

Developing and evaluating valid, reliable and usable measures of assessing competence in Cognitive Behavioural Therapy

Katherine (Kate) Patricia Muse

Green Templeton College, Oxford

Submitted for the degree of D.Phil.

Trinity Term 2014

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. Freda McManus for her many years of unflagging professional and personal support and encouragement. I am particularly grateful to Freda for her astute and pragmatic academic guidance, for her refreshing honesty, and for believing in me and giving me the confidence to do the same. I would also like to acknowledge with much appreciation the crucial role of Dr. Sarah Rakovshik, who stepped in as my supervisor at the eleventh hour. Sarah helped me to navigate the red tape and political hurdles I encountered in conducting my research and provided me with invaluable guidance in how to draw my research together into a coherent thesis.

I am grateful to the many therapists and supervisors who took time out of their busy schedules to participate in my research, particularly the staff at the Oxford Cognitive Therapy Centre and the First Step IAPT service. I was also fortunate to receive financial support from the British Association of Behavioural and Cognitive Psychotherapies and I am grateful to them for providing me with this opportunity.

I am lucky to have had a great deal of encouragement from my friends and family, which has gone a long way to helping me finish my thesis. My friends and colleagues from the department of psychiatry (Isabelle Rudolf Von Rohr, Dr. Felicity Cowdrey, Annabelle David, Adele Krusche, and Suzanne Straebler) deserve a special mention as they have been a constant source of support, help and guidance over the last three years. I would also like to thank my friends Hannah Heerey, Sarah Lloyd-Moss, and Natalie Moxey for providing me with much needed distraction from time to time. Last, but certainly not least, I am incredibly grateful to my parents (Sue and Mike Muse) and to my partner Alex Young for their patience, understanding and moral support.

Thesis Overview

Tools for measuring competence in delivering Cognitive Behavioural Therapy (CBT) provide a means of assessing the training of new CBT therapists and ensuring the quality of treatment provision within routine practice, provide a framework for delivering formative feedback, promote ongoing self-reflection, and are essential to establishing treatment integrity in research trials. As such, identifying an optimal strategy for assessing the competence with which CBT is delivered is crucial to the continued progression of the field. However, research in this area has been somewhat limited to date. Thus, there are at present no evidence-based best practice guidelines outlining the way CBT competence should be assessed. Furthermore, many of the assessment measures currently available have been widely criticised, indicating a need for improved tools for assessing CBT competence. To begin addressing this issue, the first two chapters of this thesis focus on reviewing and evaluating current assessment methods. Chapter one provides a systematic review of current methods of assessing CBT competence and chapter two outlines a qualitative exploration of experts' understandings and experiences of assessing CBT competence. Findings from these studies provide tentative recommendations for practitioners and researchers assessing CBT competence. These initial studies also highlight ways in which the assessment of CBT competence could be improved and therefore provide a platform for guiding subsequent thesis chapters which focus on further developing existing assessment measures. Specifically, chapters three to six focus on the development and evaluation of a novel CBT competence rating scale: the Assessment of Core CBT Skills (ACCS). The ACCS builds upon currently available scales (especially the Cognitive Therapy Scale- Revised: CTS-R) to provide an assessment framework for assessors to deliver formative and summative feedback regarding therapists' performance within observed CBT treatment sessions and for therapists to rate and reflect on their own performance. Development of the ACCS involved three key stages: 1- theory-driven scale development (chapter three), 2- an 'expert' review of the content validity, face validity, and usability of the scale (chapter four), and 3- an evaluation of the scale involving a pilot study examining its psychometric properties (chapter five) and a focus group examining its usability and utility (chapter six). Results from these studies indicate that the ACCS is a useful learning tool, is easy to use, has good psychometric properties, and offers an acceptable alternative to the CTS-R. Finally, chapter seven examines whether assessors require training in how to use the ACCS, concluding that simply reading the ACCS manual may be sufficient to achieve acceptable levels of reliability and usability. The results from the thesis are then drawn together in the final concluding comments in chapter eight, which discusses the findings within the broader context of the assessment of CBT competence.

Table of Contents

	Page No.
Chapter One: A systematic review of methods for assessing competence in Cognitive Behavioural Therapy	5 - 72
Chapter Two: Expert insight into the assessment of competence in Cognitive Behavioural Therapy: A qualitative exploration of experts' experiences, opinions and recommendations	73 - 152
Chapter Three: The development of a novel scale for assessing competence in the delivery of Cognitive Behavioural Therapy: The Assessment of Core CBT Skills (ACCS)	153 – 226
Chapter Four: Obtaining feedback from expert and novice clinicians: Examining content validity, face validity and usability of the ACCS	227 – 333
Chapter Five: A pilot study examining psychometric properties of the Assessment of Core CBT Skills (ACCS) scale	334 – 403
Chapter Six: A focus group evaluating the usability and utility of the Assessment of Core CBT Skills (ACCS) scale	404 – 479
Chapter Seven: Is assessor training necessary? Examining the impact of training on usability and inter-rater reliability of the Assessment of Core CBT Skills (ACCS) rating scale	480 - 537
Chapter Eight: Concluding comments	538 - 543

CHAPTER ONE:

A Systematic Review of Methods for Assessing Competence in Cognitive Behavioural Therapy [1]

Chapter Overview

Effective assessment of Cognitive Behaviour Therapy (CBT) competence is crucial to the success of the current drive to expand CBT training and service provision and to the widespread dissemination of CBT into routine practice. However, a lack of consensus about how CBT competence should be assessed has resulted in the use of numerous different methods, many of which have been widely criticised. This review describes and evaluates the various methods of assessing CBT competence. A systematic literature search identified 64 articles pertaining to a method of assessing competence in the provision of standard CBT interventions to adults experiencing mental health problems. Ten methods for assessing CBT therapist competence were identified from these articles and are presented within Miller's (1990) framework for assessing clinical skill. The advantages and disadvantages for each method are examined in relation to reliability, validity and feasibility. Recommendations for assessing therapist competence are made and areas that warrant future research are highlighted.

[1] A modified version of this chapter has also been published:
Muse, K., & McManus, F. (2013). A systematic review of methods for assessing therapist competence in cognitive behavioural therapy. *Clinical Psychology Review*, *33,* 484–499.

Contents

Introduction 8

What is CBT Competence? 9

Method 10

Inclusion Criteria 10

Literature Search Procedures 11

Results 15

(1) Level 1: Knowledge-Based Assessments 15

1.1 Multiple choice questionnaires. 15

1.2 Essays. 17

(2) Level 2: Assessments of Practical Understanding 18

2.1 Short-answer clinical vignettes. 18

2.2 Case reports. 18

(3) Level 3: Assessments of Practical Application of Knowledge (Skill) 19

3.1 Standardised role-plays. 20

(4) Level 4: Clinical Practice Assessments 25

4.1 Assessor-rated treatment sessions. 25

4.2 Supervisory assessments. 33

4.3 Therapist self-assessment. 34

4.4 Patient surveys. 35

4.5 Patient outcome. 36

Discussion 37

Summary of Assessment Methods 37

Further Psychometric Evaluation 38

A Multi-method Approach to Assessment 39

Feasible Measures 39

Implementation Protocols 39

Benchmarks 40

Formative and Summative Assessment 40

Conclusions 41

References 42

Appendices 55

Appendix 1. Literature Search Strategies 55

Appendix 2. Summary of Studies Employing Assessments of Knowledge and Practical Understanding of CBT 57

Appendix 3. Summary of Studies Employing Assessments of Clinical Practice 61

Introduction

Significant progress has been made in developing evidence-based psychological treatments for a variety of disorders and problems. Prominent amongst these treatments is Cognitive Behavioural Therapy (CBT) which has been shown to be effective in treating a wide range of psychological disorders (Butler, Chapman, Forman, & Beck, 2006; Stewart & Chambless, 2009). Identifying an optimal strategy for assessing the competence with which CBT is delivered is important to the continued progression of the field for a number of reasons. First, the strong evidence base for CBT and the high economic burden of untreated mental illness has prompted increased demand for the dissemination of CBT (Barlow, Levitt, & Bufka, 1999; Clark, 2011; McHugh & Barlow, 2010). For example, the UK has seen large-scale government investment in the Increasing Access to Psychological Therapies dissemination program (Clark et al., 2009; Department of Health, 2008) and a number of other European countries are considering similar initiatives (Berge, 2011). Similarly, in the USA the Veterans Health Administration recently embarked on a significant program promoting the implementation of evidence-based psychological treatments, such as CBT (Karlin et al., 2012). Effective methods of assessing CBT competence are essential to the success of such dissemination programs as they provide a means of assessing the training of new CBT therapists and ensuring the quality of treatment provision within routine clinical practice (McHugh & Barlow, 2010; Rakovshik & McManus, 2010).

Second, competence assessment plays a crucial role in the empirical evaluation of CBT as research trials cannot draw valid conclusions regarding the efficacy of CBT protocols unless the competence with which the protocols are delivered can be established (Waltz, Addis, Koerner, & Jacobson, 1993; Weck, Bohn, Ginzburg, & Ulrich, 2011b). Third, as evidence suggests that therapist competence may play a role in determining treatment outcomes (Brown et al., 2013; Ginzburg et al., 2012; Karlin et al., 2012; Kuyken & Tsivrikos, 2009; Shaw et al., 1999; Strunk, Brotman, DeRubeis, & Hollon, 2010; Trepka, Rees, Shapiro, Hardy, & Barkham, 2004), assessment of CBT competence could provide a vehicle for ensuring that CBT is optimally effective for patients. However, reviews examining the relationship between CBT therapists' competence and patient outcome conclude that results are variable and the use of unreliable methods of assessing CBT competence has been suggested as one possible explanation for this (Crits-Christoph et al., 1991; Perepletchikova & Kazdin, 2005; Webb, DeRubeis, & Barber, 2010). Hence improved understanding of the assessment of CBT competence could facilitate much needed future research examining the association between competence and outcome in the domain of CBT and may provide further insight into the 'active ingredients' responsible for the relationship (Dobson & Singer, 2005). Finally, effective measurement of CBT competence is necessary to provide

targeted feedback regarding therapists' strengths and weaknesses in order to promote self-reflection and guide future learning (Bennett-Levy, 2006; Laireiter & Willutzki, 2003; McManus, Rosen, & Jenkins, 2010).

Despite the importance of effective measurement of CBT competence, a lack of consensus regarding the way in which CBT competence should be assessed has resulted in the development of multiple different assessment methods, many of which have been widely criticised (e.g. Barber, Sharpless, Klostermann, & McCarthy, 2007; McGlinchey & Dobson, 2003; Sharpless & Barber, 2009; Waltz et al., 1993). Hence this review aims to provide a description and evaluation of existing methods for assessing CBT competence. This will enable recommendations about the most effective methods to be made and will identify priorities for future research into the development of reliable, valid and cost-effective methods of assessing CBT competence.

What is CBT Competence?

Before examining the utility of different methods for assessing therapist[2] competence in CBT it is necessary to clarify what is meant by the term competence. In line with Barber, Sharpless, Klostermann, and McCarthy's (2007) concept of 'limited-domain intervention competence' and Kaslow's (2004) notion of 'intervention competence', competence is defined as the degree to which a therapist demonstrates the general therapeutic and treatment-specific knowledge and skills required to appropriately deliver CBT interventions which reflect the current evidence base for treatment of the patient's presenting problem. While professional knowledge and skills (e.g., ethical practice) are recognised as important aspects of therapist competence in any treatment modality, these are not the focus of the present review. Furthermore, as the specific knowledge and skills that constitute competence vary according to therapeutic domain (Sharpless & Barber, 2009; Waltz et al., 1993), it is also necessary to define CBT competence. Pilling and Roth (2007) identify the 'core competences' required to deliver effective CBT and their framework outlines over fifty specific competences grouped within five domains: generic therapeutic competences (e.g., knowledge of mental health problems, ability to engage patient); basic CBT competences (e.g., knowledge of CBT principles, ability to explain CBT rationale); specific CBT techniques (e.g., guided discovery, use of thought records); problem-specific competences (e.g., interventions outlined in disorder-specific treatment manuals); and metacompetences (e.g., capacity to select and apply most appropriate CBT method) . Given the

[2] The term 'therapist' is used throughout to refer to appropriately qualified mental health professionals from any professional background (e.g., nursing, clinical psychology, psychiatry etc.) delivering a CBT intervention.

number and varying levels of competences outlined, the authors' note that in its current format the framework cannot be used to assess competence. Instead they suggest that competence measures identify and focus on assessing a subset of core competences (Roth & Pilling, 2008). However, the framework does give a comprehensive definition of what CBT competence is.

In reviewing CBT competence assessment, it is also necessary to differentiate the related concepts of adherence and competence. Adherence refers to the extent to which the therapist delivers the intervention as outlined in the relevant treatment model / manual, whilst competence refers to the skill with which the interventions were implemented. The successful delivery of CBT requires both adherence to the treatment model and competent delivery of the specified techniques (Dobson & Singer, 2005; McGlinchey & Dobson, 2003). Thus competence is not sufficient without adherence (e.g., skilful delivery of techniques from a different treatment model) in the same way that adherence is not sufficient without competence (e.g., rigidly following a CBT treatment manual without any skill). Although conceptually distinct, in practice there is much overlap between adherence and competence (McGlinchey & Dobson, 2003; Waltz et al., 1993). Hence there is uncertainty as to whether adherence and competence should be measured independently or concurrently, and this may depend on the context of the assessment. For example, assessing adherence and competence independently is necessary in clinical research trials (Moncher & Prinz, 1991; Perepletchikova & Kazdin, 2005). However, in the context of therapist training and routine practice it is recommended to consider both components in combination (Dobson & Singer, 2005; Fairburn & Cooper, 2011). As this review focuses on the assessment of CBT competence, examining methods of assessing adherence alone is beyond the scope of the study. The present review will therefore examine methods of assessing competence both independently and in combination with adherence, but not those which measure adherence alone.

Method

Inclusion Criteria

To be included in the current review an article must have met the following criteria: (a) not a review or commentary, (b) publication in English, (c) pertains to a quantifiable CBT competence measure (with or without adherence) administered in English, (d) relate to individual face-to-face CBT (i.e. not group, family, online CBT) for an adult population experiencing mental health problems.

Literature Search Procedures

Peer-reviewed articles published between January 1980 and December 2011 were searched using PsychInfo, Scopus, Science Direct, Psycarticles, Web of knowledge, Medline and PubMed databases. Databases were searched using the following broad strategy: any of the following words relating to competence: "therapist competence/y", "therapeutic competence/y", "clinical skill", "clinical competence/y", "assessing competence/y", "competence/y assessment" or "intervention competence/y" in combination with one or more of the following words relating to CBT: "CBT", "cognitive behavior/ur", "cognitive behavior/ural","behavior/ural therapy", "behavior/ur therapy" or "cognitive therapy" (see Appendix one for full search strategies).[3] After removing duplicated references, the database searches identified 952 articles. A further 11 articles were identified using 'snowball' methods including examining reference lists, key author searches and related article searches (Greenhalgh & Peacock, 2005). A two-stage process for selecting relevant articles was employed independently by two reviewers (Kate Muse and Dr Freda McManus), with the final selection of papers for inclusion being made by agreement between the two reviewers (Centre for Reviews and Dissemination, 2009). First, titles and abstracts of all articles were screened in relation to inclusion criteria, 836 of which were excluded as 'clearly irrelevant' (e.g. competence to stand trial, not CBT, non-adult population etc.). Second, full copies of the remaining 127 studies were retrieved and assessed for eligibility, 64 of which met inclusion criteria (see Figure one)[4].

[3] Additionally an email alert system was set up for each of the searches, so that any articles published between December 2011 and May 2014 were brought to the attention of the author. No novel assessment measures were identified within these articles and thus none were added to the review.

[4] Studies included in the review are marked with an asterisk in the reference list.

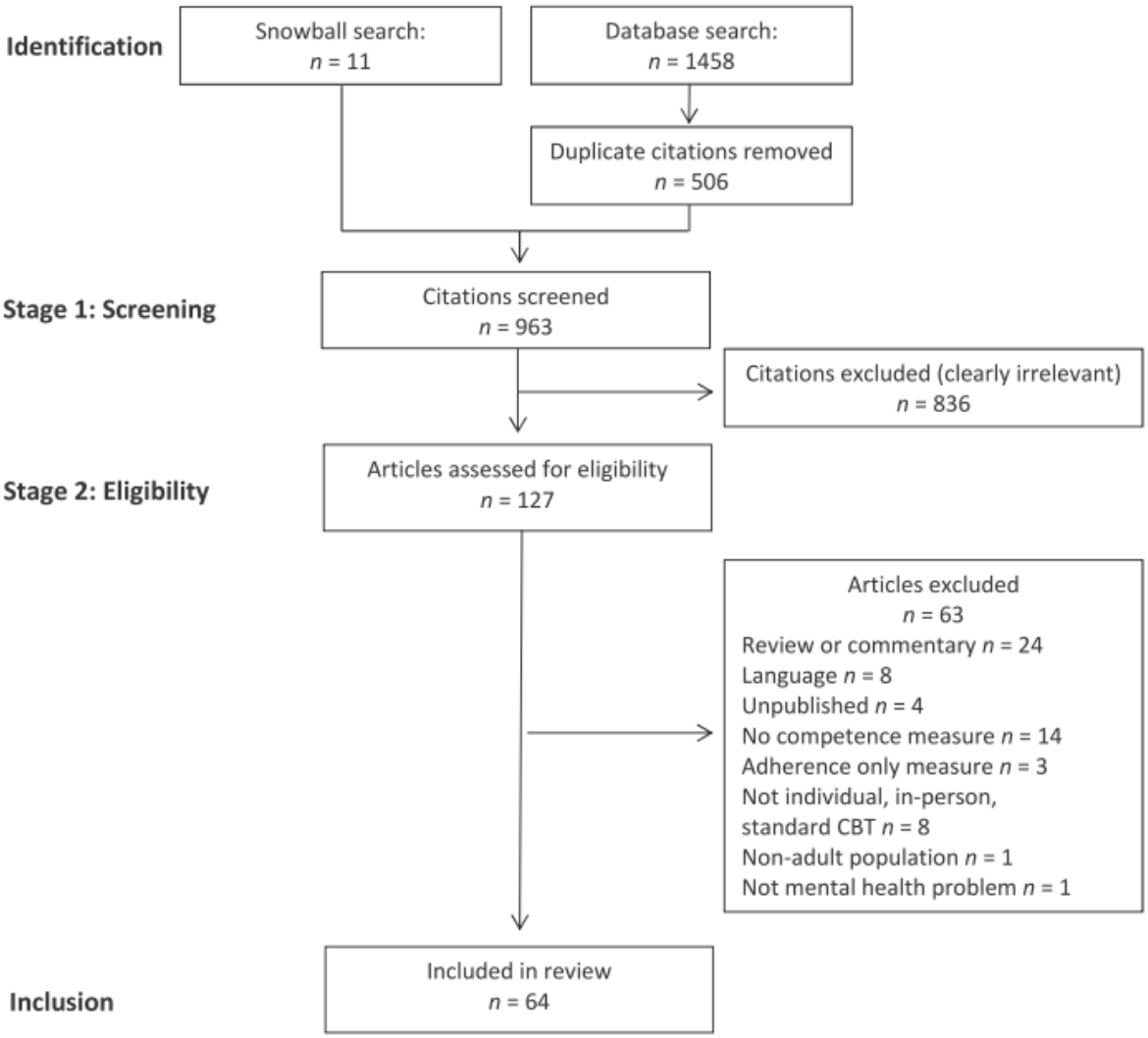


Fig. 1. Flow diagram of study selection procedures.

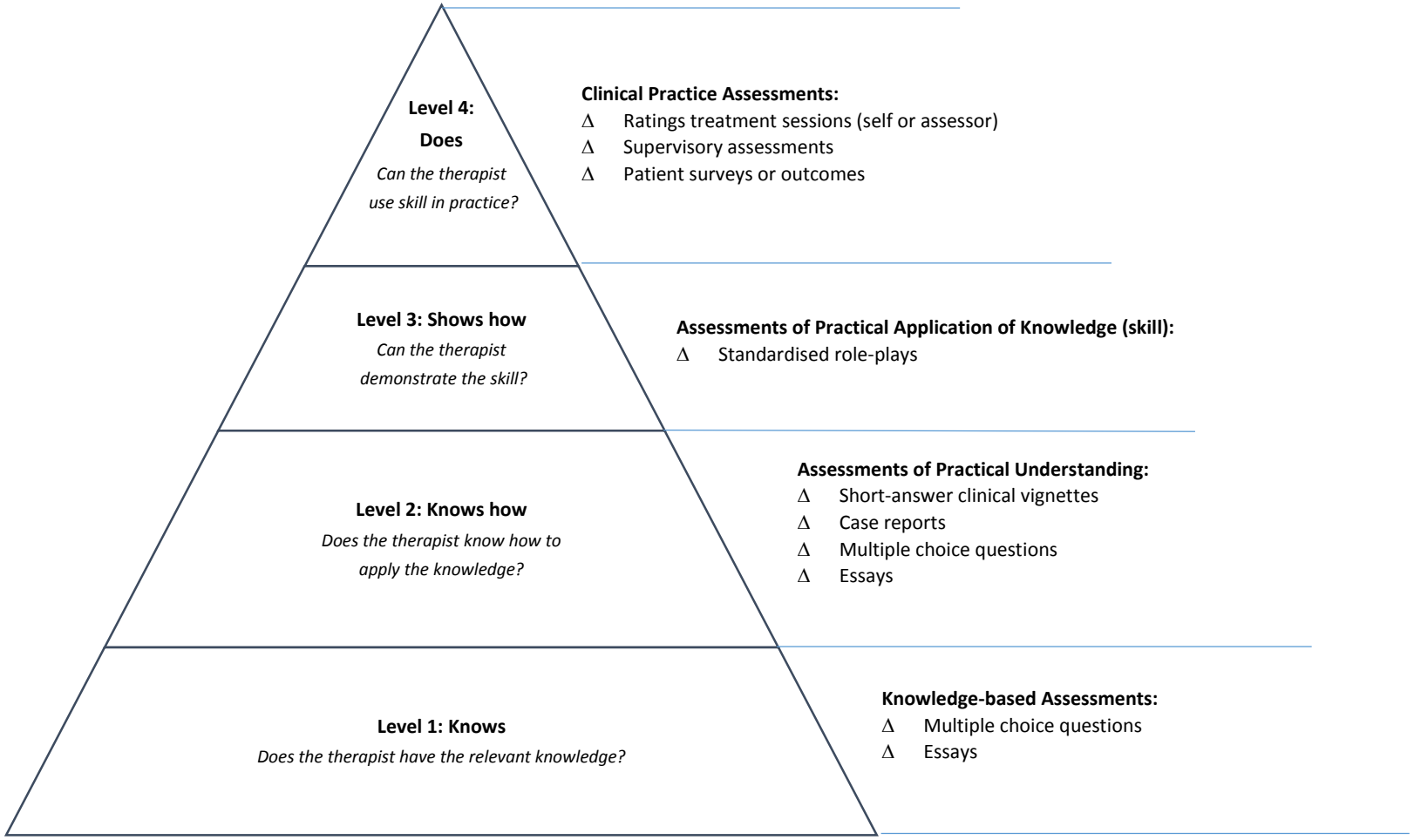


Figure 2. A framework for CBT therapist competence measures, based on Miller's (1990) clinical skills hierarchy

Table 1
Operational Definition of Terms from Barker, Pistrang, and Elliot (2002).

