that reflect these indicators include the following:

- Before we end, let's summarise what we've discussed today.
- What will each of you take away from today's meeting?
- Let's take a moment to review anything we have discussed that has been particularly helpful for each of you.
- Is there anything that you found unhelpful about today's session, or anything I got wrong?
- We have talked about some difficult themes today. Before we end the session, I want to check how you are both feeling, and that it feels OK to end the session at this point.
- I think we have identified some really important themes today and I want to commend you for the work that you've put in to this.

(NB: the above are offered as suggestions of potential illustrations of competence only.)

Part 3. Additional information

Scale development: BCTS-D

References

Barbato, A. & D'Avanzo, B. (2008). Efficacy of Couple Therapy as a Treatment for Depression: A Meta-Analysis. *Psychiatr Q*, 79: 121–132. DOI 10.1007/s11126-008-9068-0.

Baucom, D.H., Whisman, M.A. & Paprocki, C. (2012). Couple-based interventions for psychopathology. *Journal of Family Therapy* (2012) 34: 250–270. doi: 10.1111/j.1467-6427.2012.00600.x

Blackburn, I.M., James, I.A., Milne, D.L., Baker, C., Standart, S., Garland, A. and Reichelt, K. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29: 431-46.

British Association for Behavioural & Cognitive Psychotherapies (2012). *BABCP Minimum Training Standards For the Practice of Cognitive Behavioural Therapy (CBT), 2012*. (www.babcp.com/files/Minimum-Training-Standards-V6-0413.pdf).

Dreyfus, H.L. & Dreyfus, S.E. (1986). *Mind over Machine. The Power of Human Intuition and Expertise in the Era of the Computer*. New York: The Free Press.

Epstein, N.B. & Baucom, D.H. (2002). *Enhanced Cognitive-Behavioral Therapy for Couples*. Washington, DC: American Psychological Association.

Gambrill, E. (2005). *Critical Thinking in Clinical Practice. Improving the Quality of Judgments and Decisions*, 2nd ed. Hoboken, New Jersey: Wiley.

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

Padesky, C.A. (1993). *Socratic questioning: Changing minds or guiding discovery?* Invited keynote address presented at the 1993 European Congress of Behaviour and Cognitive Therapies, London. (Available from http://padesky.com/clinical_corner.htm.)

Whisman, M.A. & Baucom, D.H. (2012). Intimate Relationships and Psychopathology. *Clinical Child Family Psychology Review*, 15: 4-13. DOI 10.1007/s10567-011-0107-2.

Scale development: BCTS-D

Contact details for the scale developers

For further information about the BCTS-D, this accompanying manual and on-going research into the development of the scale, please contact:

Sarah Corrie: sarah.corrie@nhs.net

Melanie S. Fischer: melanie_fischer@med.unc.edu

Michael Worrell: michael.worrell@nhs.net

Donald H. Baucom: don_baucom@unc.edu

Appendix L: Information sheet for study two



Doctorate
in Clinical
Psychology

Department of Psychology
Egham Hill
Egham
TW20 0EX
TEL: 01784 414012
EMAIL: isabelle.rudolfvonrohr.2013@live.rhul.ac.uk



PARTICIPANT INFORMATION SHEET

An evaluation of a novel BCT competence rating scale

Principal Researcher: Isabelle Rudolf von Rohr, Supervisor: Dr. Helen Pote, Dr. Michael Worrell

You are being invited to take part in a study that aims to evaluate the psychometric properties (reliability, validity, usability) of a new Behavioural Couples Therapy competence rating scale for Depression (BCTS-D). Please read the following information explaining why the study is being conducted and what your involvement would be. If you have any questions about the study after reading this information sheet, please feel free to contact the principal researcher (Isabelle Rudolf von Rohr) using the contact details outlined on page 3.

Why have I been contacted?

You have been invited to take part in the study because you **are affiliated with** the Post Qualification course in Behavioural Couples Therapy at the Central London training centre for CBT.

What is the purpose of the study?

There is a need for methods of assessing BCT competence, particularly those, which assess whether therapists can demonstrate the skills necessary to effectively deliver BCT. We have developed a BCT competence rating scale, which can be used to provide formative and summative feedback regarding therapists' performance within observed treatment sessions. The central aim in developing this scale is to provide a tool, which is valid, reliable and usable.

Please be aware that the current study consists of three separate parts which might not all be relevant to you. What part we will invite you to participate in, is solely based on a convenience sampling method.

The first part of the study aims to investigate the psychometric properties (e.g. reliability, validity) of the rating scale in order to evaluate and further refine the scale. The second part is a review aiming to gain feedback on the new rating scale in order to examine (i) usability, (ii) face validity (inclusion of appropriate items that are a credible and plausible measure of competence) and (iii) content validity (adequate sampling of all aspects of competence specified in the scale definition) of the scale.

The study is being conducted by Isabelle Rudolf von Rohr as a part of a DClinPsych at the Department of Psychology, Royal Holloway University of London. If you agree to participate in this project, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be published with open access, meaning available to every internet user. It is our intention to publish the results of the study in

Scale development: BCTS-D

order to make the results of the scale evaluation publically available. Please note that **no participant would be personally identifiable within any of these published research findings.**

What does participation involve?

We will also ask everyone to complete a short demographics form, which will take approximately five minutes to complete. Your session recording will not be accessed or viewed by any individuals other than the Central London training centre for CBT staff.

As you are aware, as part of your BCT training course you will be asked to submit recordings of BCT treatment sessions to your clinical supervisors for feedback on your performance. Participation in **this part of the** study would involve these recordings being viewed and rated using the BCTS-D by two people:

1. Your supervisor within the Central London training centre for CBT
2. A Senior Psychotherapist within the Central London training centre for CBT

Only anonymised demographics forms and rating sheets (i.e. paperwork identified by a unique ID number rather than your name) will be sent to Isabelle Rudolf von Rohr- principal investigator- at the Department of Psychology, Royal Holloway University of London to ensure anonymity.

Whoever takes part in the second part of the study, will be asked to read through the rating scale and then to complete a feedback questionnaire about the scale. Taking part in this study is expected to last approximately one hour. Please note that all questionnaires will be identified by a unique ID number, rather than your name. **And the third part would involve taking part in a group discussion about the scale lasting for about an hour.**

Am I required to take part?

It is entirely up to you if you wish to take part. If you do decide to take part, you are free to change your mind at any time. You can withdraw during any phase of the study, without giving a reason and without any penalty, by letting the researcher know. If this is the case, any data collected from you will no longer be included in subsequent analyses and will be destroyed.

Will my taking part in the study be kept confidential?

All information which is collected from you during the course of the research would be kept strictly confidential within the limits of the law. You will be allocated a unique number, ensuring that all materials related to your participation (e.g. completed questionnaires) will contain a unique number rather than your actual name before it's passed on to the research team. **The group discussion will be audio recorded and transcribed for subsequent analysis. Again the any information collected will be kept confidential.**

In accordance with British Psychological Society research guidelines, all data for the study will be securely stored for 5 years and will be destroyed after this time. Data will be accessed only by members of the research team. However, individuals from Royal Holloway University of London and other regulatory authorities may require access to relevant data for the purpose of audit and monitoring only.