Term	Operational definition
Reliability	The extent to which a measure is free from errors and thus is able to reproduce consistent results from repeated samples and different researchers over time.
a. Internal consistency	An assessment of inter-item reliability based on the correlation amongst individual items in the measure, used to determine the degree to which items assess the same underlying construct. Usually measured using Cronbach's alpha (α), which ranges from 0 = *items independent* to 1 = *items identical*, with a minimum of .80 - .85 typically representing satisfactory agreement (Bland & Altman, 1997).
b. Inter-rater reliability	The level of agreement between different assessors' ratings on an observational measure, based on the correlation of scores between or among two or more assessors. Typically assessed using Pearson's correlation coefficient (*r*) and Intraclass correlation (ICC). The following benchmarking scale is often used to interpret agreement coefficients: < 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).
c. Test-retest reliability	The degree to which the measure produces consistent results over time, based on the correlation between observations with the same individuals on at least two occasions. Typically assessed using Pearson's correlation coefficient (*r*) (see benchmarks above).
Validity	The extent to which the scale measures what it purports to measure.
a. Face validity	Whether the measure includes appropriate items that are a credible and plausible measure of the construct.
b. Content validity	Whether the measure adequately samples the different aspects of the construct specified in its definition.
c. Predictive validity	Whether the measure is correlated with a future criterion, i.e. whether the measure predicts improvements in patients' symptoms.
d. Discriminant validity	Whether the measure discriminates between those with / without the construct. Information regarding the ability of the measure to discriminate between CBT and other forms of psychotherapy is included. However, as it is possible to adhere to a treatment protocol without delivering the treatment competently, it is possible for a measure to accurately reflect the treatment modality but fail to recognise whether the treatment was delivered in a competent manor.
e. Sensitivity to change	The capacity of a measure to assess meaningful change in the construct it measures. Also referred to as 'responsiveness'.
Feasibility	How easily the measure can be implemented (i.e. issues such as resources, time, cost, and expertise required).

Results

The 64 articles included in the review identified ten methods of CBT competence assessment, which are presented below within Miller's (1990) hierarchical framework for assessing clinical skill (see Figure two). While it is recognised that there is overlap between categories, Miller's (1990) framework separates skills assessments into four hierarchical levels: knowledge (knows); practical understanding (knows how); skills (shows how), and clinical practice (does). Each assessment method is described briefly and then evaluated in terms of reliability (internal, inter-rater, test-retest), validity (face, content, predictive, discriminant) and feasibility (see Table one for operational definitions).

(1) Level 1: Knowledge-Based Assessments

The basic foundation of CBT competence is a sound understanding of the scientific, theoretical and contextual basis of CBT (Roth & Pilling, 2007). In line with prevailing opinion that cognitive knowledge is best assessed through written test forms (Downing, 2002), therapists' knowledge of CBT is typically assessed using multiple choice questionnaires (MCQs) and essays (see Appendix two for a summary of studies employing knowledge assessments). As well as assessing knowledge of basic CBT facts, MCQs and essays can be used to assess knowledge of the clinical application of CBT and thus can also be considered assessments of practical understanding. For parsimony these methods will however only be presented within this section.

1.1 Multiple choice questionnaires.

Therapists' knowledge of the theory and application of CBT can be assessed using MCQs. MCQs commonly comprise of a statement followed by a true-false option or a lead-in question followed by a number of responses, from which one or more correct answer(s) is selected. Although the cost in developing MCQs can be high, once developed they are cheap, quick and easy to implement, do not require expert assessors, put little burden on therapists and can assess a large range of content. MCQs have been used to measure CBT knowledge acquisition following training (Decker, Jameson, & Naugle, 2011). However, training providers typically create their own MCQs to assess specific course content (e.g. Maunder, Milne, & Cameron, 2008; Myles & Milne, 2004; Sholomskas et al., 2005; Weingardt, Cucciare, Bellotti, & Lai, 2009; Weingardt, Villafranca, & Levin, 2006; Westbrook, McManus, Clark, & Bennett-Levy, 2011) and the use of standardised questionnaires is rare. The only standardised MCQs identified were the cognitive therapy awareness scale (CTAS: Wright et al., 2002) and cognitive behavioural therapy knowledge quiz (CBT-KQ: Myles, Latham, & Ricketts, 2003). The CTAS contains 40 true / false statements

relating to basic CBT constructs (e.g., description of thought records). Despite having been designed for assessing patients' understanding of CBT, it has recently been used in US psychiatry residency programs to measure medical students' CBT knowledge (Sudak, Beck, & Wright, 2003). The utility of the CTAS as a measure of therapist competence has not yet been examined (see Fujisawa et al., 2011 for examination of a Japanese version of the scale). However, given its intended use with patients, the items assess therapists' recognition of relatively basic CBT concepts and thus it is a relatively superficial CBT knowledge assessment. The CBT-KQ includes 26-items, each with four response options, and provides a more in-depth assessment of CBT knowledge. Knowledge assessed includes general CBT issues, theoretical underpinnings of behavioural and cognitive approaches and the practice of CBT. Although the CBT-KQ has only been used in one published study (Kennedy-Merrick, Haarhoff, Stenhouse, Merrick, & Kazantzis, 2008), promising pilot data for the CBT-KQ (Myles et al., 2003) and an earlier version of the questionnaire (the foundation cognitive behavioural therapy multiple choice questionnaire [FCBT-MCQ: Myles & Milne, 2001]) have been presented at conferences.

There is a need for further development of standardised MCQs with demonstrated validity and reliability to assess CBT knowledge. Future development in this area would need to address a number of issues. First, it is important to develop questions with sufficient contextual detail (e.g. case vignettes) to enable assessment of higher-order cognitive processes which are characteristic of clinical practice, such as understanding, evaluation and application of knowledge, rather than simply testing recall of basic facts which are removed from clinical reality (Case & Swanson, 2001). Second, the ability to obtain a correct answer due to recognition (i.e. 'cueing') would need to be addressed. For example, this could be addressed by increasing the number of possible responses or using extended matching items (several questions, all with the same long list of possible answers) (Schuwirth & van der Vleuten, 2004). Third, the bank of questions would need to be large enough to enable different questions to be given pre- and post-training, thus reducing practice effects. Fourth, although potentially costly at the developmental stage, it may be beneficial to explore the use of computer-based delivery and administration as this may increase ease of implementation (e.g. automatic electronic marking) and enable the inclusion of media within questions (e.g. pictures, video clips, clinical findings, case presentations etc.) (Clauser & Schuwirth, 2002; Mills, Potenza, Fremer, & Ward, 2002). Finally, this review did not identify any studies examining the relationship between MCQ performance and treatment outcome. However, a recent conference presentation reported a significant relationship between therapists' CBT-KQ scores and their patients' post-treatment depression scores (r = -.38, p <. 02) (McManus, Keen, et al., 2010). Given that performance on MCQs has not been found to correlate

with patient outcome in other therapeutic domains (e.g., Chevron & Rounsaville, 1983), further investigation of the degree to which therapists' MCQ performance relates to clinical skill and ultimately treatment outcomes is needed to establish the predictive validity of MCQs.

1.2 Essays.

Essay questions are better suited to the assessment of higher-order cognitive processes and allow for more contextualised answers than MCQs (Epstein, 2007; Schuwirth & van der Vleuten, 2004). As they require effortful retrieval of information in order to construct an answer, they also preclude the ability to obtain a correct answer due to recognition (Schuwirth & van der Vleuten, 2004) and promote improved memory for content (Larsen, Butler, & Roediger, 2008). Hence training courses routinely use essays both to enhance and consolidate learning and to determine trainees' ability to review theoretical and empirical CBT literature, critically evaluate it and link it to clinical practice (e.g. Keen & Freeston, 2008; McManus, Westbrook, Vazquez-Montes, Fennell, & Kennerley, 2010). Essays range in length from three to ten thousand words, and are usually rated on a 0 – 100 scale by an experienced therapist (> 50% representing a pass). A number of pitfalls regarding the use of essay questions as a method of assessing competence have been highlighted, the most significant of which are relatively low reliability and high financial and labour costs (Epstein, 2007; Schuwirth & van der Vleuten, 2004). That being said, evidence to date indicates that essays which have been used to assess CBT knowledge generally result in acceptable levels of reliability (Keen & Freeston, 2008; McManus, Keen, & Freeston, under review). This may in part be due to the use of structured marking procedures which are thought to improve reliability, such as clear making criteria, multiple markers and independent monitoring of the grading process (Leigh et al., 2007). More than one essay is however needed to provide a reliable assessment of a therapists' knowledge of CBT, with research indicating a minimum requirement of one examiner marking between five and seven essays, or two examiners marking between four and five essays (results from McManus et al. [2012] and Keen and Freeston [2008] respectively). Despite the need for multiple essays and expert assessors, McManus et al. (under review) estimate a cost per therapist of between £198 (one examiner/five submissions) and £395 (two examiners/five submissions), suggesting that essays may still provide a cost-effective means of assessing therapists' CBT knowledge. The present review did not identify any studies examining the relationship between therapists' essay performance and patient outcome. However, a recent conference presentation by McManus, Keen, et al. (2010) reported a significant relationship between therapists' essay scores and their patient's post-treatment depression scores (r = -.18, p <. 05), thus providing preliminary evidence of the predictive validity of essay examinations.

(2) Level 2: Assessments of Practical Understanding

A competent therapist must not only have a broad understanding of CBT theory and research but must also have the ability to use this knowledge to inform when, how and why CBT interventions should be implemented (Miller, 1990; Roth & Pilling, 2007). Essays and MCQs which are grounded in clinically relevant contextual information can be used to assess therapists' understanding of the clinical application of CBT. Practical understanding can also be assessed using short-answer clinical vignettes and case reports (see Appendix two for a summary of studies assessing practical understanding).

2.1 Short-answer clinical vignettes.

Practical understanding of CBT can be assessed using open-ended, short-answer questions about the treatment of patients presented in written, audio or visual vignettes. Although restricted in breadth, depth and complexity of clinical information, vignettes ensure the provision of standardised, replicable information. Using open-ended questions also precludes the ability to correctly answer questions due to recognition and allows in-depth, complex responses. The downside of this however is increased difficulty producing reliable results and higher assessment costs (Schuwirth & van der Vleuten, 2004). Although short-answer clinical vignettes are a widely used training tool, they are rarely used as formal assessments of therapist competence. The present review identified only one standardised short-answer clinical vignette: the Video Assessment Task (VAT: Myles & Milne, 2004). The VAT involves presenting therapists with a video recording of a scripted dialogue describing the difficulties typically experienced by a patient with panic disorder. Therapists are asked to answer three questions regarding (a) symptom identification, (b) ascertaining the problem and (c) consideration of up to six CBT techniques to use in treatment. The psychometric properties of the VAT have not yet been formally examined, although Myles and Milne (2004) report very good inter-rater reliability for all three questions (a: r = .97, b: r = 1.0, c: r = .94).

2.2 Case reports.

Case reports are written reports in which a therapist explains and reflects upon the process of assessment, formulation and treatment with an individual patient. They are usually marked by an experienced therapist on a 0 – 100 scale (>50% indicating adequate competence) in relation to therapists' ability to combine knowledge of CBT theory and research within clinical practice, to recognise and conceptualise clinical problems, and to identify and reflect upon a relevant treatment program. Case reports form part of the UK CBT therapist accreditation process and are often used as a teaching and assessment tool in CBT training settings (e.g. Barnfield,

Mathieson, & Beaumont, 2007; Department of Health, 2011; Keen & Freeston, 2008; McManus et al., under review; McManus, Westbrook, et al., 2010). Through promoting self-reflection and providing therapists with detailed feedback, case reports provide a useful formative assessment tool. There are, however, a number of limitations to using case reports as a basis of formal, summative assessment of CBT competence. First, therapists' ability to achieve high marks may be dependent on a number of patient factors (e.g. patient difficulty, complexity etc.) which have been shown to impact on therapist performance in psychotherapy (Foley, O'Malley, Rounsaville, Prusoff, & Weissman, 1987; Strunk et al., 2010). This could disadvantage therapists working in complex treatment settings and may result in therapists selectively presenting 'straightforward' cases. Second, reliability of case reports is relatively low, with studies reporting that between 49.8% (McManus et al., under review) and 80.5% (Keen & Freeston, 2008) of score variability is accounted for by measurement error. Third, it is estimated that a minimum of either one examiner assessing between four and 16 reports or two examiners assessing between three and 12 reports (results from McManus et al. [2012] and Keen and Freeston [2008] respectively) are necessary to reliably assess an individual therapists' competence, resulting in a cost per therapist of between £175 and £1050 (McManus et al., under review). Fourth, studies have reported no significant progression on case report scores following CBT training (Keen & Freeston, 2008; McManus, Westbrook, et al., 2010), suggesting that they may not be sensitive to the impact of training. Fifth, although the present review did not identify any published studies investigating the predictive validity of case reports, a recent conference presentation reported no significant relationship between therapists' case report scores and patients' post-treatment depression scores (McManus, Keen, et al., 2010). Hence, it is possible that the ability to write about CBT in the form of case reports does not correspond to the ability to deliver effective CBT. In sum, both the reliability and validity of this method have yet to be established and future research could explore whether this could be achieved through further refinement of implementation and evaluation procedures, such as the use of combined written and oral presentations (Petti, 2008).

(3) Level 3: Assessments of Practical Application of Knowledge (Skill)

A sound understanding of CBT and its clinical application is necessary but not sufficient in demonstrating competence. It is also essential that a therapist has the skills to draw on and apply this knowledge in clinical situations (Roth & Pilling, 2007). Miller (1990) distinguishes between skill-based assessments which occur within carefully constructed artificial clinical simulations ('shows how') and those which occur within real clinical practice settings ('does' independently in practice). The following section focuses on assessments using clinical simulations, namely standardised role-plays, with clinical practice-based assessments being reviewed in section four.

3.1 Standardised role-plays.

A role-play is an artificial simulation of a clinical scenario in which a therapist interacts with an individual playing the role of a standardised patient. Therapists' ability to carry out performance-based tasks (e.g. setting an agenda or completing a thought record) within the role-play are then rated by an observer, either via 'live' observation or from a recording, using pre-defined competence criteria. Medical training settings routinely use role-plays to assess students' clinical competence in the form of objective structured clinical examinations (OSCEs), a series of short encounters with standardised patients that are viewed and assessed by an independent observer (Epstein, 2007; McNaughton, Ravitz, Wadell, & Hodges, 2008). Although role-plays are also widely used within CBT training settings, these are largely informal methods of learning and practicing new skills, rather than formal assessment tools. The only standardised role-play identified by the present review was used by Sholomskas et al. (2005) to evaluate the impact of different training interventions on therapists' ability to demonstrate key CBT interventions. The role-play consisted of a one hour treatment scenario with a scripted patient played by an experienced therapist. The session was recorded and therapists' ability to demonstrate techniques outlined in the CBT for substance abuse treatment manual was rated by a trained independent observer using items drawn from the Yale Adherence and Competence Scale (YACS: Carroll et al., 2000, see section four below). The validity and reliability of this role-play has, however, not been established. Although little progress has been made in developing standardised role-plays, Fairburn and Cooper (2011) argue that this method may provide the most sensitive, focused and practical assessment of the application of CBT skills. Role-plays can be used to assess a broad range of skills with patients of varying difficulty (including 'extremes' such as aggression or hostility) and provide standardised assessments which can be easily replicated (e.g. pre- and post-training) and implemented with trainees who are not yet working clinically. Furthermore, role-plays overcome a number of practical difficulties faced by clinical practice-based assessments (e.g. informed consent, service constraints etc.). However, there is a danger that role-plays may lead to the simplification of clinical scenarios and create a less authentic subjective experience, thereby reducing the degree to which the behaviour of the therapist and the individual playing the patient is representative of clinical practice (Sharpless & Barber, 2009). These issues may be particularly pertinent when involving the use of computer technology, such as presenting pre-recorded clinical scenarios or using virtual reality interactions. Furthermore, the feasibility of role-plays is undermined by the substantial time, expense and resource commitment required both in development and implementation (Kaslow et al., 2009).

Table 2

Summary of Assessor-rated Measures of CBT competence within Treatment Sessions

Scale	**Skills assessed**	**Items**[1]	**Rating**
Cognitive Therapy Scale: CTS (www.beckinstitute.org)	General therapeutic skills	1. Agenda 2. Feedback 3. Understanding (previously empathic skills) 4. Interpersonal effectiveness 5. Collaboration 6. Pacing/use of time	0 = *poo*r to 6 = *excellent* scale with item-specific descriptions for even numbered anchors Range: 0 - 66
	CBT specific skills (transdiagnostic)	7. Guided discovery (previously empiricism) 8. Focus on key cognitions or behaviours 9. Strategy for change (previously conceptualisation) 10. Application of cognitive-behavioural techniques 11. Homework [Global skill rating] [Additional considerations: ability to deal with special problems, factors justifying a departure from the standard therapeutic approach, patient difficulty rating]	
Cognitive Therapy Scale-Revised: CTS-R (Blackburn et al., 2001)	General therapeutic skills	1. Agenda 2. Feedback 3. Collaboration 4. Pacing/use of time 5. Interpersonal effectiveness	0 = *incompetent (non-compliance)* to 6 = *expert (compliance + high skill)* scale with item-specific descriptions for each anchor Range: 0 - 72
	CBT specific skills (transdiagnostic)	6. Guided discovery 7. Conceptualisation 8. Identifying key cognitions 9. Application of change methods 10. Application of behavioural techniques 11. Homework 12. Facilitation of emotional expression	

Scale	Skills assessed	Items[1]	Rating
Cognitive Therapy Adherence and Competence Scale: CTACS (Barber et al., 2003)	General therapeutic skills: a. Structure	1. Agenda 2. Mood check 3. Bridge from previous session 4. Inquired about problem 5. Homework review 6. Homework assignment 7. Capsule summaries 8. Patient summary/ feedback 9. Focus/ structure	Amalgamation of quality rating from o = *poor* to 6 = *excellent* and appropriateness rating from 0 = *none* to 6 = *thorough,* or *not applicable* with item-specific descriptions for even numbered anchors Range: 0 - 126 (adherence is rated on a separate 0 - 6 adherence scale)
	b. Collaborative therapeutic relationship	10. Socialisation 11. Warmth/ genuineness/ congruence 12. Collaboration	
	c. Development & application of case conceptualisation	13. Eliciting automatic thoughts 14. Eliciting core beliefs and schemas 15. Eliciting meaning/ understanding/ attributions 16. Addressing key issues 17. Case conceptualisation	
	CBT specific skills[2]	18. Guided discovery 19. Asking for evidence/ alternative views 20. Use of alternative techniques (specified by assessor) 21. Overall performance rating	
Multicentre Collaborative Study for the Treatment of Panic Disorder- Global Competence Item: MCSTPD-GCI (Huppert et al., 2001)	Global skill (panic disorder)	1. Global competence	1 = *clearly inadequate*, 3 = *fair,* 5 = *good*, 7 = *excellent* Range: 1 - 7

Scale	Skills assessed	Items[1]	Rating
Cognitive Therapy Scale-Psychosis: CTS-Psy (Haddock et al., 2001)	General therapeutic skills (as applied in psychosis) CBT specific skills (as applied in psychosis)	1. Agenda 2. Feedback 3. Understanding 4. Interpersonal effectiveness 5. Collaboration 6. Guided discovery 7. Focus on key cognitions 8. Choice of intervention 9. Homework 10. Quality of intervention [Overall rating]	Six micro skills are provided for items 1 to 9, each of which is marked as 1 = *present* / *appropriately omitted* or 0 = *absent*, thus allowing for between 0 and 6 points per item. Item 10 is rated on a 1 = *barely acceptable level of skill* to 6 = *excellent* scale. Range: 0 - 60
Manual-Assisted Cognitive Behaviour Therapy Rating Scale: MACT-RS (Davidson et al., 2004)	Skill in delivering self-harm treatment techniques	1. Structure 2. Pacing 3. Collaboration 4. Appropriate techniques 5. Skilful execution of techniques 6. Helpfulness of session 7. Empathy 8. Client/problem difficulty 9. Linking sessions 10. Using the manual 11. Homework assignments	Each item is rated on a 1 - 7 scale with item-specific anchors provided at the low, mid and high scale points. Range: 11 – 77

Scale	Skills assessed	Items[1]	Rating
Yale Adherence and Competence Scale (YACS: Carroll et al., 2000)	Substance use disorder general therapeutic skills a) assessment	1. Assess alcohol use 2. Assess cocaine use 3. Assess other substances 4. Assess psychopathology 5. Assess general functioning	Each item is first rated on a 0 = *not at all* to 5 = *extensively* adherence scale and, where the item did occur in the session, a 0 = *not at all* to 5 = *extensively* competence rating is completed. Range: 0 - 105
	b) general support	6. Praise patient efforts 7. Explore feelings 8. Explore level of family support 9. Optimistic reassurance 10. Show natural spontaneity	
	c) goals of treatment	11. Explore patient's treatment goals 12. Increase discrepancy between behavior and goals 13. Assess commitment to abstinence 14. Reflective listening 15. Feedback about urine results	
	Substance use disorder CBT specific skills	16. Skills training 17. Debrief past high risk situations 18. Cognitions 19. Plan for future high risk situations 20. Difference between slip vs. relapse 21. Conditioning	

[1] Where variations occur, information is presented from the most up to date version published by the scale authors. Items presented in [square brackets] are not typically included within the total numerical score.

[2] Items examining acceptance/respect, attentiveness, accurate empathy and sharing the conceptualisation with the patient were removed from the final version of the CTACS.

(4) Level 4: Clinical Practice Assessments

In assessing CBT competence, it is important to establish that therapists are able to use independent judgment and critical thinking to appropriately and effectively deliver CBT within the cultural and organisational context of clinical practice settings (Miller, 1990; Roth & Pilling, 2007). There are a number of methods of assessing therapist skill within clinical practice: (1) assessor-rated treatment sessions, (2) supervisory assessments, (3) therapists' self-assessments, (4) patient surveys, and (5) patient outcome (see Appendix three for a summary of studies employing assessments of clinical practice).

4.1 Assessor-rated treatment sessions.

CBT competence is often assessed by observing one or more treatment sessions, either 'live' or via session recordings, and then rating the skill with which CBT is delivered using a standardised rating scale. The present review identified seven scales (see Table two) which are presented below under two broad categories: (1) transdiagnostic scales: those which focus on competences which are not specific to any one diagnosis and (2) disorder-specific scales: those which focus on competences required to deliver disorder-specific protocols. A brief discussion of the individual scales will be followed by an overall evaluation of the utility of this assessment method.

4.1.1 Transdiagnostic scales.

The most recent version of the Cognitive Therapy Scale (CTS, or Cognitive Therapy Rating Scale: CTRS, www.beckinstitute.org) is an 11-item scale (see Table two for scale summary). However, the CTS has undergone a number of modifications, including a 13-item version (Trepka et al., 2004), study-specific versions (Brosan, Reynolds, & Moore, 2007; Brosan, Reynolds, & Moore, 2008), shortened versions (Bryant, Simons, & Thase, 1999; Ryum, Stiles, Svartberg, & McCullough, 2010; Williams, Moorey, & Cobb, 1991) and use of the 'global rating' item alone (Borge et al., 2008; Hoffart, Borge, Sexton, & Clark, 2009). The CTS is derived from the cognitive therapy for depression competency checklist (CCCT: see appendix in Beck, Rush, Shaw, & Emery, 1979)[5] and has been criticised for being overly focused on the CBT for depression protocol (Barber et al., 2007). However, as well as being widely used within the context of depression (Bryant et al., 1999; Dimidjian et al., 2006; K. S. Dobson, Shaw, & Vallis, 1985; Friedman et al., 2004; Hollon et al., 1992; Jacobson et al., 1996; Jacobson & Gortner, 2000; Jarrett et al., 1999;

[5] Although the CCCT checklist has been employed as a measure of competency in a small number of studies (e.g. Bouchard et al., 2007; Pace & Dixon, 1993; Scholey & Woods, 2003), it has largely been superseded by the CTS and is therefore not presented independently within the review.

Lopez & Basco, 2011; Shaw et al., 1999; Simons, Gordon, Monroe, & Thase, 1995; Simons et al., 2010; Simons & Thase, 1992; Strunk, Brotman, DeRubeis, & Hollon, 2010; Trepka et al., 2004; Vallis, Shaw, & Dobson, 1986; Vallis, Shaw, & McCabe, 1988; Ward et al., 2000), the scale has also been used to assess the competence of CBT therapists working with patients with psychosis (Durham et al., 2003; Garety et al., 2008; Sensky et al., 2000), social phobia (e.g. Borge et al., 2008; Hoffart et al., 2009), antisocial personality disorder (Davidson et al., 2009) and generalised anxiety disorder (Westra, Arkowitz, & Dozois, 2009) and is routinely used by training courses examining CBT delivery across a range of disorders (Forand, Evans, Haglin, & Fishman, 2011; Keen & Freeston, 2008; McManus et al., under review; McManus, Westbrook, et al., 2010; Milne, Baker, Blackburn, James, & Reichelt, 1999; Williams et al., 1991).

A CTS score above 39 is commonly considered a minimum competence threshold. This cut-off is however somewhat arbitrary as it was calculated as one standard deviation below the mean score of a group of CBT therapists rated on the original 11-item CTS (Shaw et al., 1999; Shaw & Wilson-Smith, 1988). Hence further research is required to establish an empirically proven 'tipping point' for competence. Furthermore, the same cut-off has been applied across different scale versions, resulting in a more lenient competence threshold for higher item versions. It has been suggested that using mean item scores, rather than total CTS scores, may limit variability in competence thresholds and facilitate comparison across different CTS versions (Rakovshik & McManus, 2010).

The CTS has demonstrated a high degree of internal consistency (α range = .84 - .95 [Dobson, Shaw, & Vallis, 1985; McManus, Westbrook, et al., 2010; Shaw et al., 1999]). The scale was originally divided into 'general' and 'specific cognitive therapy' subscales (Vallis, Shaw, & Dobson, 1986). However, research has demonstrated that all items load highly onto both subscales, the two scales are highly correlated (Vallis et al., 1986) and there is a high degree of item convergence (Dobson et al., 1985). Factor analysis revealed different 'skill' and 'structure' subscales, but the 'skill' factor accounted for 64.8% of score variance and included all but 3 items (Vallis et al., 1986). Other studies have evaluated 'general interview procedure', 'interpersonal effectiveness' and 'specific CBT techniques' subscales (McManus, Westbrook, et al., 2010; Trepka et al., 2004; Westbrook, Sedgwick-Taylor, Bennett-Levy, Butler, & McManus, 2008), which are also highly inter-correlated (Trepka et al., 2004). Hence the CTS appears to measure one construct, with rationally defined subscales which are not independent. Trepka et al. (2004) found a significant correlation between CTS total scores and self-rated depression outcomes in a treatment completing sample and Shaw et al. (1999) report that the 'structure' subscale accounted for a significant amount of variance in clinician-rated depression, thought this did not

generalise to patient-ratings. Although these studies provide some support for the predictive validity of the CTS, further research is needed, especially in disorders other than depression.

CTS scores discriminate between CBT sessions rated as high quality (valid/acceptable) vs. low quality (unrepresentative/unacceptable) (Vallis et al., 1986) and have been shown to increase following CBT training (e.g. McManus, Westbrook, et al., 2010; Simons et al., 2010; Westbrook et al., 2008; Williams et al., 1991). Inter-rater reliability for total CTS scores is however variable, ranging from poor (ICC = .01 - .08 [Jacobson & Gortner, 2000]) and moderate agreement (ICC = .47 - .59 [McManus, Westbrook, et al., 2010; Vallis et al., 1986]), through to good (ICC = .69 - .77 [Crits-Christoph et al., 1998; McManus, Westbrook, et al., 2010; Strunk et al., 2010]) and very good agreement (ICC = .83 - .94 [Dimidjian et al., 2006; Lopez & Basco, 2011; Westra et al., 2009]; *r* = .82 - .94 [Brosan et al., 2008; Dobson et al., 1985; Williams et al., 1991]). Similarly, inter-rater reliability for individual CTS items ranges from poor to very good agreement (ICC = .27 - .56, *r* = .40 - .87 [K.S. Dobson et al., 1985; Vallis et al., 1986; Williams et al., 1991]). There are a number of possible explanations for this variability, including differences in the number of raters, use of aggregate ratings, differing levels of CBT expertise held by assessors and varying degrees of training in using the CTS (see Appendix three for a summary of studies employing CTS). Difficulties establishing inter-rater reliability may also in part be due to differing levels of inference required to rate CTS items (Whisman, 1993) and inadequate definition of scale points (Blackburn et al., 2001). Additionally, the CTS has been criticised for item overlap, multiple concepts being addressed by single items and neglecting key competences (Blackburn et al., 2001; Whisman, 1993), thus undermining the content validity of the scale.

Blackburn et al.'s (2001) 12-item revision of the CTS (cognitive therapy scale-revised: CTS-R, see Table two for scale summary) sought to overcome the original scale's limitations by providing more specific anchor descriptions, incorporating adherence in addition to competence, adding a facilitation of emotional expression item and reducing overlap by collapsing three general skill items into one 'interpersonal effectiveness' item (Milne, Claydon, Blackburn, James, & Sheikh, 2001). The authors also initially added a 'charisma/flair' item and an optional 'non-verbal behaviour' item but these were later removed (Reichelt, James, & Blackburn, 2003). A validated competence threshold score has not yet been established for the CTS-R, though a minimum standard of 36 has been widely adopted in the field (James, Blackburn, & Reichelt, 2001). The CTS-R demonstrates high internal consistency (α range = .75 - .97 [Blackburn et al., 2001; James, Blackburn, Milne, & Reichfelt, 2001; Reichelt et al., 2003]) and thus appears to measure one construct. The scale developers report good inter-rater reliability for the initial 13-item CTS-R (average ICC = .63 [Blackburn et al., 2001]) and inter-rater reliability for the final 12-

item CTS-R ranging from moderate without rater training (*r* = .44) to good following rater training (*r* = .67) (Reichelt et al., 2003). However, Gordon (2006) reports only fair agreement (ICC = .38) and a discrepancy of more than five points in over half of rating pairs. Inter-rater reliability for a number of individual CTS-R items also remains poor (ICC = -.14 - .84 [Blackburn et al., 2001]), even following rater training (pre-training *r* = .07 - .59, post-training *r* = .26 - .62 [Reichelt et al., 2003]). In sum, initial evidence indicates that the CTS-R may not have fully overcome limitations inherent in the CTS. However, direct comparison between the CTS and CTS-R as well as further examination of the psychometric properties of the CTS-R outside the scale developers are required before firm conclusions can be drawn.

The cognitive therapy adherence and competence scale (CTACS: Barber, Liese, & Abrams, 2003) is a 21-item scale (see table two for scale summary) derived from the CTS and the CBT for cocaine abuse manual (Beck, Wright, Newman, & Liese, 1993) to assess competence in the provision of CBT for cocaine abuse. The authors propose that the scale can also be used to assess CBT more broadly, but the utility of the scale has not yet been evaluated outside of substance abuse. Only two studies carried out by the scale developers, in the context of CBT for substance abuse, have examined the utility of the CTACS. These studies report high internal consistency (α range = .93 - .96 [Barber, Foltz, Crits-Christoph, & Chittams, 2004; Barber et al., 2003]) and inter-rater agreement ranging from good to very good for total scores (ICC = .73 - .94 [Barber, Foltz, Crits-Christoph, & Chittams, 2004; Barber et al., 2003]) and fair to very good for individual items (ICC = .36 - .92 [Barber et al., 2003]). The CTACS also discriminates between CBT and supportive-expressive dynamic therapy and drug counselling (Barber et al., 2004). However, the CTACS total score is confounded by inclusion of an 'overall competence' item. Moreover, competence scores are an amalgamation of two ratings: quality of intervention delivery (0 = *poor* to 6 = *excellent*) and appropriateness of the intervention (0 = *none* to 6 = *thorough* or *not applicable*). The high correlation between these ratings (rs > .90 [Barber et al., 2003]) indicates that they assess the same construct and thus undermines the validity of retaining both ratings. Hence further refinement of the CTACS as well as further examination of the utility of the scale may be needed.

4.1.2 Disorder-specific scales.

As transdiagnostic scales assess competences which underpin most CBT interventions it has been argued that they fail to assess the disorder-specific strategies and procedures thought to be central mechanisms of change in CBT (Fairburn & Cooper, 2011). Hence in recent years a number of disorder-specific scales have been developed to establish the competence with which specific CBT protocols were delivered. First, Huppert et al. (2001) describe a single-item measure

of competence in the delivery of CBT for panic disorder (Multicentre Collaborative Study for the Treatment of Panic Disorder- Global Competence Item, MCSTPD-GCI)**.** The validity and reliability of the MCSTPD-GCI has not been examined. However, Huppert et al. (2001) report that therapists with above and below average patient outcomes were rated similarly, which the authors suggest may be because a single item lacks the sensitivity to adequately assess competence. Second, the Manual-Assisted Cognitive Behaviour Therapy-Rating Scale (MACT-RS: Davidson et al., 2004) is an 11-item scale which assesses competence in the delivery of manual-assisted CBT for the prevention of para-suicide. The MACT-RS demonstrates good inter-rater reliability (r = .66) and there is some evidence of predictive validity in that MACT-RS scores were significantly associated with change in observer-rated patient outcome, although this was only apparent at the 12-month follow-up and did not extend to reduction in self-harm episodes or patients' self-ratings (Davidson et al., 2004). The 'client problem/difficulty' item appears to confound patient difficulty and therapist competence, with scores ranging from 1 *ineffective not due to client* to 7 *client difficult but doesn't appear to be due to therapist's lack of skill*. It is unclear, for example, where a competent therapist whose patient was not difficult would fit on the scale. Additionally, the item which rates whether the patient appeared to find the session helpful ('helpfulness of session'), may be confounded by factors other than therapist competence (e.g. the patient may be highly motivated to please the therapist). Third, the Yale Adherence and Competence Scale (YACS: Carroll et al., 2000) is a 21-item measure of therapist adherence and skill in the implementation of behavioural treatments for addiction which includes three 'general' subscales assessing competences common across modalities and three 'treatment specific' scales assessing competences specific to clinical management, twelve step facilitation and CBT for substance use (Carroll, 1998). The YACS has received little psychometric evaluation within the context of CBT, although it has demonstrated good inter-rater reliability for the general subscales (ICC = .80 - .85) and CBT subscale (ICC = .88) and is able to differentiate between CBT and CM / TSF treatments (Carroll et al., 2000). In sum, all of these scales have received limited psychometric evaluation or use outside of the research setting for which they were developed. Hence the utility of the MCSTPD-GCI, YACS and MACT-RS as measures of therapist competence is uncertain.

A more widely used disorder-specific scale is the Cognitive Therapy Scale-Psychosis (CTS-Psy: Haddock et al., 2001), an adaptation of the CTS designed to assess competence in the delivery of CBT for psychosis. The CTS-Psy uses a modified scaling system in which six micro skills are provided per domain, each of which carries one point and is marked as being present, absent or appropriately omitted. The scale focuses on the same general CBT skills outlined in the CTS, although the authors removed items assessing case conceptualisation and pacing/efficient use of

time and divided the application of CBT techniques into two items assessing the choice and quality of CBT interventions (See Table four for scale summary). Gordon (2006) suggests that the CTS-Psy may be useful for work with other complex cases where differences in the nature of psychological processes (e.g. instability of mood and arousal) similarly impact on the way CBT is delivered and thus uses the CTS-Psy in the context of borderline personality disorder as well as psychosis. The CTS-Psy has been shown to discriminate between CBT and supportive counselling (Lewis et al., 2002; Tarrier et al., 2004; Valmaggia, Van der Gaag, Tarrier, Pijnenborg, & Sloof, 2005) as well as therapists who have vs. have not received cognitive-behavioural training (Haddock et al., 2001). The scale authors also report very good inter-rater agreement for total scores (e.g. ICC = .93 - .94 [Haddock et al., 2001; Lewis et al., 2002; Tarrier et al., 2004] and r = .99 [Valmaggia et al., 2005]) and moderate to very good agreement for individual items (e.g. ICC = .41 - .95; Haddock et al., 2001). Results from Haddock et al. (2001) must however be interpreted with caution as the study is limited by small sample size (N = 14) and inclusion of therapists with a low level of skill (only 8/14 therapists achieved a score >50%). Furthermore, contradictory findings are reported by Gordon (2006), who found only fair inter-rater agreement (ICC = .28) and a discrepancy between markers of >5 in 55% of rating pairs. The superiority of the CTS-Psy over the CTS in assessing competence in the delivery of CBT for psychosis has also been questioned. Gordon (2006) for example reports high agreement in the average score obtained by CBT therapists on the CTS-R and CTS-Psy (r = .79, p <.001), suggesting the scales measure the same construct. Moreover, the CTS-Psy focuses on the same generic CBT skills outlined in the CTS and thus does not assess strategies specific to CBT for psychosis. Hence the CTS/CTS-R and CTS-Psy are viewed by some as being equally applicable within the context of CBT for psychosis (Fowler, Rollinson, & French, 2011).

Additionally, promising new research examining protocol specific competence assessment scales is currently being undertaken by David M Clark and colleagues (personal communication). In sum, although the use of disorder-specific scales holds promise, scale development is at an early stage. Hence further research is needed to establish the reliability and validity of these scales before they can be used more widely. Additionally, although disorder-specific scales may be relevant for use within controlled research settings where patients are selected on the basis of diagnosis, questions have been raised as to the applicability of diagnosis-specific scales within wider clinical practice, such as training settings in which clinicians learn how to deliver a variety of CBT protocols or clinical practices where therapists work with patients experiencing a range of mental health problems (Barber et al., 2007).

4.1.3 Evaluation of assessor-rated treatment sessions.

Using standardised scales to evaluate therapists' skill within treatment sessions has the potential to be an effective method of assessing therapist competence. All of the multi-item scales provide a total score identifying overall skill as well as a score for each individual item. Hence they can be used to establish whether therapists have reached a recognised standard of competence (useful for accreditation, evaluating the overall impact of training, selecting trial therapists) and to provide detailed feedback as to therapists' strengths and weaknesses (useful for personal development, supervision and examining skill acquisition). As standardised measures, they also enable comparison (e.g. between different training courses or treatment trials) and replication (e.g. pre- and post-training). There are however are number of issues which undermine the utility of this competence assessment method. A number of these relate to the scales which are used to rate competence. The ability of many of the scales to provide valid and reliable measures of competence outside of the controlled research conditions for which they were developed has not yet been established. Difficulties defining and disentangling competence have also resulted in poor inter-rater reliability, especially for individual items. This can be improved upon by using an aggregate score of multiple assessors' ratings (Vallis et al., 1986) and by standardising assessors' interpretation of scale items through the provision of training (Barber et al., 2007; Gordon, 2006; Reichelt et al., 2003). However, financial and time constraints mean it is not always feasible to obtain ratings from multiple trained assessors. Hence improved definitional clarity and item refinement may also be necessary. There is a lack of consensus about whether adherence and competence should be measured concurrently or independently. Some scales only assesses the competence with which therapists deliver CBT (e.g. CTS, CTS-PSY MACT-RS, MCSTPD-GCI), others assesses adherence and competence separately within the same scale (e.g. CTACS, YACS) whilst the CTS-R assesses both competence and adherence within the same scale. Although conceptually distinct, in reality there is a more complex hierarchical relationship between the constructs, with adherence being a pre-requisite for the competent delivery of CBT, while adherence does not necessarily imply competent delivery (Waltz et al., 1993). Drawing a distinction between adherence and competence may therefore not be meaningful in assessing therapist competence, as what matters is that therapists are "doing the right things well" (p.384, Fairburn & Cooper, 2011). The validity of the scales is also undermined by the fact that most take limited account of the therapeutic context, such as stage of therapy and key contextual issues such as patient difficulty rarely being taken into account within scoring systems (Waltz et al., 1993). There is also a balance to be struck between including too few items, risking failing to adequately capture the multi-faceted nature of competence, and including too many items,

resulting in item overlap. Further research could usefully examine the minimum degree of specificity required, although this may to some degree depend on the reason why competence is being assessed. A less comprehensive scale may for example be adequate for accreditation purposes, whilst a more detailed scale may be required to provide feedback or examine skill acquisition. Finally, there is a lack of empirically grounded competence threshold scores above which it can be confidently concluded that a therapist competently delivered CBT.

There are also two key problems relating to the implementation of these scales. First, it is not known what level of independence and expertise is required by those making the ratings. Dennhag, Gibbons, Barber, Gallop, and Crits-Christoph (2012a) report low agreement between CTACS ratings made by supervisors and independent judges, with supervisors providing higher ratings. These findings suggest that assessors may rate competence differently depending on their level of independence. However, it is not known whose ratings are more accurate. Using independent assessors for example reduces the likelihood of ratings being influenced by demand characteristics (e.g., pressure to 'pass' students) and information other than that obtained through viewing treatment sessions (e.g., prior competence, ability in other domains) (Fairburn & Cooper, 2011). However, supervisors are more likely to adhere to the same definition of competence and may provide more accurate ratings as they have access to a greater wealth of contextual information (e.g. prior sessions, patient history etc.) (Kazantzis, 2003; McGlinchey & Dobson, 2003). Whilst research suggests that therapist competence cannot be reliably rated by trained novices (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011), it is unclear what expertise are necessary. Additionally, published studies do not provide a benchmark as they rarely specify assessors' experience, and those that do range from students (e.g. Bryant et al., 1999; Simons et al., 2010) to highly qualified clinicians with years of experience (e.g. Borge et al., 2008; Hoffart et al., 2009; Jacobson et al., 1996; Keen & Freeston, 2008; McManus, Westbrook, et al., 2010). Second, as rating a treatment session is time consuming and costly, competence is typically inferred from rating one or two isolated treatment sessions per therapist. However, recent studies suggest that a much higher sample of clinical work needs to be assessed in order to reliably assess therapist competence. It has, for example, been reported that between 15 and 24 sessions rated by a single assessor, or 19 sessions rated by two assessors, are necessary to achieve a reliable assessment of a therapist's competence (Keen & Freeston, 2008; McManus et al., under review). Similarly, Dennhag, Gibbons, Barber, Gallop, and Crits-Christoph (2012b) found that between 24 and 30 treatment sessions would need to be sampled in order to make a reliable judgment about a CBT therapist's competence. Weck, Bohn, Ginzburg, and Ulrich (2011a) attempt to improve the feasibility of this method of establishing competence by investigating rating

session segments. They report overall competence ratings based on the middle third of a session to be comparable to those based on entire sessions. However, for individual items, judgments based on session segments were less reliable than those based on entire sessions and a number of items were not assessed at all. For example, rating the first segment did not assess setting homework, whilst rating the middle and last segments did not assess agenda setting, diary review, use of questionnaires, or homework review. Rating individual sessions also makes it difficult for assessors to establish whether therapists are responding appropriately to the idiosyncratic characteristics of the patient given the course of therapy to date (McGlinchey & Dobson, 2003; Waltz et al., 1993) and may fail to assess important aspects of treatment which are not demonstrated in those specific sessions (e.g. a session from the middle of therapy may not assess formulation development or relapse prevention). Furthermore, therapists are often required to self-select sessions for evaluation. This introduces a sampling bias, as therapists may submit sessions from patients who were particularly easy to treat or a session where they feel they performed especially well, rather than sessions which are representative of their general practice.

The feasibility of using evaluations of therapists' in-session skill is also uncertain. In order to achieve reliability and validity standards, multiple independent experts in CBT who are trained in the use of the specific measure are required to watch and rate multiple CBT sessions, making the measures practically difficult to implement and prohibitively resource-intensive and costly. McManus et al. (2012) for example estimate the cost of providing a reliable assessment of clinical skill using the CTS to be between £1500 (one assessor/15 ratings) and £3800 (two assessors/19 ratings) per therapist. Furthermore, it may not always be possible to obtain recordings of treatment sessions as trainees may not yet be working clinically, recordings may not be of adequate quality, or issues such as patient consent or service constraints may prevent sessions being recorded or viewed by others.

4.2 Supervisory assessments.

Supervisory assessments of therapist competence are completed retrospectively on the basis of therapists' routine clinical work over a period of time. They are low cost, quick and easy to implement and may be more representative of therapists' general competence than measures based on a small number of isolated treatment sessions. It is considered best practice to view session material within supervision. However, use of direct observational techniques is rare in routine practice (Townend, Iannetta, & Freeston, 2002). Hence supervisory assessments may in reality be based on second-hand reports of treatment rather than direct observations. This is

especially problematic given that supervisees often fail to disclose key material from therapy sessions (Ladany, Hill, Corbett, & Nutt, 1996). Furthermore, evaluations carried out by supervisors may be influenced by confounding variables such as therapists' likability, performance in other areas (e.g. academic ability) or service demands such as a pressure to "pass" therapists.

Although therapist competence ratings are regularly completed by supervising clinicians in routine practice and training settings, these are often informal and unstandardised. The current review identified only two standardised supervisory rating scales: the Supervisor Rating Form (SRF: Barnfield et al., 2007; Mathieson, Barnfield, & Beaumont, 2010) and the Evaluation of Therapist's Behaviour Form (ETBF: Kuyken & Tsivrikos, 2009). The SRF is a 24-item scale, which includes items from the CTS-R and a number of additional items relating to CBT skills and behaviours (e.g. assessment skills, formulation development), which are rated on a 0-6 scale on the basis of supervisors' observations from a number of supervision sessions and observed therapy sessions. However the psychometric properties of the SRF have not yet been evaluated. Thus it is not known whether the SRF is a reliable or valid measure of therapist competence. The ETBF provides global competence ratings, including (1) overall evaluation of competence, (2) general cognitive therapy skills, (3) flexibility and (4) general therapeutic/interpersonal skills ratings made on the basis of viewing and discussing extracts of session recordings across a number of supervision sessions. Each item is rated on a 7-point scale, based upon where the therapist falls in relation to an 'average' CBT therapist (0 *out*, 1 *below average*, 2 *average*, 3 *above average*, 4 *much above average*, 5 *excellent*, 6 *great*). Hence the scale does not indicate whether a therapist has reached a specified competence threshold (i.e. a therapist rated as better than average is not necessarily competent, merely *more* competent *compared to others*) and is dependent upon assessors' perceptions of an 'average' therapist. The ETBF has however demonstrated good inter-rater reliability (k = .80), and adequate convergent validity with independent supervisors' 5-point scale ratings of overall performance (r_s = .49, p < .05), cognitive therapy skills (r_s = .50, p < .05), and general therapy skills (r_s = .32, p < .15) (Kuyken & Tsivrikos, 2009). Additionally, Kuyken and Tsivrikos (2009) found ETBF ratings to be associated with patient outcomes. Whilst initial findings are promising, they rely on a single study carried out in the context of CBT for depression by the team who developed the measure. Further replication and extension is therefore required.