What are the possible advantages of taking part?

Taking part in this study will give you a chance to directly influence the development of this scale and the information you provide will be beneficial in evaluating the reliability, validity and usability of the BCTS-D. It is important to have a valid and reliable competency scale supporting training and your self-reflection.

Scale development: BCTS-D

Thus, the information you provide will be beneficial in helping to highlight areas where the scale requires refinement and thus will help to improve current methods of assessing BCT competence. Further, it will help you develop your skills and give you a chance to familiarise yourself with a competency tool used in clinical practice.

What are the possible disadvantages of taking part?

Given the nature of this study, it is highly unlikely that you will suffer harm by taking part. However, if the rating scale or the questionnaire happens to include any questions which, for whatever reason, you do not wish to answer then the question can be omitted.

Who has reviewed this study?

The study is in the process of being reviewed for ethics by the Health Research Authority and the Royal Holloway University of London Research Ethics Committee.

What if there is a problem?

If you have a concern about any aspect of this project, please speak to the researcher concerned (contact details on page 3) who will do her best to answer your query. If you remain unhappy and wish to make a formal complaint, please contact the Health Research Authority (contact.hra@nhs.net) or call 020 710 48066).

Contact Details:

If you require further information or would like to ask any questions, please do not hesitate to contact either the Principal Researcher or Supervisor using the details below.

Principal Researcher:

Isabelle Rudolf von Rohr
Department of Psychology
Egham Hill
Egham
TW20 0EX
Tel: 01784 414012
Email:
isabelle.rudolfvonrohr.2013@live.rhul.ac.uk

Supervisor:

Dr. Helen Pote
Department of Psychology
Egham Hill
Egham
TW20 0EX
Tel: +44 1784 414236
Email: h.pote@rhul.ac.uk

Supervisor:

Dr. Michael Worrell
Consultant Clinical Psychologist
Psychology Department
7a Woodfield Road London
W9 2NW
Telephone: 02072669588
Mobile: 07772517750
email: michael.worrell@nhs.net

Appendix M: Informed consent form for study two



Doctorate
in Clinical
Psychology

Department of Psychology
Egham Hill
Egham
TW20 0EX

TEL: 01784 414012

EMAIL: isabelle.rudolfvonrohr.2013@live.rhul.ac.uk



CONSENT FORM

Evaluating a novel BCT competence rating scale

Principal Researcher: Isabelle Rudolf von Rohr, Supervisors: Michael Worrell, Dr. Helen Pote

Study Purpose: The study aims to evaluate the psychometric properties (reliability, validity, usability) of a new Behavioural Couples Therapy competence rating scale for Depression (BCTS-D).

Please initial box

I confirm that I have read and understand the information sheet dated 26.05.2016 for the above study. I have had the opportunity to consider the information, ask questions and have had any questions I asked answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without it affecting my training.

I understand that personal data will be stored and identified using only a number code, will be accessed only by members of the research team and will be destroyed after a period of 5 years.

I understand that data collected during the study may be looked at by individuals from Royal Holloway University of London and other regulatory authorities, for the purpose of research audit and monitoring, and where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.

I understand that the research will be written up as a thesis which will be deposited both in print and online in the University archives. The thesis will be published with open access, meaning available to every internet user.

I understand that the study is in the process of being reviewed for ethics by HRA and the Royal Holloway University of London Research Ethics Committee.

I understand how to raise a concern or make a complaint.

I agree to take part in the above study.

Name of Participant

Date

Signature

Researcher

Date

Signature

Scale development: BCTS-D

Appendix N: Feedback questionnaire for study two

Feedback questionnaire on the BCTS-D

Please answer the following questions about the **BCTS-D overall**:

Questions	Ratings: please circle
<p>Do you feel the scale provides you the opportunity to judge how competent a session went? <i>(overall judgment of competence - a numerical rating)</i></p>	<p>----- ----- ----- </p> <p>1 2 3 4</p> <p>not somewhat quite very useful useful useful useful</p> <p><u>If anything</u>, what did you feel was missing?</p>
<p><u>Does the scale give you useful feedback?</u> <i>(detailed, corrective feedback in order to promote self-reflection and guide future learning)</i></p>	<p>----- ----- ----- </p> <p>1 2 3 4</p> <p>not somewhat quite very useful useful useful useful</p> <p><u>If anything</u>, what did you feel was missing?</p>
<p>How <u>appropriate</u> do you find the scoring system? <i>Examples of problems with the scoring system include an inappropriate number or range of response categories or an unsuitable response format.</i></p> <p>If you circled a <u>three or below</u>, what revisions do you feel could improve the item?</p>	<p>----- ----- ----- </p> <p>1 2 3 4</p> <p>not somewhat quite very appropriate appropriate appropriate appropriate</p>

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

<p>Please rate the overall <u>style, appearance and layout</u> of the scale.</p> <p><i>Examples of problems with the overall appearance of the scale include the use of text which is difficult to read or a confusing layout.</i></p> <p>If you circled a <u>three or below</u>, what revisions do you feel could improve the overall style, appearance and layout of the scale?</p>	<p>----- ----- ----- </p> <p>1 2 3 4</p> <p>poor fair good very good</p>
<p>How <u>easy</u> do you think the rating scale is <u>to use</u>?</p> <p><i>Examples of problems with usability include unclear or confusing instructions, difficulty understanding what is being asked or how to respond or unnecessary complexity.</i></p> <p>If you circled a <u>three or below</u>, what revisions do you feel could improve the usability of the scale?</p>	<p>----- ----- ----- </p> <p>1 2 3 4</p> <p>not somewhat quite very easy easy easy easy</p>
<p>Are there any important aspects of BCT competence which you feel are <u>missing</u> from the scale (i.e. any key competences which the scale neglects)?</p>	<p><u>If yes</u>, what aspects of BCT competence do you feel are missing?</p>

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

<p>Do you have any <u>other comments</u> or feedback about the rating scale or any other suggestions of <u>ways the rating scale could be improved?</u></p>	<p>Comments:</p>
---	------------------

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

In the next section, you will be asked to rate each item individually on its **relevance and clarity** to the assessment of competence when delivering an intervention to a couple where at least one of them is experiencing depression (Behavioural Couples Therapy).