4.3 Therapist self-assessment.

Self-assessments are rarely employed as formal competence assessment tools within research and training settings. Instead they are primarily used to foster self-reflection and

highlight ongoing learning needs (McManus, Rosen, et al., 2010). However, the limited practical and financial burden of this assessment method makes it a more popular choice in routine practice settings. Standardised measures which were originally designed as assessor-rated measures of in-session performance, such as the YACS (e.g. Carroll et al., 2000) and CTS (e.g. Brosan et al., 2008; McManus, Rakovshik, Kennerley, Fennell, & Westbrook, 2011) have been used as self-assessment measures. Additionally, the CTS and CTS-R have been modified to create the cognitive therapy self-rating scale (CTSS: Bennett-Levy & Beedie, 2007) and student self-rating form (SSRF: Barnfield et al., 2007; Mathieson et al., 2010). The CTSS modifies the rating scale of the 13-item CTS so that therapists rate their own skill on a 10-point scale from 1 *no skill at all* to 10 *master skill level*. The CTSS demonstrates good internal consistency (α = .93, Bennett-Levy & Beedie, 2007) and has been found to increase significantly following training (Bennett-Levy & Beedie, 2007; Westbrook et al., 2008). The SSRF is a 24-item scale closely modelled on the CTS-R, which is completed on the basis of therapists' general perception of their therapeutic skills using a scale from 0 = *poor performance* to 5 = *excellent performance*. It has not yet been established whether any of these original or adapted scales provide valid or reliable measures of therapist competence. Furthermore, research indicates that there may be a tendency for therapists to either over- or under-estimate their own competence compared to independent and supervisory ratings of in-session performance (Brosan et al., 2008; Carroll et al., 2000; Mathieson et al., 2010; McManus et al., 2011), therefore calling into question the accuracy of therapist self-assessment. Whilst uncertainty regarding the accuracy of therapists' self-perceived competence undermine the utility of self-assessments as a formal competence measure, this do not undermine the use of self-assessments as a formative assessment tool. Indeed drawing comparisons between supervisor and therapist ratings may be a useful learning device (Carroll et al., 2000).

4.4 Patient surveys.

Patient surveys frequently form part of routine service evaluation and may also be used to assess therapist competence. The present review identified one rating scale designed to assess patients' perceptions of CBT therapists' competence: the patient report of therapy form (PRTF: Kuyken & Tsivrikos, 2009). The PRTF is a 10-item scale which assesses patients' perceptions of (1) therapist's general therapy skills (e.g. sympathetic, warm) and (2) acquisition of therapeutic skills (e.g. ability to cope with moods) on a 0 = *not at all* to 2 = *very much* scale. Kuyken and Tsivrikos (2009) report a high degree of internal consistency for both subscales (α = .86 and .91 respectively) and found that PRTF ratings in combination with supervisory ratings (ETBF) explained 15% of variance in patient outcome. However, patient ratings were not significantly correlated with supervisory ratings, suggesting that they measure distinct constructs. The validity

of the scale is also undermined by the fact that the scale is not CBT specific and focuses on patient skill acquisition, which is likely to be confounded by issues other than therapist competence (e.g. patient motivation). A number of more general difficulties with using patient surveys to assess competence have also been raised, including language and literacy issues, problems obtaining enough surveys per therapist to deliver reproducible results and difficulty delineating the individual therapist's contribution from that of the wider treatment setting and confounding factors such as treatment gains (Manring, Beitman, & Dewan, 2003). Thus patient surveys may be better suited to evaluation of patient satisfaction at a service level, rather than assessment of individual therapist's competence.

4.5 Patient outcome.

Patient outcome data (i.e. any standardised patient symptom/functioning measure) can be used to infer the competence of CBT therapists. This method makes intuitive sense in that reduction in distressing symptoms is the primary goal of delivering CBT and it would therefore be expected that more competent therapists would evidence better patient outcomes. It is also cost-effective and easy to implement as there are a number of standardised self-report questionnaires with excellent reliability and validity which services routinely administer pre- and post-treatment. However, inferring the competence of individual therapists from patient outcome has a number of substantial limitations. First, this method fails to account for patient variables which impact on patient outcome (e.g. patient severity, problem duration, comorbidity, life circumstances etc.) (Sánchez-Meca, Rosa-Alcázar, Marín-Martínez, & Gómez-Conesa, 2010). Hence patient outcome may present a misleading picture of competence, especially for therapists who work with complex, difficult to treat patient groups. Second, a single therapist is only able to treat a relatively small number of patients at any given time, meaning that conclusions regarding competence are likely to be based on the outcome of a small number of patients. Third, this method fails to take into account treatment specific factors. Improved patient outcomes could for example be due to the therapist competently delivering interpersonal therapy rather than CBT. Fourth, this method of assessing competence is based on a somewhat circular definition of competence, with competent therapists being defined as those with better patient outcomes and better patient outcomes being used to infer their competence (Sharpless & Barber, 2009). For these reasons, patient outcome is rarely employed as a measure of individual therapist competence, although it is used to evaluate the outcome of services and training courses as a whole (Decker et al., 2011).

Discussion

Summary of Assessment Methods

This review identified ten key methods for assessing the competence of CBT therapists which were presented within Miller's (1990) framework for assessing clinical skill as those which assess therapists' knowledge ('knows'), practical understanding ('knows how'), practical application of knowledge- skill ('shows how'), and clinical practice ('does'). In principle, knowledge of CBT is the easiest domain to operationalize and assess. However, whilst MCQ's are a widely used and feasible means of assessing basic CBT knowledge, few standardised MCQs with demonstrated validity and reliability have been developed. Essays also provide a cost-effective method of assessing more complex, in-depth CBT knowledge. Although the validity of essays is not yet established, they have been shown to be reliable when structured marking procedures are implemented. Some progress has been made in developing case reports and short-answer clinical vignettes to assess therapists' understanding of the clinical application of CBT knowledge. However, the reliability, validity and utility of case reports have been called into question and there is a lack of standardised short-answer clinical vignettes which have undergone psychometric evaluation. Additionally, the inclusion of clinically relevant contextual information within essays and MCQs would enable the assessment of therapists' practical understanding, though this has largely been neglected to date. Given the current lack of psychometric evaluation of these knowledge-based assessment methods, caution must be exercised when using these methods to assess therapists' ability to competently deliver CBT, especially within the context of high-stakes, summative assessments.

Assessing more complex, skills-based aspects of competence has proved more challenging. Little progress has been made in assessing therapists' CBT skills within clinical role-plays. However, standardised role-plays have the potential to provide a sensitive, focused and practical assessment of CBT therapists' skill and thus warrant further investigation. Although a variety of methods have been proposed for assessing therapists' skill within clinical practice, many have significant limitations and all require further systematic evaluation. Patient outcome appears on the surface to provide a compelling method for assessing therapists' clinical practice. However, the degree to which the competence of individual therapists can be accurately inferred from limited and confounded patient outcome data is questionable. Questions have also been raised regarding the feasibility and validity of inferring therapist competence from patient surveys. Scales originally designed to be rated by assessors on the basis of therapists' skill in the delivery of CBT within treatment sessions have also been used as self-assessment tools, both in

original and adapted forms. However, it has not yet been established whether these scales provide a valid or reliable assessment of competence and the ability to accurately rate one's own competence remains uncertain.

Assessments based on supervision sessions which include direct observation of treatment sessions may provide a more robust assessment of therapists' clinical practice. Further psychometric evaluation of the scales used by supervisors to rate competence is however required and questions regarding the ability of supervisors to provide accurate, unbiased competence ratings need to be addressed. The most widely used method of assessing therapists' clinical practice is assessor-rated scales which are used to determine therapists' skill in the delivery of CBT within treatment session(s). Although there are a number of standardised scales, the validity and reliability of many of the scales has not yet been established outside of the controlled research trials for which they were developed. Difficulties defining and disentangling competence, poor inter-rater reliability and limited feasibility also remain concern. There is also a question as to whether the scales should focus on competences required to deliver disorder-specific protocols or those which underpin most CBT interventions. Although disorder-specific scales reflect the nuances of a specific CBT protocol, implementing these scales may be unfeasible in training and routine clinical practice settings where therapists deliver CBT to patients experiencing a wide range of mental health problems. In sum, although it has proved difficult to operationalize and assess skills-based aspects of competence, it is important to avoid oversimplification or limited focus on domains which are most easily measured. Hence developing feasible, cost-effective and psychometrically sound methods of assessing CBT therapists' skill, both within standardised settings and routine clinical practice is paramount.

Further Psychometric Evaluation

To effectively assess CBT competence it is vital that the psychometric properties of the assessment tools being used have been established prior to their implementation. This review highlights the paucity of research examining the reliability and validity of current methods of assessing CBT competence. Indeed the lack of sufficient empirical examination and difficulties demonstrating adequate reliability and validity undermine the utility of implementing the instruments identified in the current review and limits the conclusions that can be drawn. Hence further psychometric evaluation of the tools used to assess CBT competence is required. A particular priority for future studies will be to examine the predictive validity of competence assessments. This is important given that the ultimate goal of delivering competent CBT is to

alleviate patients' symptoms and it would be expected that there would be some relationship between competence in CBT and patient outcome.

A Multi-method Approach to Assessment

Given that no single method is able to provide a comprehensive assessment of all aspects of CBT competence and that many of the methods have not yet been systematically evaluated, it may be prudent to use a multi-method approach when assessing CBT therapist competence. Such an approach may need to address all of the domains outlined within Miller's (1990) competence assessment framework in order to provide an adequate assessment of and feedback regarding all aspects of therapist competence. A number of CBT training courses currently use a variety of different assessment methods to establish the competence of their trainees (Keen & Freeston, 2008; McManus et al., under review; McManus, Westbrook, et al., 2010) and the IAPT high intensity training program requires therapists to submit a multi-method practice portfolio (Department of Health, 2011). The use of multiple methods is however costly. McManus et al. (under review) for example estimate that the minimum cost per trainee to deliver CTS, case report and essay assessments that meet standard reliability criteria would be between £1872.93 and £3377.10. It is therefore important to examine the optimal combination of methods necessary to provide a complimentary and thorough assessment of CBT therapist competence.

Feasible Measures

It is prohibitively expensive and labour intensive to achieve reliable assessments of CBT competence using many of the methods discussed in the present review. Indeed high cost and labour constraints have been suggested as a reason why few treatment studies assess competence (Perepletchikova, Hilt, Chereji, & Kazdin, 2009) and this may also provide some explanation as to why validated CBT competence measures are rarely used within routine practice settings (Townend et al., 2002). Although attaining a satisfactory balance between the need for reliable and valid assessments of CBT competence and limits on availability of resources is a difficult challenge, it is vital to develop more cost-effective methods of assessing competence which can be easily rolled out across a number of settings.

Implementation Protocols

There is currently a large degree of variation in the way in which different CBT competence assessment methods are implemented, which influences the reliability and validity of the resulting competence assessment. Protocols outlining the way in which competence assessment methods should be utilised are therefore required. These will need to address two

key areas. First, it must be established what the assessments should be based upon (e.g. assessment of a full course of therapy vs. multiple evaluations of specific aspects of treatment) and how the material should be selected (e.g. therapist selection vs. random selection). Second, it is necessary to determine who is best placed to assess competence. This includes the level of independence of the assessor (e.g. external rater vs. supervisor), the degree of training required to use the measure (e.g. rating manual vs. workshop) as well as the minimum experience and understanding of CBT required (e.g. novice vs. experience CBT therapist). This is especially relevant for assessments where a right or wrong answer is not clearly distinguishable. A multiple choice questionnaire for example could be marked accurately and reliably by an individual with no understanding of CBT as there is no uncertainty about whether the answer is correct. Conversely determining whether a therapist has chosen and skilfully delivered appropriate intervention strategies given the idiosyncratic presentation of their patient is a more complex task.

Benchmarks

There is a need for validated benchmarks that outline what outcome is required on each competence assessment tool in order to conclude that the therapist being assessed has demonstrated an acceptable level of competence. Sharpless and Barber's (2009) competence framework argues that therapists progress through five key developmental stages: novice, advance beginner, competence, proficiency and finally expertise. This framework outlines stage-specific clinical benchmarks and could therefore provide a useful basis for further development of competence thresholds. The threshold which a therapist is required to reach will also need to be specified according to the purpose of assessment. When determining whether candidates are suitable for inclusion on a CBT training course a therapist may for example only be required to demonstrate the knowledge and skills of a novice or advanced beginner, whereas those seeking accreditation may be required to demonstrate competence or proficiency in CBT delivery.

Formative and Summative Assessment

A number of different assessment methods may be required for different assessment purposes. In particular a distinction may be drawn between formative and summative assessments. Formative assessments provide ongoing corrective feedback in order to promote self-reflection and guide future learning, whilst summative assessments provide an overall judgment of competence for qualification or accreditation purposes. It is therefore essential that formative assessments provide sufficient in-depth feedback regarding specific aspects of competence, whilst the use of methods with a high degree of validity and reliability are a primary

concern for 'high stakes' summative assessments (Epstein, 2007). Given the different strengths and limitations of each assessment method, different methods may be better suited to each purpose. For example, whilst methods such as self-assessment, case reports and patient surveys may not be robust enough for summative assessment purposes, they may provide therapists with useful corrective feedback. Conversely, more robust methods such as assessor-rated treatment sessions, role-plays, MCQ's and essays may not provide sufficient in-depth feedback regarding specific strengths and weaknesses to drive future learning and thus may be better suited to summative assessments.

Conclusions

It is clear that there are significant limitations with existing CBT competence assessment methods and that further refinement of existing measures and / or the development of novel assessment tools with validated benchmarking and clear implementation protocols is needed. Of particular importance is the development of improved methods of assessing the higher skills-based levels of competence: 'skill' and 'does'. To date, the 'gold standard' has been ratings of therapists' in session performance but there remains a lack of empirically evaluated measures with adequate reliability and validity. Furthermore, reliability and feasibility within the resource constraints of clinical services remain challenging. Structured supervisor assessment scales and the development of OSCE's also offer potential additional methods which warrant further investigation. It is important that the development of CBT competence assessment tools gives consideration to the feasibility of the method and balances this with the need for valid and reliable measurement. Given the complex, multi-faceted nature of CBT competence, multi-method assessments may ultimately be necessary in order to provide adequate assessment of all aspects of CBT competence, with limited packages being implemented in different settings according to resource availability.

References[6]

*Barber, J. P., Foltz, C., Crits-Christoph, P., & Chittams, J. (2004). Therapists' adherence and competence and treatment discrimination in the NIDA collaborative cocaine treatment study. *Journal of Clinical Psychology, 60*(1), 29-41. doi: 10.1002/jclp.10186

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

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology-Research and Practice, 38*(5), 493-500. doi: 10.1037/0735-7028.38.5.493

Barker, C., Pistrang, N., & Elliot, R. (2002). *Foundations of quantitative measurement*. Chichester: John Wiley & Sons.

Barlow, D. H., Levitt, J. T., & Bufka, L. F. (1999). The dissemination of empirically supported treatments: A view to the future. *Behaviour Research and Therapy, 37*(1), 147-162. doi: 10.1016/s0005-7967(99)00054-6

*Barnfield, T. V., Mathieson, F. M., & Beaumont, G. R. (2007). Assessing the development of competence during postgraduate cognitive-behavioral therapy training. *Journal of Cognitive Psychotherapy, 21*(2), 140-147. doi: 10.1891/088983907780851586

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New-York: Guilford Press.

Beck, A. T., Wright, F. W., Newman, C. F., & Liese, B. (1993). *Cognitive therapy of substance abuse*. New York: Guildford Press.

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

Bennett-Levy, J., & Beedie, A. (2007). The ups and downs of cognitive therapy training: What happens to trainees' perception of their competence during a cognitive therapy training course? *Behavioural and Cognitive Psychotherapy, 35*(1), 61-75. doi: 10.1017/S1352465806003110

Berge, T. (2011). *Improving Access to Psychological Therapies in Europe*. Paper presented at the Symposium presented at 41st Annual Congress of the European Association of Behavioral and Cognitive Therapies, Reykjavik, Iceland.

[6] References marked with an asterisk indicate studies included in the review.

*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. doi: 10.1017/S1352465801004040

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal, 314*, 572. doi: 10.1136/bmj.314.7080.572

*Borge, F.-M., Hoffart, A., Sexton, H., Clark, D. M., Markowitz, J. C., & McManus, F. (2008). Residential cognitive therapy versus residential interpersonal therapy for social phobia: A randomized clinical trial. *Journal of Anxiety Disorders, 22*(6), 991-1010. doi: 10.1016/j.janxdis.2007.10.002

*Bouchard, S., Gauthier, J., Nouwen, A., Ivers, H., Vallieres, A., Simard, S., & Fournier, T. (2007). Temporal relationship between dysfunctional beliefs, self-efficacy and panic apprehension in the treatment of panic disorder with agoraphobia. *Journal of Behavior Therapy and Experimental Psychiatry, 38*(3), 275-292. doi: 10.1016/j.jbtep.2006.08.002

*Brosan, L., Reynolds, S., & Moore, R. G. (2007). Factors associated with competence in cognitive therapists. *Behavioural and Cognitive Psychotherapy, 35*(02), 179-190. doi: doi:10.1017/S1352465806003304

*Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy, 36*(5), 581-587. doi: 10.1017/S1352465808004438

Brown, L. A., Craske, M. G., Glenn, D. E., Stein, M. B., Sullivan, G., Sherbourne, C., . . . Rose, R. D. (2013). CBT Competence in novice therapists improves anxiety outcomes. *Depression and Anxiety, 30*(2), 97-115. doi: 10.1002/da.22027

*Bryant, M. J., Simons, A. D., & Thase, M. E. (1999). Therapist skill and patient variables in homework compliance: Controlling an uncontrolled variable in cognitive therapy outcome research. *Cognitive Therapy and Research, 23*(4), 381-399. doi: 10.1023/A:1018703901116

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioural therapy: A review of meta-analyses. *Clinical Psychology Review, 26*, 17 - 31.

Carroll, K. M. (1998). *A cognitive-behavioral approach: Treating cocaine addiction*. Rockville, MD: National Institute on Drug Abuse.

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

Case, S. M., & Swanson, D. B. (2001). *Constructing written test questions for the basic and clinical sciences* (3rd ed.). Philadelphia: National Board of Medical Examiners.

Centre for Reviews and Dissemination. (2009). Systematic reviews: CRD's guidance for undertaking reviews in health care. Retrieved from http://www.york.ac.uk/inst/crd/index_guidance.htm

Chevron, E. S., & Rounsaville, B. J. (1983). Evaluating the clinical skills of psychotherapists: A comparison of techniques. *Archives of General Psychiatry, 40*(10), 1129-1132. doi: 10.1001/archpsyc.1983.01790090091014

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23*, 375 - 384.

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. doi: 10.1016/j.brat.2009.07.010

Clauser, B. E., & Schuwirth, L. W. T. (2002). The use of computers in assessment. In G. Norman, C. van der Vleuten & D. Newble (Eds.), *International Handbook of Research in Medical Education* (Vol. Vol. 2, pp. 757 - 792). Dordrecht: Kluwer.

Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., . . . Zitrin, C. (1991). Meta-Analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1*(2), 81-91. doi: 10.1080/10503309112331335511

*Crits-Christoph, P., Siqueland, L., Chittams, J., Barber, J. P., Beck, A. T., Frank, A., . . . Woody, G. (1998). Training in cognitive, supportive–expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology, 66*(3), 484-492. doi: 10.1037/0022-006X.66.3.484

*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: A Journal of Research in Psychiatry and the Allied Sciences, 34*(5), 855-863. doi: 10.1017/S0033291703001855

*Davidson, K., Tyrer, P., Tata, P., Cooke, D., Gumley, A., Ford, I., . . . Crawford, M. J. (2009). Cognitive behaviour therapy for violent men with antisocial personality disorder in the community: An exploratory randomized controlled trial. *Psychological Medicine, 39*(4), 569-577. doi: 10.1017/S0033291708004066

Decker, S., Jameson, M., & Naugle, A. (2011). Therapist training in empirically supported treatments: A review of evaluation methods for short- and long-term outcomes. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(4), 254-286. doi: 10.1007/s10488-011-0360-1

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012a). Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, 1-11. doi: 10.1080/10503307.2012.716528

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

Department of Health. (2008). *IAPT implementation plan: National guidelines for regional delivery*. www.iapt.nhs.uk.

Department of Health. (2011). *National curriculum for high intensity cognitive behavioural therapy courses*. www.iapt.nhs.uk.

*Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*(4), 658-670. doi: 10.1037/0022-006x.74.4.658

*Dobson, K. S., Shaw, B. F., & Vallis, T. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24*(4), 295-300. doi: 10.1111/j.2044-8260.1985.tb00662.x

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. doi: 10.1093/clipsy.bpi046

Downing, S. M. (2002). Assessment of knowledge with written test formats. In G. Norman, C. Van der Vleuten & D. Newble (Eds.), *International handbook of research in medical education* (Vol. 2, pp. 647 - 672). Dordrecht: Kluwer.

*Durham, R. C., Guthrie, M., Morton, R. V., Reid, D. A., Treliving, L. R., Fowler, D., & Macdonald, R. R. (2003). Tayside-Fife clinical trial of cognitive-behavioural therapy for medication-resistant psychotic symptoms. *The British Journal of Psychiatry, 182*(4), 303-311. doi: 10.1192/bjp.2.360

Epstein, R. M. (2007). Assessment in medical education. *New England Journal of Medicine, 356*(4), 387-396. doi: 10.1056/NEJMra054784

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49*(6-7), 373-378.

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. doi: 10.1016/0165-0327(87)90029-2

*Forand, N. R., Evans, S., Haglin, D., & Fishman, B. (2011). Cognitive behavioral therapy in practice: Treatment delivered by trainees at an outpatient clinic is clinically effective. *Behavior Therapy, 42*(4), 612-623. doi: 10.1016/j.beth.2011.02.001

Fowler, D., Rollinson, R., & French, P. (2011). Adherence and competence assessment in studies of CBT for psychosis: Current status and future directions. *Epidemiology and Psychiatric Sciences, 20*(2), 121-126. doi: 10.1017/S2045796011000199

*Friedman, E., Thase, M., Kornblith, S., Wisniewski, S., Briggs, M., Rush, A., . . . Veenstra, G. (2004). The Implementation of Cognitive Therapy in STAR*D. *Cognitive Therapy and Research, 28*(6), 819-833. doi: 10.1007/s10608-004-0668-1

Fujisawa, D., Nakagawa, A., Kikuchi, T., Sado, M., Tajima, M., Hanaoka, M., . . . Ono, Y. (2011). Reliability and validity of the Japanese version of the Cognitive Therapy Awareness Scale: A scale to measure competencies in cognitive therapy. *Psychiatry and Clinical Neurosciences, 65*(1), 64-69. doi: 10.1111/j.1440-1819.2010.02163.x

*Garety, P. A., Fowler, D. G., Freeman, D., Bebbington, P., Dunn, G., & Kuipers, E. (2008). Cognitive–behavioural therapy and family intervention for relapse prevention and symptom reduction in psychosis: randomised controlled trial. *The British Journal of Psychiatry, 192*(6), 412-423. doi: 10.1192/bjp.bp.107.043570

Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy, 50*(12), 747-752. doi: 10.1016/j.brat.2012.09.001

*Gordon, P. K. (2006). A comparison of two versions of the cognitive therapy scale. *Behavioural and Cognitive Psychotherapy, 35*(3), 343-353. doi: 10.1017/S1352465806003390

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *British Medical Journal, 331*(7524), 1064-1065. doi: 10.1136/bmj.38636.593461.68

Gwet, K. L. (2010). *Handbook of inter-rater reliability (2nd ed.): The definitive guide to measuring the extent of agreement among multiple raters*. Gaithersburg: Advanced Analytics.