1 : not relevant/ clear 2: somewhat relevant/ clear 3: quite relevant/ clear 4: very relevant/ clear

Further, there is space for any **comments** or feedback about each item (e.g. problems with any of the item descriptions) or any suggestions of **ways the item could be improved?**

Item 1 – Agenda Setting

Please rate how relevant:	----- ----- -----
	1 2 3 4
	not somewhat quite very
and clear:	----- ----- -----
	1 2 3 4
	not somewhat quite very
Comments / improvements:	

Item 2 – Review of Homework (from the previous session)

Please rate how relevant:	----- ----- -----
	1 2 3 4
	not somewhat quite very
and clear:	----- ----- -----
	1 2 3 4
	not somewhat quite very
Comments / improvements:	

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

Item 3 - Collaboration

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	

Item 4 – Facilitating Couple Communication

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

Item 5 – Pacing and Flow

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Item 6 – Therapist’s Interpersonal Effectiveness

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

Item 7 – Guided Discovery

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Item 8 – Formulation of depression in Context

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

Item 9 – Selection of Intervention Strategy

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	

Item 10 – Emotion-Focused Intervention

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

Item 11 – Cognitive Interventions

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Item 12 – Behavioural Interventions

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Feedback questionnaire on the BCTS-D

Item 13 – Dyadic Conceptualisation

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Item 14 – Setting Homework (for next session)

Please rate how relevant:	----- ----- ----- 1 2 3 4 not somewhat quite very
and clear:	----- ----- ----- 1 2 3 4 not somewhat quite very
Comments / improvements:	not somewhat quite very

Scale development: BCTS-D

Feedback questionnaire on the BCTS-D

Item 15 – Ending the Session

Please rate how relevant:	----- ----- -----
	1 2 3 4
	not somewhat quite very
and clear:	----- ----- -----
	1 2 3 4
Comments / improvements:	not somewhat quite very

Appendix O: Summary of interview schedule

Open ended questions	Example Prompts
<p>What has been your experience of using the BCTS-D? How have you found it?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Do you have any examples of that which you could tell me about? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience?
<p>What did you most like about the BCTS-D?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Why was it that you particularly liked xxx? • Do you have any examples of that which you could tell me about? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Is there anything else anybody particularly liked about the BCTS-D?
<p>What did you least like about the BCTS-D?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Why was it that you particularly did not like xxx?

	<ul style="list-style-type: none"> • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Do you think the issue of xxx could be resolved or overcome? Could you tell me how you think it could be resolved or overcome? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Is there anything else anybody particularly did not like about the BCTS-D?
<p>How easy or difficult was it to use the BCTS-D?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Why was it that xxx made it particularly difficult/easy to use the BCTS-D? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • What would make the BCTS-D easier to use? • Do you think the issue of xxx could be resolved or overcome? Could you tell me how you think it could be resolved or overcome? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Did anybody have any other comments about the usability of the BCTS-D?
<p>How clear did you find the items in the scale?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that?

<p><i>More specific probing questions¹:</i> Were any items unclear or confusing?</p>	<ul style="list-style-type: none"> • Why was it that xxx was unclear or confusing? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • What would help to clarify xxx? • Do you think the issue of xxx could be resolved or overcome? Could you tell me how you think it could be resolved or overcome? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Is there anything else anybody found unclear or confusing?
<p>How appropriate did you find the scoring system?</p> <p><i>More specific probing questions:</i> Did you like the generic 6-point scale, with item-specific anchor descriptions? Did you use ½ marks? Do you think this is necessary/important? Why?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Why was it that xxx was a problem? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Do you think the issue of xxx could be resolved or overcome? Could you tell me how you think it could be resolved or overcome? • What revisions do you feel could improve the scoring system? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Did anybody have any other comments about the scoring system?
<p>How relevant did you think the BCTS-D</p>	<p><u>Probing questions:</u></p>

<p>was to the assessment of BCT competence?</p> <p><i>More specific probing questions:</i> Did you feel that any of the items were not particularly relevant? Would you like to have seen a 'not applicable' or 'appropriately omitted' option for any of the items?</p>	<ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Why was it that xxx was not relevant? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Do you think the issue of xxx could be resolved or overcome? Could you tell me how you think it could be resolved or overcome? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Did anybody have any other comments about the relevance of the scale items?
<p>Did you feel that there was anything missing from the scale?</p> <p><i>More specific probing questions:</i> Are there any key competences which the scale neglects?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Why do you feel it is important to include xxx? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Could you tell me more about how you would like to see xxx incorporated within the scale? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Did anybody have any other comments about anything that was missing from the scale?

<p>What did you think about the overall style, appearance and layout of the scale?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • Could you tell me what you mean by that? • Could you say something more about that? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Could the issue of xxx be resolved or overcome? Could you tell me how you think it could be resolved or overcome? • How do you think the overall style, appearance and layout of the scale could be improved? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience? • Did anybody have any other comments about the overall style, appearance and layout of the scale?
<p>How helpful was BCTS-D in evaluating your BCT competence skills?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • In what ways? • Could you tell me what you mean by that? • Could you say something more about that? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Anything that would need changing in the scale to make it more user-friendly? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience?

<p>Did completing it inform your practice?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • In what ways? • Could you tell me what you mean by that? • Could you say something more about that? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Anything that would need changing in the scale to make it more user-friendly? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience?
<p>Could the use of the BCTS-D enhance the supervision of your BCT practice?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • In what ways? • Could you tell me what you mean by that? • Could you say something more about that? • Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none"> • Anything that would need changing in the scale to make it more user-friendly? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none"> • Did anybody else in the group experience this? • Did anybody else in the group have a different experience?
<p>Did any aspects of the rating process lead to devaluation or deterioration of your BCT competence?</p>	<p><u>Probing questions:</u></p> <ul style="list-style-type: none"> • In what ways? • Could you tell me what you mean by that? • Could you say something more about that?

	<ul style="list-style-type: none">• Do you have any examples of that which you could tell me about? <p><u>Development questions:</u></p> <ul style="list-style-type: none">• Anything that would need changing in the scale to make it more user-friendly? <p><u>Comparative questions:</u></p> <ul style="list-style-type: none">• Did anybody else in the group experience this? <p>Did anybody else in the group have a different experience?</p>
--	--

Note- At the end of the interview, all participants will be asked whether there was “anything we haven’t spoken about which you feel would be relevant or important?”

Appendix P: BCT workshop transcript

BCT workshop transcript

M: What was it like to use the scale experientially?

C1: challenging. Because not being familiar with the scale, not had the time to really understand it fully and what is it actually trying to capture that. (people agreeing with that)

M: It's gonna be a learning curve. There is a lot of new information.

What exactly are these items about. Trying to get at in terms of that experience of being challenged, say more about what was it like? If challenging was a good word for it, what was that like?

C2: suppose not being aware of what are the different kind of aspects of that scale. There is a bit of overlap. So you find yourself writing the same thing on different boxes and it probably part of the session we were listening to.

M: There is a bit of overlap on some of those items. Or you found listening that the qualitative feedback you were giving was cutting across?

C2: yeah

M: that's important for us, because we want to eliminate the overlap if that's unnecessary. Were there any areas in particular you thought that was the case having a few items saying the same thing?

C3: I guess. It gives a bit of an objective sense of where people are in terms of their competency.

M: So you are saying it felt like it did that job and gave you a bit of an objective sense of where they were?