*Haddock, G., Barrowclough, C., Tarrier, N., Moring, J., O'Brien, R., Schofield, N., . . . Lewis, S. (2003). Cognitive–behavioural therapy and motivational intervention for schizophrenia and substance misuse. *The British Journal of Psychiatry, 183*(5), 418-426. doi: 10.1192/bjp.183.5.418

*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. doi: 10.1017/S1352465801002089

*Hoffart, A., Borge, F.-M., Sexton, H., & Clark, D. M. (2009). Change processes in residential cognitive and interpersonal psychotherapy for social phobia: A process-outcome study. *Behavior Therapy, 40*(1), 10-22. doi: 10.1016/j.beth.2007.12.003

*Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1992). Cognitive Therapy and Pharmacotherapy for Depression: Singly and in Combination. *Arch Gen Psychiatry, 49*(10), 774-781. doi: 10.1001/archpsyc.1992.01820100018004

*Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive–behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*(5), 747-755. doi: 10.1037/0022-006X.69.5.747

*Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. E. (1996). A component analysis of cognitive -behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64*(2), 295-304. doi: 10.1037/0022-006X.64.2.295

*Jacobson, N. S., & Gortner, E. T. (2000). Can depression be de-medicalized in the 21st century: Scientific revolutions, counter-revolutions and the magnetic field of normal science. *Behaviour Research and Therapy, 38*(2), 103-117. doi: 10.1016/s0005-7967(99)00029-7

*James, I. A., Blackburn, I. M., Milne, D. L., & Reichfelt, F. K. (2001). Moderators of trainee therapists' competence in cognitive therapy. *British Journal of Clinical Psychology, 40*, 131-141. doi: 10.1348/014466501163580

James, I. A., Blackburn, I. M., & Reichelt, F. K. (2001). *Cognitive therapy scale- revised (CTS-R): Rating manual.* University of Newcastle.

*Jarrett, R. B., Schaffer, M., McIntire, D., Witt-Browder, A., Kraft, D., & Risser, R. C. (1999). Treatment of atypical depression with cognitive therapy or phenelzine: A double-blind,

placebo-controlled trial. *Arch Gen Psychiatry, 56*(5), 431-437. doi: 10.1001/archpsyc.56.5.431

Karlin, B. E., Brown, G. K., Trockel, M., Cunning, D., Zeiss, A. M., & Taylor, C. B. (2012). National dissemination of cognitive behavioral therapy for depression in the department of veterans affairs health care system: Therapist and patient-level outcomes. *Journal of Consulting and Clinical Psychology, 80*(5), 707-718. doi: 10.1037/a0029328

Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist, 59*(8), 774-781. doi: 10.1037/0003-066x.59.8.774

Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency assessment toolkit for professional psychology. *Training and Education in Professional Psychology, 3*(4), S27-S45. doi: 10.1037/a0015833

Kazantzis, N. (2003). Therapist competence in cognitive-behavioural therapies: Review of the contemporary empirical evidence. *Behaviour Change, 20*(1), 1-12. doi: 10.1375/bech.20.1.1.24845

*Keen, A., & Freeston, M. (2008). Assessing competence in cognitive-behavioural therapy. *British Journal of Psychiatry, 193*(1), 60-64. doi: 10.1192/bjp.bp.107.038588

*Kennedy-Merrick, S. J., Haarhoff, B. A., Stenhouse, L. M., Merrick, P. L., & Kazantzis, N. (2008). Training cognitive behavioural therapy practitioners in New Zealand: From university to clinical practice. *New Zealand Journal of Psychology, 37*(2), 8-17.

*Kuyken, W., & Tsivrikos, D. (2009). Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychotherapy and Psychosomatics, 78*(1), 42-48. doi: 10.1159/000172619

Ladany, N., Hill, C. E., Corbett, M. M., & Nutt, E. A. (1996). Nature, extent, and importance of what psychotherapy trainees do not disclose to their supervisors. *Journal of Counseling Psychology, 43*(1), 10-24. doi: 10.1037/0022-0167.43.1.10

Laireiter, A. R., & Willutzki, U. (2003). Self-reflection and self-practice in training of cognitive behaviour therapy: An overview. *Clinical Psychology and Psychotherapy, 10*(1), 19-30.

Larsen, D. P., Butler, A. C., & Roediger, H. L. (2008). Test-enhanced learning in medical education. *Medical Education, 42*(10), 959-966. doi: 10.1111/j.1365-2923.2008.03124.x

Leigh, I. W., Smith, I. L., Bebeau, M. J., Lichtenberg, J. W., Nelson, P. D., Portnoy, S., . . . Koslow, N. J. (2007). Competency assessment models. *Professional Psychology-Research and Practice, 38*(5), 463-473. doi: 10.1037/0735-7028.38.5.463

*Lewis, S., Tarrier, N., Haddock, G., Bentall, R., Kinderman, P., Kingdon, D., . . . Dunn, G. (2002). Randomised controlled trial of cognitive-behavioural therapy in early schizophrenia:

Acute-phase outcomes. *The British Journal of Psychiatry, 181*(43), s91-s97. doi: 10.1192/bjp.181.43.s91

*Lopez, M. A., & Basco, M. R. (2011). Feasibility of dissemination of cognitive behavioral therapy to Texas community mental health centers. *The Journal of Behavioral Health Services & Research, 38*(1), 91-104. doi: 10.1007/s11414-009-9209-8

Manring, J., Beitman, B. D., & Dewan, M. J. (2003). Evaluating Competence in Psychotherapy. *Academic Psychiatry, 27*(3), 136-144.

*Mathieson, F. M., Barnfield, T., & Beaumont, G. (2010). Are we as good as we think we are? Self-assessment versus other forms of assessment of competence in psychotherapy. *The Cognitive Behaviour Therapist, 2*(01), 43-50. doi: 10.1017/S1754470X08000081

*Maunder, L., Milne, D. L., & Cameron, L. (2008). Pilot evaluation of brief training in CBT for primary care practitioners. *Behavioural and Cognitive Psychotherapy, 36*(03), 341-351. doi: doi:10.1017/S1352465808004220

McGlinchey, J. B., & Dobson, K. S. (2003). Treatment integrity concerns in cognitive therapy for depression. *Journal of Cognitive Psychotherapy Winter, 17*(4), 299-318. doi: 10.1891/jcop.17.4.299.52543

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*(2), 73-84. doi: 10.1037/a0018121

McManus, F., Keen, A., Clark, G., Westbrook, D., Kennerley, H., & Fennell, M. (2010). *Measuring competence in CBT.* Paper presented at the World Conference of Behavioral and Cognitive Therapies.

*McManus, F., Keen, A., & Freeston, M. (under review). Assessing competence in cognitive-behavioral therapy: A replication and extension.

*McManus, F., Rakovshik, S., Kennerley, H., Fennell, M., & Westbrook, D. (2011). An investigation of the accuracy of therapists' self-assessment of cognitive-behaviour therapy skills. *British Journal of Clinical Psychology*. doi: 10.1111/j.2044-8260.2011.02028.x

McManus, F., Rosen, K., & Jenkins, H. (2010). Developing and progressing as a CBT therapist. In M. Mueller, H. Kennerley, F. McManus & D. Westbrook (Eds.), *Oxford guide to surviving as a CBT therapist* (pp. 301 - 322). New York: Oxford Universtity Press.

*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. doi: 10.1016/j.brat.2010.08.002

McNaughton, N., Ravitz, P., Wadell, A., & Hodges, B. D. (2008). *Psychiatric Education and Simulation : A Review of the Literature* (Vol. 53). Ottawa, ON, CANADA: Canadian Psychiatric Association.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*, 63-67.

Mills, C. N., Potenza, M. T., Fremer, J. J., & Ward, W. C. (2002). *Computer-based testing: Building the foundation for future assessments.* Mahwah, NJ: Erlbaum.

*Milne, D. L., Baker, C., Blackburn, I.-M., James, I., & Reichelt, K. (1999). Effectiveness of cognitive therapy training. *Journal of Behavior Therapy and Experimental Psychiatry, 30*(2), 81-92. doi: 10.1016/s0005-7916(99)00011-7

Milne, D. L., Claydon, T., Blackburn, I.-M., James, I., & Sheikh, A. (2001). Rationale for a new measure of competence in therapy. *Behavioural and Cognitive Psychotherapy, 29*(1), 21-33. doi: 10.1017/S1352465801001047

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266. doi: 10.1016/0272-7358(91)90103-2

Myles, P. J., Latham, M., & Ricketts, T. (2003). *The contributions of an expert panel in the development of a new measure of knowledge for the evaluation of training in cognitive behavioural therapy*. Paper presented at the Annual Conference for the British Association of Behavioural and Cognitive Psychotherapy, York University.

*Myles, P. J., & Milne, D. L. (2004). Outcome evaluation of a brief shared learning programme in cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy, 32*(02), 177-188. doi: doi:10.1017/S1352465804001183

*Pace, T. M., & Dixon, D. N. (1993). Changes in depressive self-schemata and depressive symptoms following cognitive therapy. *Journal of Counseling Psychology, 40*(3), 288-294. doi: 10.1037/0022-0167.40.3.288

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. doi: 10.1037/a0015232

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365-383. doi: 10.1093/clipsy.bpi045

Petti, P. V. (2008). The use of a structured case presentation examination to evaluate clinical competencies of psychology doctoral students. *Training and Education in Professional Psychology, 2*(3), 145-150. doi: 10.1037/1931-3918.2.3.145

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. doi: 10.1016/j.cpr.2010.03.004

*Reichelt, F., 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.

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognititve and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

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.

*Ryum, T., Stiles, T. C., Svartberg, M., & McCullough, L. (2010). The effects of therapist competence in assigning homework in cognitive therapy with cluster C personality disorders: Results from a randomized controlled trial. *Cognitive and Behavioral Practice, 17*(3), 283-289. doi: 10.1016/j.cbpra.2009.10.005

Sánchez-Meca, J., Rosa-Alcázar, A. I., Marín-Martínez, F., & Gómez-Conesa, A. (2010). Psychological treatment of panic disorder with or without agoraphobia: A meta-analysis. *Clinical Psychology Review, 30*(1), 37-50. doi: 10.1016/j.cpr.2009.08.011

*Scholey, K. A., & Woods, B. T. (2003). A series of brief cognitive therapy interventions with people experiencing both dementia and depression: A description of techniques and common themes. *Clinical Psychology & Psychotherapy, 10*(3), 175-185. doi: 10.1002/cpp.368

Schuwirth, L. W., & van der Vleuten, C. P. (2004). Different written assessment methods: what can be said about their strengths and weaknesses? *Medical Education, 38*(9), 974 - 979.

*Sensky, T., Turkington, D., Kingdon, D., Scott, J. L., Scott, J., Siddle, R., . . . Barnes, T. R. E. (2000). A randomized controlled trial of cognitive-behavioral therapy for persistent symptoms in schizophrenia resistant to medication. *Archives of General Psychiatry, 57*(2), 165-172. doi: 10.1001/archpsyc.57.2.165

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. doi: 10.1016/j.cpr.2008.09.008

*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), 837-846. doi: 10.1037/0022-006x.67.6.837

Shaw, B. F., & Wilson-Smith, D. (1988). Training therapists in cognitive-behavior therapy. In C. Perris, I. M. Blackburn & H. Perris (Eds.), *Cognitive psychotherapy: Theory and practice* (pp. 140 - 159). Heidelberg: Springer-Verlag.

*Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 73*(1), 106-115. doi: 10.1037/0022-006x.73.1.106

*Simons, A. D., Gordon, J. S., Monroe, S. M., & Thase, M. E. (1995). Toward an integration of psychologic, social, and biologic factors in depression: Effects on outcome and course of cognitive therapy. *Journal of Consulting and Clinical Psychology, 63*(3), 369-377. doi: 10.1037/0022-006x.63.3.369

*Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K., . . . Beck, A. T. (2010). Training and dissemination of cognitive behavior therapy for depression in adults: A preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology, 78*(5), 751-756. doi: 10.1037/a0020569

*Simons, A. D., & Thase, M. E. (1992). Biological markers, treatment outcome, and 1-year follow-up in endogenous depression: Electroencephalographic sleep studies and response to cognitive therapy. *Journal of Consulting and Clinical Psychology, 60*(3), 392-401. doi: 10.1037/0022-006x.60.3.392

Stewart, R. E., & Chambless, D. L. (2009). Cognitive–behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology, 77*(4), 595-606. doi: 10.1037/a0016032

*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. doi: 10.1037/a0019631

Sudak, D. M., Beck, J. S., & Wright, J. (2003). Cognitive behavioral therapy: A blueprint for attaining and assessing psychiatry resident competency. *Academic Psychiatry, 27*(3), 154-159. doi: 10.1176/appi.ap.27.3.154

*Tarrier, N., Lewis, S., Haddock, G., Bentall, R., Drake, R., Kinderman, P., . . . Dunn, G. (2004). Cognitive-behavioural therapy in first-episode and early schizophrenia. *The British Journal of Psychiatry, 184*(3), 231-239. doi: 10.1192/bjp.184.3.231

Townend, M., Iannetta, L., & Freeston, M. H. (2002). Clinical supervsion in practice: A survey of UK cognitive behavioural psychotherapists accredited by the BABCP. *Behavioural and Cognitive Psychotherapy, 30*(04), 485-500. doi: 10.1017/S1352465802004095

*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. doi: 10.1023/B:COTR.0000021536.39173.66

*Vallis, T., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54*(3), 381-385. doi: 10.1037/0022-006X.54.3.381

*Vallis, T., Shaw, B. F., & McCabe, S. B. (1988). The relationship between therapist competency in cognitive therapy and general therapy skill. *Journal of Cognitive Psychotherapy, 2*(4), 237-249.

*Valmaggia, L. R., Van der Gaag, M., Tarrier, N., Pijnenborg, M., & Sloof, C. J. (2005). Cognitive–behavioural therapy for refractory psychotic symptoms of schizophrenia resistant to atypical antipsychotic medication. *The British Journal of Psychiatry, 186*(4), 324-330. doi: 10.1192/bjp.186.4.324

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), 620-630. doi: 10.1037/0022-006X.61.4.620

*Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., . . . Addington-Hall, J. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness. *British Medical Journal, 321*(7273), 1383-1388. doi: 10.1136/bmj.321.7273.1383

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

Weck, F., Bohn, C., Ginzburg, D. M., & Ulrich, S. (2011a). Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire sessions. *Psychotherapy Research, 21*(6), 658-669. doi: 10.1080/10503307.2011.602751

Weck, F., Bohn, C., Ginzburg, D. M., & Ulrich, S. (2011b). Treatment integrity: Implementation, assessment, evaluation and correlations with outcome. *Verhaltenstherapie, 21*, 99-107.

Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of

clinical experience. *Journal of Nervous and Mental Disease, 199*(4), 276-279. doi: 10.1097/NMD.0b013e3182124617

*Weingardt, K. R., Cucciare, M. A., Bellotti, C., & Lai, W. P. (2009). A randomized trial comparing two models of web-based training in cognitive–behavioral therapy for substance abuse counselors. *Journal of Substance Abuse Treatment, 37*(3), 219-227. doi: 10.1016/j.jsat.2009.01.002

*Weingardt, K. R., Villafranca, S. W., & Levin, C. (2006). Technology-based training in cognitive behavioral therapy for substance abuse counselors. *Substance Abuse, 27*(3), 19-25. doi: 10.1300/J465v27n03_04

*Westbrook, D., McManus, F., Clark, G., & Bennett-Levy, J. (2011). Preliminary evaluation of an online training package in cognitive behaviour therapy: Satisfaction ratings and impact on knowledge and confidence. *Behavioural and Cognitive Psychotherapy, FirstView*, 1-10. doi: doi:10.1017/S1352465811000701

*Westbrook, D., Sedgwick-Taylor, A., Bennett-Levy, J., Butler, G., & McManus, F. (2008). A pilot evaluation of a brief CBT training course: Impact on trainees' satisfaction, clinical skills and patient outcomes. *Behavioural and Cognitive Psychotherapy, 36*(5), 569-579. doi: 10.1017/S1352465808004608

*Westra, H. A., Arkowitz, H., & Dozois, D. J. A. (2009). Adding a motivational interviewing pretreatment to cognitive behavioral therapy for generalized anxiety disorder: A preliminary randomized controlled trial. *Journal of Anxiety Disorders, 23*(8), 1106-1117. doi: 10.1016/j.janxdis.2009.07.014

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

*Williams, R. M., Moorey, S., & Cobb, J. (1991). Training in cognitive-behaviour therapy: Pilot evaluation of a training course using the cognitive therapy scale. *Behavioural and Cognitive Psychotherapy, 19*(04), 373-376. doi: doi:10.1017/S0141347300014075

Wright, J. H., Wright, A. S., Salmon, P., Beck, A. T., Kuykendall, J., Goldsmith, L. J., & Zickel, M. B. (2002). Development and initial testing of a multimedia program for computer-assisted cognitive therapy. *American Journal of Psychotherapy, 56*(1), 76-86.

Appendices

Appendix 1. Literature Search Strategies

Psycharticles search strategy: yielded 187 results.

Psycharticles was searched in December 2011 using the following strategy:

1. "therapist competence" OR "therapist competency" OR "therapeutic competence" OR "therapeutic competency" OR "clinical skill" OR "clinical competence" OR "clinical competency" OR "assessing competency" OR "assessing competence" OR "assess competency" OR "competence assessment" OR "competency assessment" OR "intervention competence" OR "intervention competency": in any field
2. AND "CBT" OR "cognitive behavior" OR "cognitive behaviour" OR "cognitive behavioral" OR "cognitive behavioural" OR "behavioural therapy" OR "behavioral therapy" OR "behaviour therapy" OR "behavior therapy" OR "cognitive therapy": in any field
3. AND publication year: 1980 TO 2011
4. AND publication type: journal

Science Direct search strategy: yielded 294 results.

Science direct was searched in December 2011 using the following strategy:

1. "therap* competenc*" OR "clinical competenc*" OR "therap* skill" OR "assess* competenc*" OR "competenc* assess*" OR "therapy quality" OR "intervention quality" OR "intervention competenc*" OR "clinical expertise": in all fields
2. AND "CBT" OR "cognitive behavio*" OR "behavio* therapy" OR "cognitive therapy": in title/abstract/key
3. subject: psychology
4. date range: 1980 to present

Scopus search strategy: yielded 406 results.

Scopus was searched in December 2011 using the following strategy:

1. "therap* competenc*" OR "clinical competenc*" OR "therap* skill" OR "assess* competenc*" OR "competenc* assess*" OR "therapy quality" OR "intervention quality" OR "intervention competenc*" OR "clinical expertise": in all fields
2. AND "CBT" OR "cognitive behavio*" OR "behavio* therapy" OR "cognitive therapy": in title/abstract/key
3. subject: psychology
4. limit to publication year: 1980 to 2011
5. Exclude letter/conference paper/note
6. Limit to language: English

PsycInfo search strategy: yielded 90 results.

PsycInfo was searched in December 2011 using the following strategy:

1. "therap* competenc*" or "clinical competenc*" or "therap* skill" or "assess* competenc*" or "competenc* assess*" or "therapy quality" or "intervention quality" or "intervention competenc*" or "clinical expertise": in all fields

2. "CBT" or "cognitive behavio*" or "behavio* therapy" or "cognitive therapy": in abstract, heading word, key concepts, original title, table of contents, tests & measures, title
3. 1 and 2
4. Limit 3 to English language, year 1980-current, journals

Web of Knowledge search strategy: yielded 118 results.

Web of knowledge was searched in December 2011 using the following strategy:

1. "therap* competenc*" or "clinical competenc*" or "therap* skill" or "assess* competenc*" or "competenc* assess*" or "therapy quality" or "intervention quality" or "intervention competenc*" or "clinical expertise": in topic
2. "CBT" or "cognitive behavio*" or "behavio* therapy" or "cognitive therapy": in topic
3. Timespan 1980 - 2011
4. Subject area = psychology
5. Exclude document type: meeting, other, letter, reference material
6. Languages: English

Medline search strategy: yielded 178 results.

Medline was searched in December 2011 using the following strategy:

1. "therap* competenc*" or "clinical competenc*" or "therap* skill" or "assess* competenc*" or "competenc* assess*" or "therapy quality" or "intervention quality" or "intervention competenc*" or "clinical expertise": in topic
2. "CBT" or "cognitive behavio*" or "behavio* therapy" or "cognitive therapy": in topic
3. Timespan 1980 - 2011
4. Subject area = psychology
5. Exclude document type: comment, letter, patient education handout
6. Languages: English

PubMed search strategy: yielded 185 search results.

PubMed was searched in December 2011 using the following strategy:

1. ("therapist competence") OR ("therapist competency") OR ("therapeutic competence") OR ("therapeutic competency") OR ("clinical skill") OR ("clinical competence") OR ("clinical competency") OR ("assessing competency") OR ("assessing competence") OR ("assess competency") OR ("competence assessment") OR ("competency assessment") OR ("intervention competence") OR ("intervention competency")
2. AND ("CBT") OR ("cognitive behavior") OR ("cognitive behaviour") OR ("cognitive behavioral") OR ("cognitive behavioural") OR ("behavioural therapy") OR ("behavioral therapy") OR ("behaviour therapy") OR ("behavior therapy") OR ("cognitive therapy"): in Title/Abstract

Limits: dates 1980-2011, Languages: English

Appendix 2

Summary of Studies Employing Assessments of Knowledge and Practical Understanding of CBT

Study						Assessor(s)		
Authors (year)[1]	**Research context[2]**	**Measure type**	**Knowledge assessed**	**No. of items & response options / Word limit**	**Therapist profession (N)[3]**	**Independence (N)[4]**	**Profession (N)[3]**	**No.**
Barnfield, Mathieson, & Beaumont (2007)	TS	Case report	Ability to combine theory & research within clinical work with patients, demonstrated though a brief case history, formulation and conceptualization, & treatment plan.	-	SW (5), MH Nurse (5), OT (2), Other MHP (1)	-	-	3 p/report
Keen & Freeston (2008)	TS	Essay	Ability to review CBT literature, critically evaluate it & link it to clinical practice	4000 words / 8000 words	- (52)	Superv(CS) (pool of 15)	-	2 p/essay
		Case report	Ability to select / present relevant patient material, conceptualize / write a CBT case formulation & report/ reflect on a course of CBT	4000 words	- (52)	Superv(CS) (pool of 15)	-	2 p/report
Kennedy-Merrick et al. (2008)	TS	MCQ: CBT-KQ	General CBT knowledge: general issues, theoretical underpinnings & CBT practice	26 items, 4 response options (1 correct)	Clin Psy (11), CSL (11), Psych Nurse (9), SW (8), Psy (5), Psych (5),MHW (3), GP (2) OMH (4)	n/a	n/a	n/a
Maundera Milne & Camerona (2008)	TS	MCQ: FCBT-MCQ (adapted)	General CBT knowledge, adapted to reflect specific course content (formulation driven CBT)	10 items, 4 response options (1 correct)	'predominantly' GPs (15)	n/a	n/a	n/a

Authors (year)	Research context	Measure type	Knowledge assessed	No. of items & response options / Word limit	Therapist profession (N)	Independence (N)	Profession (N)	No.
Mcmanus, Wesbrook, Vazquez-Montes, Fennell, & Kennerley (2010)[a]	TS	Essay	Ability to review CBT theory / empirical literature, critically evaluate it & link it to clinical practice	3000 words / 10000 words	CBT T trainees + Clin Psy (145), nurse (56), psych (33), CSL (14), OT (13), SW (9), GP (2), Psy (2), Other MHP (4)	Superv(TS) (-), + subsample re-rated by Superv(CS)	-	2 p/essay
		Case report	Ability to select / present relevant patient material, conceptualize / write a CBT case formulation & report/ reflect on a course of CBT	4000 words	As above	Superv(TS)	-	2 p/report
Myles & Milne (2004)	TS	MCQ: FCBT-MCQ	General CBT knowledge: therapeutic model, therapeutic principles & treatment strategies	13-items, 4 response options (1 correct)	MH nurse (76), OT (5), CSL (2), SW (2), AP (2), Psych (2), MH Nurse (1), (full data set for *n* = 36)	n/a	n/a	n/a
		Short-answer clinical vignette: VAT	Diagnosis and choice of treatment techniques	3 items (open ended)	As above	-	-	-
Sholomskas et al. (2005)	TS	MCQ	Theoretical & practical knowledge of specific treatment manual (CBT for cocaine addiction)	55 items, -	- (78)	n/a	n/a	n/a
Weingardt, Cucciare, Bellotti, & Lai (2009)[b]	TS	MCQ	Knowledge of specific course content (CBT for cocaine addiction)	73 items, -	CSL (147)	n/a	n/a	n/a

Authors (year)	Research context	Measure type	Knowledge assessed	No. of items & response options / Word limit	Therapist profession (N)	Independence (N)	Profession (N)	No.
Weingardt, Villafranca, & Levi (2006)[b]	TS	MCQ	Knowledge of specific course content (1 module from CBT for cocaine addiction)	17 items, -	CSL (166)	n/a	n/a	n/a
Westbrook, McManus, Clark, & Bennett-Levy (2011)[a]	TS	MCQ	Knowledge of specific course content (modules in basic CBT theory, assessments, understanding panic, treating panic & OCD model/treatment)	20 items (4 per module), 4 response items (1 correct)	- (94, with some missing data)	n/a	n/a	n/a

Note. MCQ = Multiple Choice Questionnaire, CBT-KQ = Cognitive Behavioural Therapy Knowledge Quiz, FCBT-MCQ = Foundation Cognitive Behavioural Therapy Multiple Choice Questionnaire, VAT = Video Assessment Task, items marked with "-" indicate that the information was not specified in the manuscript.

[1] Studies which identified as using the same or related samples (e.g., re-analysis of the same or a subset of the same results or therapists recruited from the same training course/research trial) and those carried out within the same research team / training setting are marked with the same superscript letters.

[2] Research context: RT = research trial, RT-E = embedded within a research trial, RT-E (TP) = embedded within a research trial (training phase), DS = dissemination study, TS = training setting, RCP = routine clinical practice.

[3] Therapist / assessor profession: MH Nurse = psychiatric/ mental health nurse, OT = occupational therapist, CSL = counsellor (level of training variable), AS = assistant psychologist, Psy = psychologists (exact level of training not specified), Clin Psy= clinical psychologists, PhD Psy = psychologist with a PhD, Ed Psy = educational psychologist, MA Psy = psychologist with a masters, GP = doctor with a medical degree, Psych = psychiatrist, RF = research fellow (discipline not specified), CBT T = completed some form of CBT training, usually in a higher education setting, (e.g., diploma in CBT, IAPT training), SW = clinical/ masters level social worker, BABCP = therapists meeting BABCP accreditation standards, GMHW = Graduate mental health worker, Other MHP = other mental health

practitioner. The label “trainee” indicates that the therapist is currently undergoing the specified level of training. Where names of the assessor were given but the profession of the individuals at the time of publication were not presented in the MS, this information was left as blank.

[4] Independence of assessors defined as: research = member of the team involved in conducting, analysing or reporting the study, supervisor= involved in providing training or supervision & where available further information is given in brackets: TS= therapist supervisor, i.e. provides training/supervision to the therapist(s) being assessed, CS = course supervisor, i.e. involved in providing training or supervision within the research or training setting but not to the specific therapist(s) being assessed, independent = neither a member of the research team or a supervisor/trainer, therapist = self-rating made by the therapist who is being assessed.

Appendix 3

Summary of Studies Employing Assessments of Clinical Practice

Study				Assessor(s)					
Authors (year)[1]	**Research context[2]**	**Measure type[3]**	**Therapist profession (*N*)[4]**	**Independence (*N*)[5]**	**Profession (*N*)[4]**	**Measure training?[6]**	**No.**	**Source material**	**Patient group[6]**
Barber, Foltz, Crist-christoph, & Chittams (2004)[f]	RT-E	Assessor-rated treatment session: CTACS	- (16)	Indep (-)	-	-	2 p/recording (mean score)	Two randomly selected audio recordings from every patient-therapist dyad, from sessions 2-11 & sessions 12-end	SUD
Barber, Liese, & Abrams (2003)[f]	RT-E(TP)	Assessor-rated treatment session: CTACS	Phd Psy (12), Ed Psy (2), SW (4)	Indep (2)	Clin Psy (1), MH Nurse (1)	Y	2 p/recording	Randomly selected audio recordings of sessions	SUD
Barnfield, Mathieson, & Beaumont (2007)[i]	TS	Assessor-rated treatment session: CTS-R	SW (5), MH Nurse (5), OT (2), Other MHP (1)	Superv(CS) (2), Indep (1)	Clin Psy (3)	-	3 p/recording (mean rating used)	Visual recording of a session submitted as routine course evaluation	-
		Supervisor assessment: SRF	As above, with 1 missing participant	Superv(TS)	-	-	1 p/therapist	-	-
		Self-assessment: SSRF	As above, with 1 missing participant	Therapist	See therapist profession	-	n/a	Completed pre and post training on the basis of a general perception of competence	-
Bennett-levy & Beedie (2007)[g]	TS	Self-assessment: CTSS	CBT T trainee + Psy (16), Nurse (5), psych (3)	Therapist	n/a	-	n/a	Completed periodically on the basis of a general perception of competence	Range

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Blackburn et al., (2001)[b]	TS	Assessor-rated treatment session: CTS-R	CBT T trainee + Psych (6), Clin Psy (6), Psych Nurse (8). [*n* = 11 inc. in pre-post analysis]	- (4)	- (4)	-	2 p/recording	Three visual recordings for each patient from the beginning, middle and end of therapy	Dep, SP, PD, OCD, GAD
Borge et al. (2008)[e]	RT	Assessor-rated treatment session: CTS (global item)	Clin Psy (2)	Research (2)	Clin Psy (2)	-	2 p/recording	Random selection of visual recordings	SP
Bouchard et al., (2007)	RT	Assessor-rated treatment session: CCCT (adapted)	Clin Psy trainee (2), MA Psy trainee (2)	Indep (2)	-	-	2 p/recording	Random selection of audio recordings	PD + AG
Brosan, Reynolds, & Moore (2007)[e]	RT	Assessor-rated treatment session: CTS (adapted)	Clin Psy (8), CSL (2) MH Nurse (10), Other MHP (2)	Indep (2)	CBT T (2)	N	1 p/recording (subsample = 2 p/recording)	One therapist selected audio recording of a mid-treatment session per therapist	Range
Brosan, Reynolds, & Moore (2008)[e]	RCP	Assessor-rated treatment session: CTS (adapted)	Psyc (10), MH nurse (10), other MHP (2)	Indep (1), - (1)	Dip CBT (2)	-	1 p/recording (subsample 2 p/recording)	One audio recording of a mid-treatment session submitted per therapist (5 randomly selected tapes were also re-rated)	-
		Self-assessment: CTS (adapted)	Psyc (10), MH nurse (10), other MHP (2)	Therapist (-)	See therapist info.	-	n/a	One audio recording of a mid-treatment session per therapist	-

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Bryant, Simons, & Thase (1999)[d]	RT-E	Assessor-rated treatment session: CTS (shortened)	PhD Psy (2), Psych Nurse (2)	Indep (1)	Clin Psy trainee (1)	-	1 p/recording (+3 p/recording for reliability)	Visual recordings of entire sessions from the beginning, middle and end of therapy (where possible weeks 2, 12 & 18)	Dep
Carroll et al. (2000)	RT-E	Assessor-rated treatment session: YACS	Ma Psy (-)	-	-	Y	5 p/recording (reliability)	Randomly selected visual recording of sessions	SUD
		Self-assessment: YACS	Ma Psy (-)	Therapist	See therapist profession	-	n/a	-	SUD
Crits-christoph et al. (1998)[f]	RT-E(TP)	Assessor-rated treatment session: CTS	PhD Psy, Clin Psy or ED Psy (16), SW (4)	Superv(TS) (4) + subsample re-rated by Superv(CS)	- (4)	-	2 p/recording	Three sessions per therapist (+ 2 sessions from 3 patients re-rated)	SUD
Davidson et al. (2004)	RT-E	Assessor-rated treatment session: MACT-RS	Nurse (12), Clin Psy (4), SW (2), Psych (1), OT (2)	Research (2)	-	-	2 p/recording	Therapists selected audio recordings of 3 sessions per patient	SH
Davidson et al. (2009)	TS/RCP	Assessor-rated treatment session: CTS	Psych (2), Clin Psy (2), CSL (1), CBT T (1) MH Nurse (1)	-	-	-	-	A random sample of session audio recordings	ASPD
Dimidjian et al., (2006)	RT	Assessor-rated treatment session: CTS	Psy (1)	Indep (-), superv (2)	-	-	2+ p/recording	-	Dep

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Dobson, Shaw, & Vallis (1985)[a]	RT-E	Assessor-rated treatment session: CTS	Psy (11), Psych (10)	-	PhD Psy (3), Psych (1)	Y	2 p/recording	One visual recording of a treatment session or consultation submitted per therapists as part of a selection process	Dep
Durham et al. (2003)	RT	Assessor-rated treatment session: CTS	MH Nurse (5)	Indep (1)	-	-	1 p/recording	A 'representative' sample from available session recordings + transcripts	SO
Forand et al. (2011)	RCP	Assessor-rated treatment session: CTS	Phd Psy trainee / Clin Psy trainee (*n* not specified)	Supevisor(TS)	Not specified	Not specified	1	Therapist selected visual recordings submitted as part of routine course evaluation	Range (see article)
Friedman et al., (2004)	RT	Assessor-rated treatment session: CTS	- (44)	-	-	-	-	Therapist selected 3 audio recordings of 3 consecutive sessions during training + random 10% of sessions from trial	Dep
Garety et al. (2008)	RT	Assessor-rated treatment session: CTS	Clin Psy (5) (+ Clin Psy/nurse [-])	Research (+ subsample = Indep [4])	-	-	-	Audio recordings of sessions (+ random sample re-rated)	Psyc
Gordon (2006)	TS	Assessor-rated treatment session: CTS-R, CTS-Psy	- (26)	Indep (7), superv (CS) (2)	-	Y	2 p/recording	One therapist selected audio recording of an entire session submitted as part of routine course evaluation	Pers D, PO

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Haddock et al. (2003)	RT	Assessor-rated treatment session: CTS-Psy	BABCP (-)	-	-	-	-	-	SO + SUD
Haddock et al., (2001)	TS	Assessor-rated treatment session: CTS-Psy	All CBT T trainees + MH nurse (20), OT (1), (full data for *n* = 14)	Indep (4)	Clin Psy (2) MH Nurse (1), RF (1)	Y	4 p/recording	Random sample of 2 tapes per trainee selected from the 4 audio recordings submitted by each therapist during routine training evaluation	SO, SOA, BP
Hoffart, Borge, Sexton, & Clark (2009)[e]	RT	Assessor-rated treatment session: CTS (global item)	Clin Psy (2)	Research (2)	Clin Psy (2)	-	2 p/recording	Random selection of visual recordings	SP
Hollon et al., (1992)	RT	Assessor-rated treatment session: CTS	Clin Psy (1), SW (3)	Indep (-)	-	-	-	Four therapy sessions per patient	Dep
Huppert et al. (2001)	RT-E	Assessor-rated treatment session: MCSTPD-GCI	PhD Psy (13), Psych (1)	-	-	Y	-	A random sample of session recordings were selected from a random sample of patients	PD
Jacobson et al., (1996)[c]	RT	Assessor-rated treatment session: CTS	Psy (4)	Superv (1)	-	-	1 p/recording	Assessor selected recordings	Dep
Jacobson & Gortner (2000)[c]	RT	Assessor-rated treatment session: CTS	Psy (4)	Indep (2), Superv (1)	-(3)	-	3 p/recording	assessor selected recordings	Dep

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
James, Blackburn, Milne, & Reichfelt (2001)[b]	TS	Assessor-rated treatment session: CTS-R	CBT T trainees + Psych/GP (6), Psy (6), Nurse (8)	Superv (4)	-	-	2 p/recording	Three therapist selected visual recordings (beginning, middle & end of therapy) from the first patient treated whilst enrolled in course	Dep, GAD, SP, OCD
Jarrett et al., (1999)	RT	Assessor-rated treatment session: CTS	Clin Psy (2), Psych (1)	Indep (1)	-	-	-	Two sessions per patient	Dep
Keen & Freeston (2008)[b]	TS	Assessor-rated treatment session: CTS	- (52)	Superv(CS) (pool of 15)	-	Y	2 p/recording	Therapists selected 2 visual recordings as part of routine course evaluation	Range
Kuyken, & Tsivirkos (2009)	RCP	Supervisor assessment: ETBF	- (18)	Superv(TS), (2)	-	-	1 p/recording (subset = 2 p/recording)	Knowledge of therapists' work as a whole (namely extracts of session recordings selected by the therapist and discussed in supervision)[7]	Dep
		Patient survey: PRTFBF	- (18)	n/a	n/a	n/a	n/a	Completed at the end of session one.	Dep
Lewis et al. (2002)	RT	Assessor-rated treatment session: CTS-Psy	-	Indep (2)	-	-	2 p/recording	Random selection of audio session recordings	SO

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Lopez & Basco (2011)	DS	Assessor-rated treatment session: CTS	CSL (1), SW (5), PhD Psy trainee (1)	Indep (1)	-	-	Min. 2 p/recording (mean rating)	A random sample of session recordings	Dep
Mathiesona, Barnfield & Beaumont (2010)[i]	TS	Assessor-rated treatment session: CTS-R	CBT T trainee + SW (14), nurse (12), OT (5), Other MH (3)	- (3)	Clin Psy (3)	-	3 p/recording	Visual recording of a session submitted as routine course evaluation	-
		Supervisor assessment: SRF	As above	Supervisor(TS)	-	-	1 p/therapist	Observations from supervision sessions and several observed therapy sessions	
		Self-assessment: SSRF	As above	Therapist	See therapist profession	-	n/a	Completed pre and post training on the basis of a general perception of competence	
McManus et al. (2011)[g]	TS	Assessor-rated treatment session: CTS	CBT T trainees + Clin Psy (27), nurse (20), psych (9), CSL (5), OT (1), SW (1), Other MHP (1)	Supervisors	BABCP (16)	-	1 p/recording	Trainee selected audio or video recordings of a mid-treatment session, submitted as routine course evaluation	Range
		Self-assessment: CTS	As above	Therapist	See therapist info.	-	n/a	As above	Range
Mcmanus, Wesbrook, Vazquez-Montes, Fennell, & Kennerley (2010)[g]	TS	Assessor-rated treatment session: CTS	CBT T trainees + Clin Psy (145), nurse (56), psych (33), CSL (14), OT (13), SW (9), GP (2), Psy (2), Other MHP (4)	Superv(TS) (-), + subsample re-rated by Superv(CS)	-	-	1 p/recording (subsample = 2 p/recording)	Six trainee selected audio or video recordings of a mid-treatment session, submitted as routine course evaluation	Range

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Milne, Baker, Blackburn, James, & Reichelt (1999)[b]	TS	Assessor-rated treatment session: CTS-R	CBT T + Clin Psy (6), Psych (6), MH Nurse (8)	Superv (4)	-	-	2 p/recording (average ratings)	Sessions from the beginning, middle and end of therapy seen by therapists at the start of training	Axis 1 - usually anx/dep
Pace & Dixon (1993)	RT	Assessor-rated treatment session: CCCT	CSL (7)	Indep (2)	CSL Trainee (2)	Y	1 p/recording	Assessors viewed one randomly selected session per therapist through a one-way mirror	Dep
Reichelt, James & Blackburn 2003[b]	TS	Assessor-rated treatment session: CTS-R	All CBT T trainees (24)	Superv (CS) (10)	- (10)	Y	2 p/recording (some mean ratings)	Two visual recordings selected by each therapist & submitted as part of routine course evaluation	Axis 1: Range
Ryum et al. (2010)	RT	Assessor-rated treatment session: CTS (selected items)	Clin Psy (6)	Indep (2)	-	-	-	Visual recordings of an early treatment session (usually # 6)	Pers D
Scholey & Woods (2003)	RT	Assessor-rated treatment session: CCCT	-	-	-	-	-	One session (occurring after the second session) selected per participant	Dem + Dep
Sensky et al., (2000)	RT	Assessor-rated treatment session: CTS	Psych Nurse (2)	Research (1)	Psych (1)	-	1 p/recording	Random sample of audio recordings	SO

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Shaw et al., (1999)[a]	RT-E	Assessor-rated treatment session: CTS	Psy (6), Psych (2)	Superv (2)	PhD Psy (2)	-	2 p/recording	Visual recordings 9/20 complete sessions (# 1, 2, 4, 6, 7, 10, 15, 18 & 19)	Dep
Simons et al., (2010)	DS	Assessor-rated treatment session: CTS	MA Psy (8), Clin Psy (4)	Research (1), Indep (-)	Not specified (1) Clin. Psy trainee (-)	-	Min. 2 p/recording (consensus ratings)	Therapist selected 1 visual recording of a complete session for 3 patients	Dep
Simons, Gordon, Monroe, & Thase (1995)[d]	RT	Assessor-rated treatment session: CTS	-	-	-	-	-	-	Dep
Simons & Thase (1992)[d]	RT	Assessor-rated treatment session: CTS	-	-	-	-	-	-	Dep
Strunk, Brotman, Derbeis, & Hollon (2010)	RT-E	Assessor-rated treatment session: CTS	PhD Psy (5), MH Nurse (1)	Research (2)	Clin Psy (2)	Y	2 p/recording (consensus ratings)	Visual (or audio when unavailable) of the first 4 sessions for each patient	Dep
Tarrier et al. (2004)	RT	Assessor-rated treatment session: CTS-Psy	Clin Psy (3), Nurse (2)	Indep (2)	-	-	2 p/recording	Random selection of audio session recordings	SO

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Trepka, Rees, Shapiro, Hardy, & Barkham (2004)	RT-E	Assessor-rated treatment session: CTS	Clin Psy (6)	Indep (1)	- (1)	-	1 p/recording	One randomly selected audio recording of a complete session (from beginning or end of therapy) per patient	Dep
Vallis, Shaw, & Dobson (1986)[a]	RT-E (TP)	Assessor-rated treatment session: CTS	PhD Psy / MA Psy (9)	Trainer (3) + Indep (for a subsample) (4)	MA Psy (1), PhD Psy (6)	-	2 – 5 p/recording (some mean ratings)	Visual recording of every session + random subsample re-evaluated	Dep
Vallis, Shaw, & McCabe (1988)[a]	RT-E (TP)	Assessor-rated treatment session: CTS	Psy (6), Psych (2)	Trainer / Indep (6)	- (6)	-	2 p/recording	Visual recording of every session + random subsample re-evaluated	Dep
Valmaggia, Van Der Gaag, Tarrier, Pijnenbord, & Sloof (2005)	RT	Assessor-rated treatment session: CTS-Psy (adapted)	-	Research (1), Indep (1)	-	-	2 p/recording	Two audio recordings per condition/therapist randomly selected	SO
Ward et al., (2000)	RT	Assessor-rated treatment session: CTS	Psy (12)	Indep (1)	Psy (1)	-	1 p/recording	Recordings of all sessions with the 2nd & 5th patient allocated to each therapist, of which 2 sessions per therapist were selected at random	Dep/ Dep + Anx

Authors (year)	Research context	Measure type	Therapist profession (*N*)	Independence (*N*)	Profession (*N*)	Measure training?	No.	Source material	Patient group
Westbrook et al. (2008)[g]	TS	Assessor-rated treatment session: CTS	CBT T trainee + GMHW (12), MH Nurse (8), Nurse (1), OT (1), Other MHP (2) (full data set for *n* = 18)	Supervisor(CS)	-	-	-	Therapist selected audio recording of a session submitted as routine course evaluation	Anx/ Dep
		Self-assessment: CCTS	As above	Therapist	See therapist info.	-	n/a	Completed pre and post training on the basis of a general perception of competence	Anx/ Dep
Westra, Arkowitz, & Dozois (2009)	RT	Assessor-rated treatment session: CTS	PHD Psy (1), Clin Psy (3)	-	Clin Psy Trainee (3)	Y	2 p/recording (consensus rating used)	A random sample of session recordings (3 per patient from the beginning, middle & end of treatment).	GAD
Williams, Moorey, & Cobb (1991)	TS	Assessor-rated treatment session: CTS (shortened)	Psy (8), Not specified (3)	Superv (2)	- (2)	-	2 p/recording (some mean ratings)	Two therapist selected audio recordings per therapist, submitted during routine course evaluation	Range

Note. Where there is more than one treatment arm, only information regarding the CBT arm of the study is reported in the table. Items marked with "-" indicate that the information was not specified in the manuscript.

[1] Studies which use the same or related samples (e.g. re-analysis of the same or a subset of the same results or therapists recruited from the same training course/research trial) and those carried out within the same research team / training setting are marked with the same superscript letters.

[2] Research context: RT = research trial, RT-E = embedded within a research trial, RT-E (TP) = embedded within a research trial (training phase), DS = dissemination study, TS = training setting, RCP = routine clinical practice.

[3] Competence Measures: CTS = Cognitive therapy scale, CTS-R =Cognitive therapy scale-revised, CTS-Psy = Cognitive therapy scale psychosis, CCCT = competence checklist for cognitive therapists, CTACS = cognitive therapy adherence and competence scale, YACS = Yale adherence and competence scale, MCSTPD-GCI = Multicentre Collaborative Study for the Treatment of Panic Disorder- Global Competence Item, MACT-RS = Manual-Assisted Cognitive Behaviour Therapy Rating Scale, CCTS = Cognitive therapy self-rating scale, ETBF = Evaluation of Therapist's Behaviour Form, PRTF: Patient Report of Therapy Form, SRF = Supervisor rating form, SSRF = Student self-rating form

[4] Therapist / assessor profession: MH Nurse = psychiatric/ mental health nurse, OT = occupational therapist, CSL = counsellor (level of training variable), AS = assistant psychologist, Psy = psychologists (exact level of training not specified), Clin Psy= clinical psychologists, PhD Psy = psychologist with a PhD, Ed Psy = educational psychologist, MA Psy = psychologist with a masters, Psy UG = undergraduate psychology degree, GP = doctor with a medical degree, Psych = psychiatrist, RF = research fellow (discipline not specified), CBT T = completed some form of CBT training, usually in a higher education setting, (e.g. diploma in CBT, IAPT training), SW = clinical/ masters level social worker, BABCP = therapists meeting BABCP accreditation standards, GMHW = Graduate mental health worker, Other MHP = other mental health practitioner. The label "trainee" indicates that the therapist is currently undergoing the specified level of training. Where names of the assessor were given but the profession of the individuals at the time of publication were not presented in the MS, this information was left as blank.