C3: yes. Sometimes you overrate yourself or underrate yourself. But I guess if you have a tool like this, that might be useful instrument to use for self-reflection towards development as a BCT therapist. Rather than rely on external supervision, you can use this tool as your own way of guiding your own development.

M: ok that's really helpful. I appreciate this is the first time you are working with this scale in its current form. So the initial experience is that it's really challenging. You try and keep lots of things in mind and have to decide does that belong in that section does that belong in the other section – making those distinctions. I think this might be a feature right throughout the experience of using the scale. Tell us more about what it was like doing it. It was challenging. What else?

C4: think part of the challenge for me was, when it was introduced it was said you score from the manual, which is a 98 page document that you use to score from and then we had to focus in on the actual score sheet, which is a useful kind of shortcut, but that made me wonder whether I was using it correctly. We haven't had enough time to absorb all the subtleties, because BCT is much less kind of manualised compared to CBT. There is more room for judgement. It's hard to know how you interpret that in terms of the scale.

Isabelle Rudolf von Rohr

Issues with rating: Challenging using the scale

kate muse

The need for familiarity and practice to feel confident enough to use the scale

Isabelle Rudolf von Rohr

Issues with rating: A lot of information

kate muse

Lack of understanding about what scale items are measuring.

kate muse

Overlap between scale items makes rating difficult.

Isabelle Rudolf von Rohr

Overlap of items: There is overlap of the items – writing same thing in different boxes

Isabelle Rudolf von Rohr

Positive comment: Objective sense of where in development of competencies

kate muse

Scale gives insight into therapists' development / ability.

kate muse

Self-presentation biases are challenging for self-ratings.

kate muse

Scale helps to promote more accurate self-reflection.

Isabelle Rudolf von Rohr

Positive comment: Useful for self-reflection, guide your own development

kate muse

Lack of clarity about how to use the manual / scoring sheet.

Isabelle Rudolf von Rohr

Issues with rating: Very long manual

Isabelle Rudolf von Rohr

Issues with rating: Used scoresheet as a short cut, wondering if scoring correctly

kate muse

Lack of certainty / agreement about what competent BCT actually is.

Isabelle Rudolf von Rohr

Issues with rating: More room for judgement on each item compared to CBT

M: So there is a lot in there. There is a lot more room for making judgments and so there was a lot more uncertainty, sounds like in relation to that you've got a 100 page manual to refer to, which is a lot trickier than referring to the helpfully designed score sheet. But you have got to go from one to the other. So you are saying because it's less manualised it leaves you with a lot more room for decisions about your ratings and that's demanding.

C4: yes. As you get more familiar with it I think it will become easier.

M: we hope right. Does it match this experience with the CTS-R. Do you still use the manual for the CTS-R?

C4: yes / sometimes

M: all the time?

C4: No

C5: So I was worried about how I am going to cope with a 98 page manual. Without looking at the manual today I feel I did really well.

M: so let's hear from more. Even if just to say you agree with what's been said so far. We want a sense of how strongly shared this experience was: of it being challenging that was my word I introduced, or that there was a lot of uncertainty or more decisions. Let's just hear. Even if you feel like you are repeating. Is that a general sense?

C6: Yeah. Again it might be repetition of what others have said. I felt there was a heck of a lot uncertainty again because I am unfamiliar with the criteria or the sort of anchor points or the scoring system. But also because I have got nothing to compare it to. I think once you get familiar with the BCTS-D you have listened to a few recordings and you may have done some group scoring and got some consensus on when somebody does that they tend to score roughly like this. And for me personally because I have only just done the training, I don't really have that. I feel like I'm going in blind. Allocating scores according to this, but I've got nothing. No framework in terms of ratings or other sessions.

M: I wonder actually if this. Does that match your experience of learning BCT as it is less manualised. Is there more uncertainty when you are first starting or even as you are continuing to do that in a particular session wondering what am I doing, am I doing it right?

C7: There is a lot more to consider. So when you are saying things like cognitive interventions. Even though I have done the training and have started a case study I am still in that stage of thinking it should be useful to have little snapshots, what were they again, a list of all of them. What were the cognitive and behavioural interventions, just like a little prompt in case I have forgotten when I was scoring that and why I was scoring lower because I forgot one particular one. Bearing in mind where I am at in stage of my training that might be something useful.

M: How about those who perhaps have had a bit more experience. We have supervisors on our programme. How does it feel in terms of your ease with it or the sense of uncertainty. Perhaps how you felt earlier when you first started doing it.

kate muse

The need for familiarity to be confident in using the scale.

Isabelle Rudolf von Rohr

Positive comment: Scoring sheet really helpful

kate muse

Manual is overwhelming and long- meaning that people don't use it.

kate muse

The need for familiarity and practice to feel confident enough to use the scale

Isabelle Rudolf von Rohr

Issues with rating: Uncertainty with how to rate certain item
- more room for judgment

kate muse

Practice makes perfect! Need to practice to feel confident in using the scale.

Isabelle Rudolf von Rohr

Incorporate in scale: Helpful to have cognitive, behaviour and emotion-focused interventions listed on scoring sheet to remind me

kate muse

Lack of familiarity with BCT in general makes using the scale challenging.

C8 (supervisor): in terms of delivering the treatment? Or doing the ratings?

M: both in the sense. I was picking up there might be an overlap there. I was just wondering from a supervisor's perspective does it feel like there a lot of uncertainty for you? Does it match with what other people are saying?

C8: I kind of like that.

M: You like the uncertainty?

C8: Well the flexibility (people agreeing)

M: Flexibility. That's what I like about it too.

C8: yeah. Which is true for the treatment and also true for the ratings. You have scope to rate someone highly for doing certain interventions really well in the session. You are less likely to be penalising heavily for sort of neglect of the cognitive interventions because that might be entirely appropriate for that couple at that stage of treatment.

M: I really like the same and I was hoping that would be there and there would be greater flexibility and how you might use the scale. Both matches BCT itself and also maybe just starting to overcome some of the problems that I perceive in the CTS-R of being a little rigid. There are lots of nods.

C9: I think it's good the way Sarah said not to worry about that you don't have to do all the cognitive, emotional and behavioural interventions if the session focuses on one just score them. That's a clear difference from the CTS-R and takes a lot of pressure away. Cos you kind of have an idea form the agenda, don't you, where the session is going and where the emphasis is going to be. So it's a lot easier I think.

C10: I think I find the excess pointers very helpful. on page 28 the criteria or options of that makes it a bit more broader and flexible rather than the CTS-R.

M: challenging flip side there might be a bit more flexibility with this scale matching the approach of the therapy itself. Any other broad bits of feedback what's it like to use it what's it like to listen to a session like this? Trying to use this new instrument as a way of making judgements of degree of competence and what you might focus on in supervision. Broad headline bits of feedback.

C: I think it's really useful to have those categories and it means that you can remember to note things that you have done well or well enough even if it's a really difficult session which frankly with a couple session it might well be. It could be a session with loads of distress in the session where it's kind of hard to contain things at the end where it doesn't feel hugely successful but there would be scope to rate positively.