[5] Independence of assessors defined as: Research = member of the team involved in conducting, analysing or reporting the study, Supervisor= involved in providing training or supervision & where available further information is given in brackets- TS= therapist supervisor, i.e. provides training/supervision to the therapist(s) being assessed, CS = course supervisor, i.e. involved in providing training or supervision within the research or training setting but not to the specific therapist(s) being assessed, Independent = neither a member of the research team or a supervisor/trainer, Therapist = self-rating made by the therapist who is being assessed.

[6] Patient diagnosis: Dep. = depression, SP = social phobia, PD = panic disorder (+ AG = with agoraphobia), OCD = obsessive compulsive disorder, GAD = generalised anxiety disorder, SAD = separation anxiety disorder, SO = schizophrenia, SOA = schizoaffective disorder, Dep + Anx = mixed anxiety & depression, DEM = dementia, Personality disorder = Pers D, SH= self-harm, SUD = substance use disorder, Psyc = psychosis, Anx = anxiety, ASPD = antisocial personality disorder, BP = bipolar disorder

[7] Information obtained through personal communication with study authors

CHAPTER TWO:

Expert Insight into the Assessment of Competence in Cognitive Behavioural Therapy: A Qualitative Exploration of Experts' Experiences, Opinions and Recommendations

Chapter Overview

The current study utilises qualitative methodology to explore expert's (*N* = 19) experiences of conceptualising and assessing the competence of therapists delivering cognitive behavioural therapy (CBT). The study aimed to offer insight into the way in which experts define and evaluate CBT competence and highlight ways in which the assessment of CBT competence could be further improved. Semi-structured interviews were used to explore participants' experiences and interview transcripts were then analysed using Interpretative Phenomenological Analysis in order to identify commonalities and differences in the way CBT competence was evaluated. Four super-ordinate themes were identified: (1) what to assess: the complex and fuzzy concept of CBT competence; (2) how to assess CBT competence: selecting from the toolbox of assessment methods; (3) who is best placed to assess CBT competence: expertise and independence; and (4) pitfalls: identifying and overcoming assessment biases. Priorities for future research and ways in which the assessment of CBT competence could be further improved are discussed in light of these findings.

Acknowledgements:

With thanks to all of the participants who gave up their valuable time to take part in this study and without whom this research study would not have been possible. These include Peter Armstrong, David Clark, Chris Fairburn, Melanie Fennell, Daniel Freeman, Nick Grey, Helen Kennerley, Willem Kuyken, Martina Mueller, Pamela Myles, Steve Pilling, Candida Richards, Tony Roth, Roz Shafran and David Westbrook.

Contents

Introduction 76

Method 78

Ethics 78

Participants 78

Data Collection 79

Data Analysis 81

Validity Checks 81

Results 82

(1) What to Assess: The Complex and Fuzzy Concept of CBT Competence 82

1.2. More than a CBT technician. 82

1.2. The challenge of diversity under the broad CBT umbrella. 84

1.3. Generic vs. disorder specific competence. 86

1.4. Scales used to assess CBT competence have come to define it. 87

(2) How to Assess CBT Competence: Selecting from the Toolbox of Assessment Methods. 89

2.1. Assessing knowledge and understanding. 89

2.2. You can talk the talk but can you walk the walk? 90

2.3. Gold standard vs. feasibility. 93

(3) Who is Best Placed to Assess CBT Competence: Expertise and Independence 95

3.1. Assessor experience and expertise. 95

3.2. The pros and cons of using supervisors as assessors. 97

(4) Pitfalls: Identifying and Overcoming Assessment Biases 98

4.1. Some patients will make stars of us all. 98

4.2. Therapist self-presentation biases. 99

Discussion 101

Summary of Results 101

Refining the Toolbox of Assessment Methods 101

Multi-source Competency Assessment Programmes 103

Combining Formative and Summative Assessments 103

Recognising Assessment Consequences 104

Limitations 104

Future Directions 105

References 106

Appendices 110

Appendix 1: Ethical Approval 110

Appendix 2: Information Sheet 111

Appendix 3: Consent Form 114

Appendix 4: Template for Initial Email Invitation to Participate 115

Appendix 5: Demographic Form 116

Appendix 6: An Example of an Analysed Transcript: Participant no. 6 117

The theme which the text was assigned to is indicated in the far right column headed “theme”. Please note that the referenced theme is that which was coded at time of transcription. The themes have therefore changed in subsequent analysis and write up. 140

Appendix 7: An Example of a Theme and Associated Interview Extracts 141

Appendix 8: Invitation to Provide Respondent Feedback. 152

Introduction

The assessment of competence in the delivery of cognitive behavioural therapy (CBT) has long been considered to be of central importance in process and outcome research. This is because it is necessary to demonstrate that a treatment was competently delivered in order to establish the integrity of the treatment being investigated (Waltz, Addis, Koerner, & Jacobson, 1993). The need to identify an optimal strategy for assessing CBT competence has gained further prominence in recent years due to an increased focus on the dissemination of CBT into routine practice. This movement has intensified the need for effective measures which can provide a means of assessing the training of new CBT therapists and ensuring the quality of treatment provision within routine clinical practice (McHugh & Barlow, 2010).

Despite the need for accurate and efficient methods of assessing CBT competence, progress has been somewhat limited. The literature review outlined in Chapter one identified ten key methods for assessing CBT competence, all of which require further evaluation or refinement (Muse & McManus, 2013). In principle, therapists' knowledge is the easiest aspect of CBT competence to assess. However, there are few standardised multiple choice questionnaires with demonstrated validity and reliability and the validity of essays has not yet been established. The reliability, validity and utility of case reports has also been questioned and there is a lack of standardised short-answer clinical vignettes with established psychometric properties. Assessing whether therapists have the skills to apply their knowledge in clinical situations has proved especially challenging. Some of the methods used to assess clinical skill have significant limitations and thus are of limited usefulness. For example, patient outcome data provides an indirect, confounded and limited indication of therapists' competence. Questions have also been raised regarding the feasibility and validity of inferring therapist competence from patient surveys. Additionally, the ability to accurately rate one's own CBT competence is highly questionable. Supervisory assessments have the potential to provide more robust assessment of therapists' clinical practice, although further psychometric evaluation and refinement of currently available measures is needed. The use of standardised role-plays also has potential, but little progress has been made in this domain to date. The current 'gold standard' is assessor-ratings of therapists' in-session performance. However, this method is also problematic due to the lack of empirically evaluated rating scales with adequate reliability, validity and feasibility.

In sum, a range of different methods can be used to assess different aspects of CBT competence, all of which have their own inherent advantages and disadvantages. Given the limitations of the evidence-base for the use of existing CBT competence assessment methods,

there are currently no clear guidelines as to how CBT competence should be assessed. Thus little is known about when, why and how a given assessment method, or combination of methods, are selected and used in practice. The current study uses a qualitative methodology to shed light on this issue by examining the experiences of, opinions about, and recommendations for assessing the competence of CBT therapists from the perspective of CBT experts with a background in assessing CBT competence.

A qualitative exploration of the way in which CBT competence is assessed has potential to contribute to the further development of CBT competence assessment in a number of ways. First, Chapter one argues that there is need for further refinement of existing measures and/or the development of novel tools for assessing CBT competence (Muse & McManus, 2013). When developing assessment measures it is important to consider pragmatics, such as feasibility, usability and acceptability as without these features measures are unlikely to be adopted in practice (DeVellis, 2012). Qualitative enquiry identifies recurrent, unifying concepts which characterise the experiences of a number of individuals (Patton, 2002; Smith, Flowers, & Larkin, 2009; Sofaer, 1999) and thus could usefully identify which methods experts generally view as feasible, acceptable and applicable across a broad range of situations. This would enable future efforts to refine current assessment methods to focus on those methods which are most likely to be widely implemented in practice. Drawing upon the collective experience of a number of individuals with practical involvement in assessing CBT competence may also highlight problems and pitfalls with existing tools, as well as yielding useful information as to how assessment measures could be further developed and refined.

Second, frameworks which delineate the knowledge and skills required to competently deliver CBT (e.g. Bennett-Levy, 2006; Rodolfa et al., 2005; Roth & Pilling, 2007) provide a comprehensive definition of CBT competence. However, they also draw attention to the fact that assessing CBT competence is far from straightforward. Indeed, the models suggest that assessors must consider multiple inter-related dimensions in order to determine a therapist's competence. Qualitative research is well suited to increasing understanding of complex, multifaceted phenomena, such as CBT competence (Patton, 2002; Smith et al., 2009; Sofaer, 1999) and thus has the potential to highlight the ways in which assessors draw upon these multi-faceted definitions when examining the competence of CBT therapists. Third, sharing the experiences of experts could improve awareness and understanding of the practical issues involved in assessing CBT competence. For example, qualitative enquiry could help to identify any difficulties which commonly occur during the assessment process, together with ways these problems can be overcome. Finally, as well as revealing commonalities between experts in the field, qualitative

enquiry could also be helpful in highlighting areas where experts differ in their experiences, viewpoints or opinions (Patton, 2002; Smith et al., 2009; Sofaer, 1999). By exploring both convergence and divergence between expert opinion and experience, qualitative enquiry could serve to stimulate discussion and debate about how to assess CBT competence.

The current study therefore uses qualitative methodology to explore the experiences, understandings, and views of assessing the competence of CBT therapists from the perspective of nineteen CBT experts with a background in assessing competence. In line with the open-ended nature of qualitative enquiry, the study had three broad aims: (1) to explore what participants looked for when assessing therapists' competence (i.e. how they defined CBT competence), (2) to examine the way in which participants went about assessing the competence of CBT therapists (i.e. when, why and how assessment methods were selected and used), and (3) to reveal any difficulties which participants experienced during the assessment process.

Method

Ethics

Ethical permission for the study was obtained by the University of Oxford Research Ethics Committee (see Appendix one). All participants provided written informed consent prior to the interview (see Appendix two and three).

Participants

'Experts' were broadly defined as experienced CBT therapists whose research or clinical activities focused on the assessment and evaluation of CBT competence. Individuals were identified as experts through their involvement in the development and provision of BABCP accredited CBT training courses and/or recent publications related to the training or assessment of CBT therapists. Twenty six participants were invited to take part in the study via email (see Appendix four), of which five did not respond and two declined. Thus nineteen participants took part in the study. Given the absence of an ideal sample size for qualitative research (Smith et al., 2009) and the large variation in the number of participants included in previous Interpretative Phenomenological Analysis studies (range = 1- 42; Reid, Flowers, & larkin, 2005), the present study sought to recruit a comparatively large sample size in order to ensure that the study captured the experiences and opinions of a broad range of experts across the field.

To situate the sample participants completed a short demographics questionnaire (see Appendix five). The ten female and nine male participants ranged in age from 40 to 63 years (*M* =

51, *SD* = 7.45). All participants were experienced CBT therapists who had been practicing CBT for between 13 and 34 years (*M* = 23.68, *SD* = 6.66). Sixteen participants had treated more than 200 CBT cases, with the other three having treated 50 – 200 CBT cases. Fourteen participants were clinical psychologists, four were mental health nurses and one was a psychiatrist. Participants were evenly distributed between those who worked primarily within a training setting (*n* = 7), routine clinical practice setting (*n* = 7) and research context (*n* = 6) and all participants had involvement in the provision of CBT supervision and delivery of CBT training.

Data Collection

Individual semi-structured interviews were used to provide in-depth exploration of each expert's understanding of and approach to assessing CBT competence. In line with the recommendations of Smith et al. (2009) and Kvale (1996), an interview schedule (see Table one) consisting of a series of open-ended questions and minimal prompts was used to invite participants to (1) describe how they conceptualised CBT competence, (2) provide a narrative description of their experience of assessing the competence of CBT therapists and (3) reflect upon this experience in terms of what they were looking for, what they based their assessment on, the factors they took into account during the assessment, the reasons for assessing competence and any difficulties experienced during the assessment process. The schedule was used flexibly to guide the interview process, with participants being encouraged to take a strong role in leading the discussion. Interviews were conducted by Kate Muse and lasted between 28.29 and 60.10 minutes (*M* = 43.73, *SD* = 9.63). All interviews were digitally audio recorded and transcribed verbatim using *NVivo* transcribing software. Any personal or identifying information regarding the participant or any other individuals was removed at the point of transcription.

Table 1

Summary of interview schedule

Open ended narrative questions	Example prompts
• What is your understanding of the term 'CBT competence'? I.e. what makes a competent CBT therapist?	• Could you tell me more about what you mean by that? • Have you ever encountered this? • Do you have any examples which you could tell me about?
• Can you tell me about your experience of assessing the competence of CBT therapists?	• When was this? • What was the context? • Why were you assessing therapists' competence? • How did you actually go about assessing therapists' competence? • Can you describe to me what happened? • What did you base your assessment on? • Why did you do that? • What were you looking for? • Did anything make it difficult to assess therapists' competence? Could you say something about those difficulties? • In your experience, who typically assesses therapists' competence? Do you feel they are well placed to assess competence? Why is that? • In your experience, what would you say was the 'ideal' method of assessing CBT competence? Why? • Are there any methods of assessing CBT competence which you have found don't work very well? Why?

Note- At the end of the interview, all experts were also asked whether there was "anything we haven't spoken about which you feel would be relevant or important?".

Data Analysis

Verbatim interview transcripts were analysed using Interpretative Phenomenological Analysis (IPA), a thematic approach which aims to provide insight into how a person makes sense of a given experience or phenomenon (Smith et al., 2009). IPA analysis was chosen for a number of reasons (Brocki & Wearden, 2006; Smith et al., 2009). First, IPA gives central focus to understanding how participants make sense of their unique lived experiences. Second, the analytic process is co-consutructed between researchers and partcipants (i.e. IPA recognises the active role of the researcher in interpreting the data). Third, IPA involves generating and developing ideas and theories from the data, rather than applying pre-existing theoretical perspectives to the data (i.e. it is a 'bottom-up' rather than 'top-down' process). Fourth, IPA is also ideally suited for exploration of novel and complex research areas because it provides a summary of key emergent themes whilst also offering an in-depth interpretative analysis of participants' experiences, thus highlighting both convergence and divergence within the sample.

In line with IPA guidelines (Smith et al., 2009), a cyclical analytic approach was employed using a qualitative analysis software package (NVivo). First, a close line-by-line analysis of the first interview transcript was conducted and initial descriptive and conceptual comments were noted along with any ideas for emergent patterns (themes) (see Appendix six for an example of an analysed transcript). Second, a list of emerging theme titles was created and used to group together extracts of the transcript relating to each theme. Third, the list of themes was reviewed and, where appropriate, themes were re-structured, amalgamated and subsumed within a super-ordinate theme in order to create an overarching thematic structure. The first three stages were repeated for each transcript in turn, with the emerging thematic structure being updated and adapted in light of new information. Kate Muse took the lead in analysing the data and each stage of analysis was overseen by Dr Freda McManus who verified that the analysis had been systematically achieved and was supported by the data.

Validity Checks

A number of checks were carried out in order to help improve the accuracy, credibility and validity of the research findings (Flick, 2009; Yardley, 2000). Validity checks carried out at the analytic stage included re-examining all transcripts in light of the final thematic structure, examining the number of themes assigned to each transcript to ensure that all participants were adequately represented and reviewing extracts assigned to each theme to ensure the analysis was representative of the data (see Appendix seven for an example of a theme and associated extracts). Additionally, upon completion of the study, a respondent validation check was

conducted. This involved sharing findings of the study with the each participant and inviting them to comment on the interpretations and conclusions drawn (see Appendix eight for the invitation to provide respondent feedback).

Results

Four super-ordinate themes emerged from analysis of the data: (1) what to assess: the complex and fuzzy concept of CBT competence; (2) how to assess CBT competence: selecting from the toolbox of assessment methods; (3) who is best placed to assess CBT competence: expertise and independence; and (4) pitfalls: identifying and overcoming assessment biases. Super-ordinate and sub-ordinate themes are outlined in Table two and described, interpreted and illustrated with exemplar verbatim quotations[1] below.

(1) What to Assess: The Complex and Fuzzy Concept of CBT Competence

To discuss participants' experiences and opinions of assessing the competence of CBT therapists it is first necessary to examine what they understood CBT competence to be. In line with Barber, Sharpless, Klostermann, and McCarthy's (2007) concept of 'limited-domain intervention competence' and Kaslow's (2004) notion of 'intervention competence', participants focused on therapists' ability to appropriately, effectively and skilfully deliver evidence-based CBT treatments. Hence the assessment of broader aspects of CBT competence (e.g. use of supervision, ethical practice etc.) were not discussed. Discussion about what participants looked for when assessing therapists' ability to effectively deliver CBT yielded four sub-themes, which are outlined below. These sub-themes highlight the challenges and complexities involved in defining and disentangling the complex and fuzzy concept of CBT competence.

1.2. More than a CBT technician.

Participants observed that a competent CBT therapist needs access to a large tool bag of CBT techniques, akin to the 'specific CBT techniques' outlined in Roth and Pilling's (2007) competence map (e.g. behavioural experiments, thought records etc.). However, participants raised concerns that assessments which focus too tightly on a therapist's ability to perform technical behaviours fail to adequately assess CBT competence:

[1] In the quoted extracts, editorial elision by the author is indicated by three dots (. . .) whilst editorial insertion is indicated by [square brackets].

> "*I actually do think that there's an aspect of CBT that often does get dropped in favour of it being seen as a series of techniques and that is this sense of a fundamental philosophy . . . The tragedy is that people have started to see a lot of CBT as a series of activities- do a thought record, ermm err, do this, do that, so it's all very over technique bound and people lose sight of the fact that, ermm, well look why are you doing this?"* P11

When assessing CBT competence, participants therefore looked for evidence to indicate that therapists were not overly procedural "*CBT technicians*" (P5) who mechanically applied a series of CBT techniques. Participants identified three competences which differentiate a competent CBT therapist from a CBT technician. First, CBT interventions should be embedded within general clinical skills. These general skills echo the 'generic therapeutic competences' outlined by Roth and Pilling (2007), and include interpersonal and two-way communication skills, the capacity for empathy and warmth, and the ability to foster and maintain a sound therapeutic relationship. Although not sufficient in isolation, these skills were seen as an essential foundation for CBT competence, in that the effective delivery of CBT techniques was not possible without them. Indeed the absence of these general skills was a striking indication of a lack of competence:

> "*You know, I think one of the first things one is looking for is nothing to do with CBT specifically, but do people have the ability to form a therapeutic, helpful relationship with other human beings . . . can they be warm and empathic and supportive and all of those kinds of things, is certainly one of the first things. Certainly that, if it is not there, makes you think 'eek'*." P7

Second, therapists' should make treatment choices (e.g. which, when and how to deliver CBT interventions) on the basis of a sound understanding of underlying CBT theory, principles and research:

> "[CBT competence is] s*kills growing out of an understanding of, ermm, theory and principles. So it's not just a bunch of techniques in other words. We are talking here about a coherent errmm understanding of the nature of difficulties. And coherent treatment programs directly related to that.*" P5

Third, there was a sense that a competent CBT therapist should have the capacity to select appropriate interventions and deliver them in a manner commensurate with their patient's needs, both in terms of the individual person and the disorder they present with. The ability to deliver CBT techniques within the context of all three competences in combination was viewed as being indicative of CBT competence:

> "*I'd want to know that someone's got the technical skills, but that's not enough. I'd want to know that they are aware of the theory, that's going to inform the use of the technical stuff, but that's not enough. I want, ermm, evidence that they can marry theory with real life clinical practice, but that's not enough. Because I was to see that they're doing that within a quality therapeutic alliance that is characterised by collaborative empiricism.*" P4

In sum, participant's interview transcripts indicate that CBT competence assessments should not only look for evidence that therapists are able to perform a number of technical skills but should also examine whether therapists can deliver techniques skilfully, flexibly and appropriately in line the individual patient's formulation, CBT theory and research.

1.2. The challenge of diversity under the broad CBT umbrella.

Participants commented that, due to the broad range of treatment approaches under the CBT umbrella, there is a lack of agreement within the field about what CBT is, the way it should be applied and therefore what assessors consider competent. As P12 commented "*there are lots of different ways to get to the top of the mountain*". Participants reported that opinions regarding best practice often varied between therapists working within the same protocol. However, disagreement over what competent CBT involves was thought to be especially apparent when patients do not clearly fit a specific diagnosis, have co-morbid diagnosis, or present with disorders for which there are not yet agreed protocols or for which there are several different evidence-based protocols. For example, P3 commented on the debate surrounding the treatment of patients who do not clearly fit a specific diagnosis:

> "*There's a, a more general debate about using CBT. . . with an ordinary clinical patient who may not come in with pure diagnosis, ermm, and there are different views about how CBT should be done with those sort of patients . . . Ranging from people who probably advocate that, well you do one protocol and then you do another, and if they . . . don't have something that meets, you know, psychiatric diagnosis you shouldn't be seeing them. Ermm, through to people who would use a more formulation driven CBT or schema-focused CBT to work with more or less anything. So I suppose sort of, in terms of how competence is assessed, that relates to what view you might take on that debate.* " P3

In sum, CBT was viewed as a class of treatments within which there are many different ways of working. This divergence resulted in inconsistencies in assessors' opinions of what can be considered competent delivery of CBT and thus hindered the process of conducting reliable assessments of CBT competence.