M: got you. It can go all wrong – couple therapy sessions with a difficult couple. If we think about the history of that couple these guys weren't communicate well at all. And someone had to pull some teeth out to get anywhere and that can be real difficult. But that doesn't mean that the therapist has done a bad job.

Isabelle Rudolf von Rohr

Positive comment: Scale is very flexible – not penalising if not using cognitive intervention in appropriate manner (many people agree)

kate muse

Uncertainty or flexibility? A pro or a con? Or both? Something about the need for flexibility but this is double edged as it brings with it more rater uncertainty / ambiguity?

kate muse

MORE ABOUT: Uncertainty or flexibility? A pro or a con? Or both? Something about the need for flexibility but this is double edged as it brings with it more rater uncertainty / ambiguity?

Isabelle Rudolf von Rohr

Positive comment: flexibility

Isabelle Rudolf von Rohr

Positive comment: Anchor points on the scoring sheet are very helpful

kate muse

Scoring sheet provides useful guidance

Isabelle Rudolf von Rohr

Positive comment: Helpful to have anchor points – rating easier especially if potentially difficult couple

kate muse

Scoring sheet provides useful guidance

C8: No and if say the therapist has to do the traffic cop thing you just have to adjust your expectation of what they can cover in that session. And it felt there was scope to mark accordingly.

kate muse
Flexibility to adapt to contextual factors.

M: Good. We don't want this to be a tool with which beat ourselves up or other people. That wouldn't serve anyone at all. Does it feel like that? Does it feel like it could be something that you could go home and rate myself badly... joke. Other broad reactions?

Isabelle Rudolf von Rohr
Positive comment: Scope to rate according to couple's needs

C3: I guess I'd say I need to shift my thinking if you look how you use the CTS-R for individual sessions and how you are going to use this. Because I guess there are clear similarities and clear differences. I guess for us the users to be able to be mindful of that, because sometimes there might be this

kate muse
Needing to change mindset- remember the differences between couples vs. individual CBT

Isabelle Rudolf von Rohr
Mindful of differences to CTS-R as so many things are very similar

automatic response for example you take the section related to agenda setting. The way you do agenda setting for a session for individual session will be different for what you'll be doing here. For individual sessions you might be very rigid and you can clearly prioritize the order that couple sessions they are clearly different form that. I guess being mindful of that as practitioners we'll have to keep at the back of our mind.

Isabelle Rudolf von Rohr
Clarification: Agenda setting: very different with couples than with individuals, need to keep in mind

M: absolutely. Cool. So can people now throw in in terms of what you liked and what you didn't like is there anything about the experience of doing that, the scale or process or the way we framed things. You know I don't like that or I struggled with that.

Isabelle Rudolf von Rohr
Positive comment; flexible

C: Just some of the points that were already raised was I like the fact that it is quite flexible that there is a lot of points to go through at the start of it.

kate muse
Another flexibility comment.

Isabelle Rudolf von Rohr
Issues with rating: very long

M: yes, we've added some points. It's longer than the CTS-R. man..

kate muse
Length makes it time consuming.

C: yeah. So it's kind of keeping that in mind I suppose but then also what is quite difficult for me. Which I can understand that's why it is flexible, but sometimes with like interpersonal effectiveness what might work for one couple in the way you work with might be the appropriate thing to do, but

kate muse
Again this issue of flexibility vs. subjectivity- a double edged sword.

you might miss important other bits collaboration for example that you can't really kind of rate. In essence It might have become more subjective rather than objective. It feels to me that it stands the chance of being more subjective if you use the scale if there is too much flexibility. So it's more about like how to kind of couch that.

Isabelle Rudolf von Rohr
Issues with rating: risk of item 'interpersonal effectiveness' being quite subjective

Isabelle Rudolf von Rohr
Issues with rating: if too flexible might become too subjective- and then be more do I like this therapist than how competent they are

M: So it's more a concern you have rather than you dislike it. You wouldn't want to see this be given too much room for a subjective response which might be more a global response do I like this therapist or is the couple interesting? General sense rather than hard and objective criteria. Is that a shared worry?

C: yeah.

M: nodding

C: I guess that could be true of any sort of evaluation of clinical work whether its structured or unstructured there is going to be bias. Particularly with interpersonal effectiveness criteria.

M: It's a bit of a polarity we're doing we're in kind of a dilemma the more objective and nailed down we get the more it turns into something more manualised way and if we are more rigid it takes us away from the principles of BCT. But if we go too far the other way it will be tricky to get any valid ratings of competence. There is we belief a set of competence underlying the CBT world. it's something about how we manage that tensions of between the two. Do you think we've got this about right? Or is it too much towards nailed down in a so called objective way or too much I line with flexibility? Or is it roughly around the middle.

C: General sense it is in middle. I think there are pointers as well the kind of questions they are asking as a frame to come back to. So yes there is a degree of flexibility but I do like that your decision can either way be guided by a set of guidelines on the whole. Which to me feels more comfortable. to still have a framework of how to use it.

M: that's really helpful. There is a general sense for most people that it's roughly in the middle. Striking that balance right. Any decentres to that always in qualitative research you are looking for the voice that isn't being heard. Anyone saying no you've got it wrong, tear it up start again join the revolution. I don't know what that would be in this context but anyway join it.

C: I think you certainly wouldn't want to be anymore rigid. Because of the nature of the therapy. That's as far as you want to go with prescribing what the session should look like.

M: I have to say on the subject of this, that was my concern when approaching the agenda of doing the scale. My anxiety was I don't want to do something that would in any way push BCT to become more rigid as a result of the measure we are using. It's been known in training people playing to the CTS-R. So it starts to distort what the therapy is because of the measurement tool and I would be horrified if we were doing the same thing here. I'd rather not use the scale. Because you hear it and I know people need to pass courses so they do agenda setting rigidly - And I suspect most people then stop doing it. Especially you - just kidding. once you leave the course. Then it's artificial. That's my concern, but it sounds like you are saying that it's probably at the moment at least it's working out roughly where it should be. Alright another broad think about what you like and dislike. It could be anything. I like the font, we need different paper.

C: In a general sense I really liked filling it out, but then I disliked that I didn't know what everybody else was filling in theirs. I was wondering if I was being score happy here. I would have liked to have known what other people have scored on each one that would have been helpful.

M: That's certainly something we can do. We could plot that out for you. It's only because of time, but we could compare notes here of that process thinking it will take 10minutes. But it usually takes an hour to do it well. So we thought we'd get the results and plot them out so you can see them. I think you are right. You need to calibrate how other people are seeing this, that's the key.

C: One thing that I guess is more of a comment than having a positive or negative to it. In terms of my own learning, in terms of how long it will take me to get used to it. And then when I was filling it out - obviously we didn't have time here. But had I been doing this by myself, I definitely would have gone back and listened to that tape. I don't know how often I would have gone back over to rescore it and I guess that's that uncertainty. And with anything new you are going to get used to it.

kate muse

Although it is flexibility, there are still some guidelines to follow

Isabelle Rudolf von Rohr

Positive comment: wouldn't want to prescribe what the session should look like anymore

kate muse

Flexibility vs. rigidity again.

Isabelle Rudolf von Rohr

positive comment

kate muse

Lack of confidence, am I doing it "right"... relates to issue of familiarity and training needed to use the scale.

Isabelle Rudolf von Rohr

Issues with rating: unsure I was scoring correctly

Isabelle Rudolf von Rohr

Issues with rating: it will probably take me a long time to score one session due to looking through the manual

But yeah I just wonder how quickly that will all come and also time wise, you said for you 20-30min it would take?

M: yeah 20minutes to half an hour and that's once you are familiar with something. If it's a more difficult session longer. And doing exactly what you were just saying and going back over sections and think about it. I think for ourselves this will be the next step once we are happy with the feedback. Training in the scale would look a little different where we do do that step of ok that's where we as a programme team would have rated that so you get a bit of a sense of cutting points, but it feels a bit premature doing that, because we are still feeling our way through it. I wouldn't want to prematurely lock that down by saying we think we would rate this and then you would all agree with that. We'd rather hear from you say no I think you got that wrong and it's not quite getting at the competency that's where we are at the moment but expect that's where we'll go.

C: Does that in itself highlight any discrepancies if in that process itself if there is someone calling it 2 and someone calling it 5 that might indicate that something is not quite right.

M: yeah if you are getting it a lot.

I: yes we will look at inter-rater reliability and the expectation is that people would rate similarly and if there is too much discrepancy the items would probably not capture it very well.

C: What is really good is to have the little prompts in the box, so you don't feel you have to go back to the manual for everything. So you can then be clearer about which items you know you might need to go back to the manual from time to time again until you become more familiar with them.

M: Ok so that works well. The prompts work well. Having to go through a 100 page each and every time is not very practical is it. Exactly.

C: For me it was. I felt actually I did have to go back and I think that probably says more about me and not being used the whole thing and still learning and it being very new. But I felt for me using the prompts and the score sheet maybe I could be marking something quite higher or lower – and going back, I'm really glad I did go back. There were clear instances when I realised I had missed something and did not score high enough or I made an assumption, did score highly and actually it should have been less.

M: So it was important you had that manual by the sounds of it. You wanted those bigger descriptions and questions there. That's really important.

C: yes it was helpful, but I think if I wasn't relying on that I don't think I would have been rating properly. It might be what was being said about the CTS-R that you do need to keep going back...

M: you need to keep going back. Because in my experience, I've been using the CTS-R for over a decade and I find you can build up your own sense of what you think an item should look like. Mine is working with emotional expression. I realised I built up my own ideas of what that should be and going back to the manual I realised I was a little bit of with what they thought it should be. That doesn't mean their wrong and I'm right. Just maybe I need to go back to the manual more often even if you are very familiar with it it's helpful to do that.

Isabelle Rudolf von Rohr

Positive comment: like the anchors on the scoring sheet, so didn't have to refer to the manual all the time

kate muse

Scoring sheet provides useful guidance and reduces the need to use the manual

kate muse

Scoring sheet provides useful guidance and reduces the need for the manual- but danger in this! (e.g. miss things..)

Isabelle Rudolf von Rohr

Clarification; anchor prompts not clear enough, would have marked wrong without referring to the manual

C: Yeah and then you made me think that the short prompts were helpful. Don't ask me how I would improve them, because they need to be improved, but I don't know how. You don't want to end up duplicating the manual. But I felt there some things when I went back that if I hadn't gone back and looked I would have really been off kilter. I was probably off kilter anyway. |

kate muse

As above- Scoring sheet provides useful guidance and reduces the need for the manual- but danger in this! (e.g. miss things.)

Isabelle Rudolf von Rohr

Clarification: therefore anchors need improving

M: yeah it's difficult. We know the more reliable methods is going to be always if you rate the scale with reference to the manual. So we're overshooting from a position when you don't do that. And if you know it real well, you maybe are not going to be having to cross-reference quite so much, but we are wanting you to use the manual as you go through. And obviously if you get stuck you might have to read it in more depth, but we're not ever gonna have the position where the rating form replaces the manual and you don't need to use it. There just needs to be enough to kind of queue you but it's not going to answer all the questions, otherwise you just reproduce the manual like you are suggesting.

C: exactly.

Isabelle Rudolf von Rohr

Positive comment: like that the scale reminds me of BCT competencies

C: This is not the aim of the scale because the scale is obviously there to assess, but is the by-product of learning and getting used to and more familiarised with all the different techniques and all the different points you have to bear in mind when actually doing therapy yourself. So I quite like that. |

kate muse

Useful tool for practitioners delivering BCT

M: Yes. Was it useful to do that in terms of this should also act as the CTS-R does in training. Here's a nice summary of the key principles and reading it ah reminds me this is broadly what a CBT session looks like. Does it tell you that? Does it broadly tell you what a BCT session is all about?

C: yeah. Yeah.

Isabelle Rudolf von Rohr

Clarification: sharing thoughts and feelings is missing in the communication part

C: There is one thing that might be slightly missing maybe in the communication part. There is no mention of the thoughts and feelings which for me is such a huge part. |

kate muse

Missing aspects of competence?

M: And tell us how do you feel about that?

C: I'm not sure I might have to hide in the toilet. But I thought that was a very specific thing which in the tape was sort of happening but not in a very formal way. And in some tapes it's going to be a big part. So I don't know if that needs kind of highlighting a bit more.

M: So highlighting that particular intervention and how well that was done.

C: yes, because it talked about facilitating communication as a couple, but wasn't specific about doing the sharing the thoughts and feelings in session. Which I suppose leaves you flexibility to rate highly when somebody has clearly done that in a previous session and is reviewing that at..

M: Yeah it might not come up at every session.

C: yeah or least have it as an example.

M: I mean often the principles of that intervention are expressed in later interventions as well when you are working on relationship standards you would still expect that it's built on the sharing thoughts and feelings intervention a bit. It becomes a little more implicit and the couple is able to do it. But you would prefer by the sounds of it somehow within the items that to be more explicit.

C: I think so. Just because it gives you more something to look for or hold on to. Some of the wording is quite – there is a lot of syllables- and then it can a bit unsure of what you are looking for. So maybe have the occasional reference to a specific thing where there are specific things. Just have a bit of a prompt, because if you think about using it for your own therapy learning that would be good.

kate muse

Language used in the scale is confusing. Not user friendly

Isabelle Rudolf von Rohr

Clarification: some wording in scale can be confusing

C: It covers it a little bit in the communication section but doesn't label it in quite that way. So it does talk about steering the conversation towards reciprocal listening but it's not labelled as a behavioural intervention.

Isabelle Rudolf von Rohr

Clarification: thoughts and feelings / supervisor (supervisor)

M: I wonder if it's similar.. What sort of advice do we give people if they are going to submit a CTS-R? Often my advice on how are they going to do something which displays most of the competencies we look for. Of course we often say it might be one of those sessions where you do a thought record, because it would give you a lot of prompts. And I wonder if this is picking up on something similar if you are doing the sharing thoughts and feelings exercise in the session it's going to allow you to demonstrate a lot of these things. But not every CBT session is going to be a thought record and similarly not every BCT session is going to be a sharing thoughts and feelings intervention. And if you look at Don Baucom's books one of the traps we can fall into, and one of the traps that mainly beginners fall into is just doing the sharing thoughts and feelings intervention. Just like going at the communication, because it's kind of tricky to get to other stuff. I wonder if we want to avoid that. It's a problem if that's not highlighted, but I wonder if there is something there that we don't want to bias it and we think that BCT is all about sharing thoughts and feelings. That's something clear we can do it and it might help us pass the tape.

C: I don't know. I suppose it's something about some of the language it's a little bit, it's quite hyperbolic.

kate muse

Language used in the scale is confusing. Not user friendly

M: hyperbolic. You have to first look up the dictionary. There is too many syllables here... Sounds painful.

C: I suppose the sharing thoughts and feelings in a way are a means to an end, isn't it? Because normally you then go on to use it as the basis for maybe decision making stuff or. So if we just focus on a.. other than if you were going to having an item about teaching couples skills. Maybe there is room for did you do good teaching.

Isabelle Rudolf von Rohr

Incorporate in scale: add item about how well did you teach the couple skills (supervisor)

M: I think we do need to capture that whether or not that's representative or not. There is something there about the coaching, skills giving aspect of this you've done that well. Again because that broadens the sharing thoughts and feelings a whole new skills that might be the focus of the session. Other things that you liked..

C: I just want to go back to that issue. I agreed with the lady over there who

M: That's not a lady that's Kit..

C: Maybe some specific examples or some strategies she could use to maybe jog your memory when you are using the rating scale. But essentially that stands the risk of then as a therapist to just be using those strategies all the time and not really adapting it. So I kind of liked the open-endedness of

it and the questions are a little bit in the scale phrased to make you think and not just as a check box exercise.

M: yeah we don't want it to be just a check box exercise. Again narrowing it down what we understand as interventions to be and then what our trainees are doing in the sessions. We don't want that. Other things that you like don't like?

C: I quite liked the short prompts. I don't think I would like anything to be added, because otherwise I think then I would look at it as a check box of have they prioritized, have they done this. And if they haven't I would score them down as opposed to thinking about it in more detail. So I quite like the short prompts.

M: quick prompts. Yeah it's going to be more functional.

C: yeah. I liked it.

M: thank you. Anything about the, just the way it's all set out? The use of the scoring is up to 6 the lay-out of it. The quality of it as a tool? Whether you feel comfortable with that? Familiar with that? Because it's a choice. We don't have to do it like that. We don't have to do it as a six. There's a range of other ways. From traffic light systems to all sorts. Does it feel fit for purpose? Does it feel fine? You are all CBT therapists and you like counting. Anything else? The layout of the thing?

C several: yeah. It's fit for purpose.

C: I guess if it was me if I could fold it out in a form. One of those fold out papers and everything was on the one page. But that's quite hard to do.

M: you are demanding.

C: I'm just saying.

M: It sounds actually quite nice. If we did have it as a fold out, you could get more information that way. I can imagine sitting there listening to it and it might be a bit easier.

C: So you just don't have to turn over the page. It might be easier

M: that could be improved with maybe a design guide sitting down and helping how that works out. Or if we publish it and make loads of money maybe we could get people to do that. How do you think in terms of the scale you might be using in supervision? Do you think it would be a helpful thing you could imagine using to inform your supervision? Or receiving supervision? Is that something that might facilitate how you approach supervision? Or make the whole process worse?

C: I definitely feel like it would be constructive feedback as opposed to aimlessly ripping apart the session. I think it will feel quite nice to receive it in that way.

M: Ok so from a trainee point of view it will be quite good to get this sort of feedback and that will be helpful?

C several: yes definitely.

Isabelle Rudolf von Rohr

Positive comment: liked anchor points

Isabelle Rudolf von Rohr

Scale provides opportunity for constructive feedback

kate muse

Scale provides useful formative function

M: great. How about we've got some supervisors in terms of a tool to facilitate other people's learning does it feel it's going to be useful for that? Or getting in the way of that?

S: yes it will be useful.

C: Because I think you can think about it on a micro- and macro level. And the micro-level is the different scales all the 15 items. And thinking about what those items could be. But also the three domains, when you are giving feedback. It could be this bit is very good. The structure or whatever is very good, but what's missing is or the weaker part is the intervention part. And then you can focus on what part of the intervention is perhaps needs a little bit of thinking about. So I think you can use it in many ways|

M: So the structure of the three domains is helpful

C: I guess there might be interpersonal conflict and it might be very helpful because there might be areas they might disagree on and this can become a reference point to both of them.

M: absolutely. I can't imagine you having interpersonal conflicts with your supervisees. They do what you tell them to do. But yeah having an external reference point rather than just your opinion or reaction this is what you are saying is quite good.

C (supervisor): And I guess we won't be able to listen to all the tapes and all the sessions. So the supervisee could come with their own scale having rated it. Saying look I think this item I have done really badly. Both ways.|

M: yeah. This is the intended use. This is what you must have done in CBT training. We do want people to be rating themselves and in supervision I think part of the thing you do is you help them adjust the way they are rating themselves. A lot at the start are way off and are either rating themselves Oh I think I am really good at this, and they are not where they think they are, but that's less common than oh I am absolutely hopeless at this and I am really struggling with this in your eyes. And actually you are not as bad. I think simply because of the nature of couple work and couple's are often in a lot of distress as a therapist you are thinking oh I am doing terrible job. And the feedback might be the couple is really struggling and this is really hard work, but you are not doing a terrible job. It's just not easy and it's going to remain difficult for quite a while by the sounds of it but you are doing a reasonable job.

C: yes giving that feedback with the scale, will make it more objective and concrete doesn't it. It's not just a reassurance. You have actually rated it on all these things even though it was a nightmare session.|

M: Definitely.

Izzy: Do you think is there anything missing from the scale? Anything that you think would be helpful for it to be in there for it to be a useful tool and to capture and measure the competencies and any aspects of competencies.

C: I wonder if there could be something in the scale to capture the essence of the session.

M: like a small meter or something?

Isabelle Rudolf von Rohr
Positive comment: objective useful feedback

kate muse
Scale provides useful formative function

Isabelle Rudolf von Rohr
Positive comment: helpful tool if interpersonal conflict with supervisor

kate muse
Scale provides useful formative function

Isabelle Rudolf von Rohr
Positive comment: do self-ratings on the scale (self-reflection)

Isabelle Rudolf von Rohr
Positive comment

kate muse
Scale provides useful formative function

Isabelle Rudolf von Rohr
Incorporate in scale: add item about the essence of the session – global rating

C: hehe. Indeed. What flavour?

Izzy: What do you mean with essence?

C: So the essence. In general what the session has been like. Here we are looking at specifics. But in general did it feel like an appropriate session.

kate muse
The need for a global rating.

Izzy: like a global rating?

M: Would that be a rating or would that be a qualitative description?

Isabelle Rudolf von Rohr
Incorporate in scale: could rate the essence on a 6 point rating scale again – from both view points

C: You could rate on a point 6 scale again. You could rate whether this session felt you know it was a comfortable session and the session worked within the principles of BCT. Some sessions I only work with elements of CBT or ACT. So it could be that you might only work with elements of BCT but does it really truly reflect a whole BCT session.

kate muse
The need for a global rating.

M: so it sounds like there are a couple of things you are saying. First a global rating of is this a good session of BCT broadly, but you're also wanting room for some sort of qualitative feedback, some sort of summary of what this session kind of felt like. This is what it was like to listen to that or this is my sense of that session would be helpful.

C: From the raters point of you?

C: From both. I think both is a good way to then compare and is a good learning opportunity if two say how might we review this.

C: Of course there could be a mismatch between competence level and how the session felt in all fairness. Especially early on in couple work with highly distressed couples and a lot of couple hostility. It might not feel that pleasant to listen to.

M: Or if the therapist. Let's say the couple is not that distressed and they come in after they've just had a chat and they felt real comfortable. The therapist didn't do any BCT but it felt nice. I like the couple they were chatting, but he didn't actually do anything.

C: Would it make sense to put in context of the whole treatment So is progress being made, is that a logical step that's been taken given what's gone on before.

Isabelle Rudolf von Rohr
Incorporate in scale; consider the context of treatment within the scale

M: Normally what we would do certainly for a formal rating. You would ask some background information, what session number was this, what were the aims of the session, what's the formulation for the couple, so you have that contextual thing. And if you are the rater you would also have some self-rating from the trainee and their view of this was my aim for this session and this is how broadly I felt that went. So we get that comparison.

C: But I guess it's not in the scale I suppose. I just wondered.

M: Yeah that's interesting we've kind of tacked from the CTS-R, we've tacked those bits on. It's not a formal property of the scale, but it's extra contextual information. But yeah maybe we could think about it maybe possible to incorporate that somewhere.

C: yeah I think that would be helpful at the end bit when you summarise if you had a prompt there to kind of think about what level of competence, where they are the therapist and whether they have just qualified. Just to add a bit of context. That might give feedback or makes the feedback a little bit more richer. But I know you probably would write that in there, but to kind of give a prompt

M: It's a nice prompt though. If we expand this you certainly don't want the ratings to be done very quickly without that sort of keeping in mind the trainee: where they at etc.

kate muse

The need for a global rating.
Also the need for contextual information about the therapist.

Isabelle Rudolf von Rohr

Incorporate in scale: helpful to know context: where therapist was in his training and also where they were in therapy

Appendix Q: HRA ethical approval letter



Health Research Authority

Miss Isabelle Rudolf von Rohr
80 Haydock Road
Bicester
OX26 1BG

Email: hra.approval@nhs.net

14 September 2016

Dear Isabelle,

Letter of HRA Approval

Study title: Scale evaluation: Behavioural Couple Therapy Scale -
Depression Version (BCTS-D)
IRAS project ID: 199914
Sponsor Royal Holloway, University of London

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendix R: RHUL self-certified ethics review



Ethics Review Details

You have chosen to self certify your project.	
Name:	Rudolf von Rohr, Isabelle (2013)
Email:	PAVA061@live.rhul.ac.uk
Title of research project or grant:	Scale evaluation: BCTS-D
Project type:	Royal Holloway postgraduate research project/grant
Department:	Psychology
Academic supervisor:	Helen Pote
Email address of Academic Supervisor:	h.pote@rhul.ac.uk
Funding Body Category:	No external funder
Funding Body:	
Start date:	03/10/2016
End date:	30/06/2017

Research question summary:

Behavioural Couples Therapy (BCT; Baucom, Epstein, La Taillade & Kirby, 2008) is a talking therapy, which has been included in the NICE guidelines to treat depression (2009) and is being rolled out as part of the national IAPT (Increasing Access to Psychological Therapies) programme. This study aims to evaluate if a scale rating the competency of students training to become behavioural couples therapists, is psychometrically sound (Behavioural Couple Therapy Scale – Depression Version (BCTS-D)). Additionally, it will be evaluated if the users find it a credible and plausible measure.

The project will collect its data from the Post Qualification course in BCT at the Central London training centre for CBT. The study is supervised and monitored by the key developer (Don Baucom) and a team of BCT supervisors (Michael Worrell, Sarah Corrie). Further, this project is part of a Doctorate of Clinical Psychology at Royal Holloway and is supervised by Dr.Helen Pote.

The scale evaluation project mainly consists of three parts: (1) Review already collected data within a training workshop for BCT supervisors. Analyse inter-rater reliability and gain some insight on content validity, usability and how to improve the scale (2) Establishing the reliability and validity of the scale within a real world BCT training context (3) Feedback study to evaluate face validity, content validity and usability.

Research method summary:

The therapists' competences within the BCT training course are routinely assessed on three different occasions during their yearlong training. As part of the training programme both BCT supervisors and BCT students will have familiarised themselves with the rating scale, and the students have three therapy tapes rated on the BCTS-D. Both students and supervisors will be asked to complete a short demographics questionnaire and a short questionnaire about the scale to provide feedback on the content validity for each item and the usability of the scale overall.

The students will be asked for their consent to use their anonymised data within a research context. As this study involves NHS staff participating in a training course capacity to consent will be assumed. Informed consent will be sought with the use of a detailed participant information sheet and consent form. It will be emphasised orally and in the participant information that not consenting to take part in the study will in no way affect their academic standing with the course.

As part of routine treatment practice patients sign a consent form at the start of therapy, which allows the student therapist to record the session and use it in supervision, and to use anonymised data in further research.

Given the nature of this study, it is highly unlikely that participants will suffer harm by taking part. However, if the questionnaire happens to include any questions which, for whatever reason, the participant does not wish to answer then the question can be omitted.

Risks to participants

Appendix S: Summary sheet of therapeutic context

Ages	
Genders	
Presenting Problems	
Number of sessions Was the end planned / unplanned?	
Assessment (briefly summarise) Standardized measures used Report scores of CSI long form and all minimum data set scores for both partners at intake	
Formulation (attach formulation diagrams or descriptive/narrative summaries that have been used in therapy to the Log Book. Make sure you remove any identifying information)	
Intervention (outline specific BCT interventions that were used)	
Outcome (include an excel spread sheet of session by session progress scores including CSI, PHQ-9 and GAD-7)	
Critical Review (Reflect briefly on what you learned through your work and supervision on this case)	