Table 2

Summary of Themes

Superordinate themes	**Summary of superordinate themes**	**Subordinate themes**
1. What to assess: The complex and fuzzy concept of CBT competence	Reflects the challenges involved in conceptualising and defining the key competences assessors look for when assessing CBT competence.	**1.1** More than a CBT technician **1.2** The challenge of diversity under the broad CBT umbrella **1.3** Generic vs. disorder specific competence **1.4** Scales used to assess CBT competence have come to define it
2. How to assess CBT competence: Selecting from the toolbox of assessment methods	Provides insight into the narrow range of methods used to assess therapists' knowledge and skills and draws attention to the discrepancy between the ideal multi-method approach and what is feasible in practice.	**2.1** Assessing knowledge and understanding **2.2** You can talk the talk but can you walk the walk? **2.3** Gold standard vs. feasibility
3. Who is best placed to assess CBT competence: Expertise and independence	Highlights the complexities involved in identifying the optimal level of experience and independence needed to assess CBT competence.	**3.1** Assessor experience and expertise **3.2** The pros and cons of supervisors as assessors
4. Pitfalls: Identifying and overcoming assessment biases	Outlines issues which are thought to bias CBT competence assessments and suggests ways these could be overcome.	**4.1** Some patients will make stars of us all **4.2** Therapist self-presentation biases

1.3. Generic vs. disorder specific competence

Participants recognised the tension between assessing generic aspects of CBT competence ('basic CBT competences') and assessing competence within the context of a specific disorder or treatment protocol ('problem-specific competences'; Roth and Pilling [2007]). Some participants argued that, although basic CBT competences should also be assessed, it is not possible to adequately assess CBT competence without focusing on the specific protocol within which the therapist is working:

> "*For me, competence is treatment specific and there are many different therapies under the label CBT . . . I can't see there being a useful single measure of competence in CBT. Now I can see that there are going to be commonalities. In other words, if you had a measure of competence for CBT for depression or CBT for GAD, that there'll be some features in common. There might be, sort of, some core constructs that you'd want to assess that would be shared by these approaches. But there would also be some protocol-specific features that . . . wouldn't be relevant to another protocol.*" P13

Other participants agreed that a protocol-specific approach was feasible and necessary within a research context, where it is important to "*make sure everyone is really singing from the same hymn sheet"* (P4). However, though the importance of taking both the disorder and protocol into account when assessing competence was noted, there was uncertainty as to whether disorder or protocol-specific competence measures could realistically be implemented outside of a research context. Participants raised concerns that, as there are many different treatment protocols, this approach would result in multiple different scales or "*unwieldy instruments*" P11. Consequently this approach would be prohibitively costly to implement and would impose an unrealistic burden on assessors:

> "*You have to take the protocol into account. But if you take the protocol into account completely then the person assessing it has to know every protocol in detail. And there are quite a lot of them . . either for different disorders or within the same disorder. And I don't think you'll, you'd be able to find anyone who knows all of the protocols well enough to be able to assess someone else."* P1

Participants also observed that assessing protocol-specific competence may be less applicable in routine practice settings, where patient co-morbidity and complexity are common, and where therapists do not always deliver CBT in accordance with a specific protocol:

"*People on the whole that . . . I supervise . . . tend not to be applying CBT in a very protocol driven fashion . . . So there'll be adaptation on what they do according to the kind of problem they're working with, but it's mainly those transdiagnostic things that we're concerned about. And I would think they* [transdiagnostic things] *would be the most important in assessing competence really because, . . . if you've got those building blocks in place, it allows for flexibility and, when you meet a complex case or you meet somebody who doesn't fit in a particular box, you've still got your back to basics core skills that you can apply to try to tackle the problem.*" P8

"*We know there are evidence-based protocols which are very powerful . . . We also know that in routine clinical practice life is more complicated than that. I think that there is a risk that in homing in on, ermm, very tightly defined models and protocols, ermm, we may lose things which are actually quite important. For example, the capacity to generalise to other problem areas, the capacity to deal with co-morbidity, complexity, ermmm, different populations, different settings, all that kind of thing."* P5

In sum, all participants noted the importance of assessing basic CBT competences. However, there was uncertainty about whether the disorder and protocol should merely be taken into account during assessment or whether this should be a central focus, with participants opinions being somewhat dependent on the purpose and context of the assessment.

1.4. Scales used to assess CBT competence have come to define it.

Participants commented that one of the main difficulties faced when assessing competence was that defining and operationalising the specific skills involved in the competent delivery of CBT is far from straightforward:

"*The kind of behaviours that we're trying to observe are inherently difficult to define I think. You know, it's a bit like the old cliché of 'I don't know very much about art but I know what I like' and clearly there's something like that goes on with therapy skills, you know. It is hard to say exactly what it is but I know it when I see it.*" P7

It was evident from participants' transcripts that rating scales, particularly the Cognitive Therapy Scale/Revised (CTS-R: Blackburn et al., 2001; CTS: Dobson, Shaw, & Vallis, 1985), were viewed as useful assessment tools in part because they helped to overcome this difficulty by operationalising the specific skills a CBT therapist should display within a competent therapy session. However, some participants observed that rating scales had actually come to define CBT competence for

assessors and therapists alike. For example, when outlining his definition of CBT competence, P3 commented on how readily items of the CTS/R came to mind:

> "*Because I'm used to assessing CBT competence using scales* [CTS/R] *that is how I tend to categorise it . . . I'm virtually going through the boxes as I'm talking to you!"* P3

It is widely recognised that assessment drives learning and that learners will employ strategies to optimise their chances of successfully passing summative assessments (Van der Vleuten, Schuwirth, Scheele, Driessen, & Hodges, 2010). It is, therefore, perhaps not surprising that participants highlighted the key danger in seeing the CTS/R as synonymous with CBT competence was that it could lead therapists to work towards achieving good CTS/R ratings, rather than focusing on delivering quality CBT. As P7 stated "*students sometimes are more concerned with getting a good CTS score than they are with doing a good CBT session*". Participants viewed this strategy as counterproductive as it typically led therapists to deliver CBT in an artificial way in an attempt to try and *"tick all of those boxes"* (P3) on the scale, thus resulting in poorer treatment delivery:

> "*The trainees also get a bit caught up in 'I've done the agenda, I've done this, I've done that'- rather than just doing the therapy, you know."* P16

> "*Not every session should show all of the competences every time and I think what happens definitely in training students is that they begin to do things artificially. So I know that some of my sessions I wouldn't score highly on every single one of those measures. But actually I would think that was an OK session. And a session where I really worked to score highly on those* [rating scales] *I wouldn't think was necessarily optimal for the patient*." P2

Participants noted that therapists' concern with achieving high marks on the rating scales could sometimes be overcome by providing advice, support and training in how to cope with being assessed, thus enabling them to "*let it go a bit*" (P18). Participants also identified the need for trainers and supervisors to recognise and resist the temptation to teach therapists how to pass assessments, rather than how to deliver effective CBT:

> "*One of the risks is the temptation to teach to your assessment, rather than teaching to need. Ermm. And a lot of our students in fact will say 'teach me to pass this exam'. No I'll teach you to do therapy. If you do therapy you'll pass the exam . . . we shouldn't teach to the assessment.*" P9

Overall, participants identified rating scales which specify the knowledge and skills involved in the competent delivery of CBT as being necessary to assess CBT competence. However, they also highlighted the need for assessors and therapists alike to recognise that these are merely

assessment tools and should, therefore, not be the primary driving force behind CBT training or treatment delivery.

(2) How to Assess CBT Competence: Selecting from the Toolbox of Assessment Methods

As outlined in Chapter one, ten key methods can be used to assess CBT competence. The three sub-themes in this section provide insight into when and how participants used these methods to assess CBT competence. Miller's (1990) framework for assessing clinical skill separates skills assessments into four hierarchical levels: knowledge ('knows'); practical understanding ('knows how'); skills ('shows how'), and clinical practice ('does'). The first sub-theme outlines how participants went about assessing therapists' knowledge ('knows') and practical understanding of CBT ('knows how'). The second sub-theme discusses how participants assessed therapists' ability to draw on and apply this knowledge ('shows how' and 'does'). The final sub-theme moves on to consider the disparity between participants' ideal 'gold standard' approach to assessing CBT competence and what was realistic and feasible in practice.

2.1. Assessing knowledge and understanding.

Participants agreed with Roth and Pilling's (2007) assertion that the basic foundation of CBT competence is a sound understanding of the scientific, theoretical and contextual basis of CBT and therefore noted the importance of assessing therapists' knowledge of CBT theory, principles and research and their ability to relate this knowledge to clinical practice. As outlined in Chapter one, four key methods can be used to assess therapists' knowledge and practical understanding of CBT: multiple choice questionnaires (MCQs), short-answer clinical vignettes, case reports, and essays. Interestingly, none of the participants in this study reported using MCQs or short-answer clinical vignettes to assess therapists' knowledge. Essays were however used within formal educational settings and case reports were used for accreditation purposes (e.g. for the British Association of Behavioural and Cognitive Psychotherapies) as well as within the context of CBT training. Outside of formal education and accreditation settings, participants primarily assessed therapists' CBT knowledge within supervision sessions by examining case formulations and through verbal discussion of therapists' cases. This gave participants an insight into whether the therapist's clinical work was grounded in a sound understanding of underlying CBT theory, principles and research:

> "*I assess it* [CBT knowledge] *on their talking about CBT . . . do they seem to understand why they are doing what they are doing? Do they seem to be able to formulate this person's problem in a cognitive-behavioural model and then use that to decide which CBT interventions to carry out? So, sort of, knowledge about CBT practice from what they talk about."* P1

> "*You're thinking more often about the way which people discuss their cases, whether it evidences that they are able to formulate problems using the CBT model. That their discussion and thinking about the case and the way they are working with the person is informed by clear CBT model and that the kinds of things they are doing and planning to do fit with that, they make sense within the model and, and within the formulation for the client."* P8

Thus, although a number of methods can be used to assess therapists' knowledge of CBT, these were either not used at all or were employed only within formal education or accreditation settings. Instead, participants relied on case discussion and reviewing case formulations to assess therapists' understanding of the scientific and theoretical basis of CBT.

2.2. You can talk the talk but can you walk the walk?

Participants noted that demonstrating CBT knowledge was not sufficient to infer CBT competence. Instead, examining whether therapists have the skills necessary to draw on and apply this knowledge in clinical situations was considered central to CBT competence assessments:

> "*You can assess, ermm, both of those things* [theoretical knowledge and knowledge of clinical principles] *through . . . essays and written case presentations. What you . . . get then is whether they actually know in theory what's what and how to do it. But you don't get, you know, a real, accurate indication of whether they can actually do it in practice. And to me that is the essence of competence. It's not just knowing the theory, it's actually knowing how to use it.*" P5

Chapter one identified six methods for assessing CBT skills: self-assessment, patient surveys, patient outcome, supervisory assessments, assessor-rated treatment sessions, and standardised role-plays. Self-assessment was not used by participants as a summative assessment tool as it was believed to be inaccurate and prone to self-presentation biases. This belief is supported by a number of studies which have found that therapists' appraisals of their own competence typically differ from those made by assessors (Brosan, Reynolds, & Moore, 2008; Carroll et al., 2000; Mathieson, Barnfield, & Beaumont, 2010; McManus, Rakovshik, Kennerley, Fennell, & Westbrook, 2011). Participants did, however, note that self-assessment was useful in enhancing therapists' capacity for self-reflection and therefore used this method in combination with assessor-rated treatment sessions as a formative tool. The utility of this learning strategy is supported by an increasing body of literature highlighting self-reflection as an important mechanism for therapist development (Bennett-Levy, 2006; Laireiter & Willutzki, 2003).

Although not commonly used, participants did sometimes administer patient satisfaction surveys or engage in general discussions with patients about what they had learned and their satisfaction with treatment. However, these were most often used informally within training settings, rather than as formal assessment tools. Participants noted that using patient outcome data to infer therapists' skill in delivering CBT has a number of flaws. Most significantly this method does not take into account treatment-specific factors (i.e. improved outcome may not be due to delivering CBT) and is confounded by patient variables, such as patient difficulty and comorbidity. Despite the perceived fallibility of this method, participants noted the importance of assessing patient outcome, as the primary objective of CBT is to reduce the impact of patients' distressing symptoms. Hence participants often examined patient outcome data alongside other assessment methods:

> *"I mean symptom outcome . . . should be the focus. It doesn't mean that the therapist isn't competent, obviously it's not controlled. We don't know, it could be they were recovering anyway and whatever they would have done they would have got better. But if you want to have a marker, it is a decent one to have."* P10

Participants routinely carried out informal competence assessments on the basis of supervision sessions. Although it is considered best practice to view session material within supervision, participants acknowledged that case discussion alone is often relied upon to infer competence. This supports Townend, Iannetta, and Freeston's (2002) finding that the use of direct observational techniques is rare in routine practice. There was however a sense of scepticism about whether what a therapist says they do "*talking the talk*" is an accurate representation of what they actually do in practice "*walking the walk*" (P11). This scepticism was in part due to a recognition that supervisees may not always recognise or want to report important information during supervision. A study by Ladany, Hill, Corbett, and Nutt (1996) found that supervisees do often fail to disclose key material from therapy sessions to their supervisors, providing confirmation of this observation. However, participants also considered that talking about CBT skills and demonstrating those skills are distinct aspects of competence, akin to Bennett-Levy's (2006) declarative and procedural knowledge systems:

> "*I've met people who . . . have a wonderful declarative knowledge. They can talk beautifully and persuasively. You sit them in a room and they are not that good. There are other people who struggle ermmm with the declarative, kind of, stuff and are procedurally really good.*" P6

For the purposes of formal assessments, participants identified assessments based on direct observations of therapists' performance as being a more valid way of assessing clinical skill. As P14

commented "*I think you do need to observe what they did one way or another, at least on a few occasions*". Two key methods for assessing therapists' performance were identified: role-plays and assessor-rated treatment sessions. Participants had some experience of rating therapists' ability to conduct CBT assessments or implement specific CBT interventions on the basis of therapists' performance with an individual playing the role of a standardised patient. However, role-plays were more commonly employed informally within supervision in order to highlight therapists' strengths and weaknesses and enhance their skills development:

> "*Normally in supervision I do quite a bit of role-play where the person plays the client and I play the therapist, which enables them* [the therapists] *to learn how in practice you might deal with the situation. But I sometimes, as a sort of test, I do it the other way around where I play a client and ask them to, you know, ask me questions to see if they can ask they right sort of questions. So that's another way of assessing competence actually.*" P3

Participants commented that role-plays overcome many of the difficulties faced when assessing therapists' performance on the basis of viewing session recordings. For example, they reduce the impact of patient variability, remove bias in the selection of recordings and overcome practical and ethical difficulties faced when recording, viewing and storing real treatment sessions. However, participants also recognised that, as role-plays are constructed, artificial simulations, therapists' ability to perform within a role-play may not translate to an ability to perform within a real clinical encounter:

> "*It* [the role-play based method] *gets around patient variability . . . it gets round the length of therapy because you can have role-plays selected from various parts of the therapy, to do a whole little group of them which would sample across the therapy. Ermm. And so you can overcome those problems with the role-play method. It has this disadvantage that you don't know if the person is behaving as they would, or as they do with their actual clients, as a genuine problem."* P13

Some participants favoured assessments of therapists' performance which are based within real clinical practice settings. As P9 commented *"I think real practice is, is essential really. You don't pass your driving test without driving on the road*". The way participants assessed therapists' clinical practice varied depending on the context and purpose of assessment. Participants sometimes conducted 'in vivo' assessments by sitting in on therapists' treatment sessions, although practical difficulties meant this was rare. More typically participants rated audio or video recordings of therapists' treatment sessions. Recordings were usually mid-therapy sessions involving change strategies, although sometimes sessions were purposively sampled from the beginning, middle and end of treatment. In addition to session recordings, participants noted the

importance of having contextual information such as the session number, patients presenting problem, formulation, materials used during the session (e.g. thought records) and sometimes patient outcome data. On occasions case-reports were also used alongside treatment sessions to provide contextual detail and insight into therapists' understanding of CBT. Participants rated the skill with which therapists delivered CBT within treatment sessions using standardised rating scales. Within some research contexts the scales focused on protocol-specific competences, whereas in routine practice and training settings these typically focused on transdiagnostic or generic CBT competences. In particular, participants reported a heavy reliance on the Cognitive Therapy Scale/Revised (CTS-R: Blackburn et al., 2001; CTS: Dobson et al., 1985). Although the CTS/R was widely used, participants perceived it to be "*a bit of a wobbly instrument"* P11. Criticisms included a lack of applicability beyond depression, failure to account for disorder or protocol-specific competences or stage of therapy, under emphasis of the therapeutic relationship, unclear instruction manual and rating scale, ambiguity and lack of behavioural specificity in scale item descriptors, a high degree of assessor inference, poor inter-rater reliability and a lack of opportunity to provide detailed formative feedback.

In sum, assessing whether a therapist has the skills to draw on and apply their knowledge of CBT in clinical situations was considered essential when assessing CBT competence. In informal supervisory contexts participants often used case discussion as an indication of CBT skills. However, when formally assessing CBT competence, participants acknowledged that direct observations of therapists' performance were necessary. Role-plays were identified as a potentially useful method of observing therapist performance. However, formal, standardised role-plays were rarely employed and participants were uncertain whether performance in role-plays would translate to performance within real clinical practice settings. Instead participants typically used standardised scales (especially CTS/R) to assess therapists' performance on the basis of viewing treatment session recordings.

2.3. Gold standard vs. feasibility.

A multi-method approach involving large samples of an individual's performance is widely advocated to ensure that reliable and generalisable inferences can be made about an individual's competence (Kaslow et al., 2009; Sharpless & Barber, 2009; Van der Vleuten et al., 2010). Consistent with this viewpoint, participants considered the 'gold standard' for assessing CBT competence to be a multi-method assessment involving a number of observations of therapist skill (e.g. role plays, recordings of treatment sessions), sampled across a range of patients presenting with a variety of problems, in combination with an assessment of knowledge of underlying CBT theory and clinical principles (e.g. case discussion, written case reports, essays) and examination of

patient outcome data. Participants commented that this approach would enable the assessment of multiple aspects of CBT competence (i.e. therapists' knowledge, application of knowledge and skill) and would compensate for the inherent weaknesses of each method. Participants 12 and 13, for example, discussed the benefits of a multi-method approach:

> *"I mean if you talk about gold standard and how you would you ideally assess it I think it would be a mixture of live observation by a recording or some sort. Ermm, and case discussion, because we need to think about the case formulation, whether or how well they are theory grounded and, and look at their reasoning which is not necessarily absolutely apparent in the . . . live session that we would have. Get patient feedback and look at patient outcome ermmm. I think all of those, a combination of all of those together would be the best way of assessing it."* P6

> *"I suppose ideally you want you want to triangulate and you want to get data from different sources. So you want to see therapy tapes, you want to see, ermm, the treatment plans written down, you want to see the formulations written down, you want to see . . . their patient outcomes.* P12

Although participants advocated this 'gold standard' multi-method approach to assessing competence, they also recognised that it was prohibitively time consuming and expensive and thus of limited feasibility:

> *"I would imagine the gold standard of assessing competence is going to be observation by a number of expert observers of a number of samples of work with different methods of assessing them. So some observations of live practice, some observations of theoretical knowledge. And that's just not going to be practical for your local service."* P1

Indeed participants' reports of their experiences of assessing competence revealed that the way they had to assess competence in reality was often in stark contrast to this ideal 'gold standard' approach. Participant nine for example commented that competence was often assessed on the basis of only a few snapshots of the therapist's performance:

> *"I think you get a snapshot. So even with three* [session recordings] *you've got three snapshots. Ermm now does that say that that person is a competent therapist? Well you can't say that. What you can say is that in the opinion of these markers they've demonstrated competence on these three occasions. Now, to say this is a competent therapist we're going to have to demonstrate that that is the predominant position, so we, to be really hands on heart saying perhaps with confidence, we would need to see a lot*

> *more than that. Ermm, and unless the world changes utterly, that's just not going to happen."* P9

Other participants noted that, despite misgivings, they were often forced to rely on case discussion alone:

> *"I certainly think that the best way of doing that* [assessing CBT competence] *is what BABCP calls 'live' observation . . . What one actually settles for a lot of the time of course is the person's report of what they actually do and there are fairly obvious problems with that, but nevertheless, it's all one has to work with sometimes."* P7

> "*I think I'd be relying more on the self-report of the therapist. In therapy trials you've just got less time to go and listen to all of the tapes."* P10

There was a large disparity between the multi-method assessment involving large samples of an individual's performance which participants viewed as the ideal method of assessing CBT competence and the way in which they actually went about assessing CBT competence in practice. As this discrepancy was largely due to time and resource constraints, participants suggested that more feasible and cost-effective multi-method assessment packages may be necessary to bridge this gap.

(3) Who is Best Placed to Assess CBT Competence: Expertise and Independence

When considering the way in which CBT competence should be assessed, it is important not only to consider which methods should be used but also who is best placed to carry out competence assessments. Participants' discussions about who should assess competence centred on two key sub-themes, both of which highlight the complexity of this issue and illustrate a degree of uncertainty within the field about who is best placed to assess competence. The first sub-theme considers the amount of experience and understanding of CBT required by assessors, whilst the second sub-theme discusses what level of assessor independence is most beneficial.

3.1. Assessor experience and expertise.

Opinions about the level of CBT expertise required to adequately assess CBT competence varied widely. For example, P13 argued that novice assessors, with little or no experience of delivering CBT, should be able to assess CBT competence, providing the assessment tool being used had adequately operationalised the expected behaviour and that the assessor had been provided with training in how to use the tool. Under these circumstances, competence assessments could be carried out by "*someone intelligent, conscientious but not necessarily a therapist*" P13. In line with findings reported by Weck, Hilling, Schermelleh-Engel, Rudari, and Stangier (2011), other

participants held the contrasting viewpoint that, while novices can successfully identify whether certain techniques have been implemented (i.e. rate adherence), only those with CBT expertise can accurately identify whether techniques (1) have been skilfully implemented and (2) are appropriate within the context of the individual patient's presentation:

> "*A relatively junior person no doubt could do some basic ratings, they could spot when someone's doing something. Whether they could spot the level? So I think the distinction might be that you could probably get someone with less experience to pick up on adherence but actually identifying competence is a trickier matter*." P11

This view seemed to rest on the belief that assessing CBT competence is a difficult and complex skill which requires experience and expertise in CBT delivery and needs to be fostered and developed through experiential assessor training. Thus, participants endorsed the type of assessor training packages providing experience in and guidance about how to use competence assessment tools that are often used in order to achieve reliable ratings of therapists' skill (Barber et al., 2007; Reichelt, James, & Blackburn, 2003):

> "*I think it will help to have more experience of actually rating, the more experience you get of rating things and measuring competence the better you'll get. But more importantly alongside that is getting feedback and having some sort of training."* P17

Performance-based assessments of therapist skill (e.g. role-plays, treatment sessions) were regarded as involving a particularly high level of assessor skill, as was providing developmental feedback (as opposed to simply scoring scale items). For example, P6 reflects on the skills involved in assessing competence:

> "*The skill required to assess competency is enormous because you have to take procedural skill and then translate it into declarative knowledge so that you can describe what the person is doing in a way that you can put into very clear words and then use the words to guide what becomes another procedural skill . . . It's really quite technically, quite difficult to do . . . It's even why, why do I like what I've just heard? It can be a very simple snippet and it will take quite a lot of thinking.*" P6

Although these participants were in agreement that assessors needed themselves to be competent CBT therapists, there remained uncertainty as to the optimal level of assessor expertise. For example, some participants identified leading experts with years of experience in delivering CBT as being best placed to assess competence, whilst others voiced concerns that

therapists at the very top of the field may have unusually high expectations and therefore bias assessments.

In sum, all participants commented that the person carrying out CBT competence assessments should be provided with adequate training, regardless of their CBT expertise. However, participants' opinions about how much CBT experience assessors require ranged from inexperienced novices through to experts in the field. Hence there appears to be uncertainty regarding the optimal level of therapeutic experience required to assess CBT competence.

3.2. The pros and cons of using supervisors as assessors.

Participants reported routinely assessing their own supervisees' competence. This approach was viewed as having two key benefits. First, if participants knew the person they were assessing, they had more contextual information on which to make a judgment. In particular they had a better understanding of the specific patient's treatment context, as well as knowledge of the therapist's work more broadly. Second, when participants were supervising the therapist being assessed, they were better able to deliver feedback in a way that was appropriate to that supervisee. However, participants also highlighted a significant disadvantage to using supervisors to assess competence. Namely that assessment may be biased by social demands and influences, the dynamics of the relationship and the impact of negative feedback, all of which made it difficult for assessors to be objective. Similar concerns about the undesirable consequences of negatively evaluating supervisees have also been highlighted by supervisors within medical assessment settings (Dudek, Marks, & Regehr, 2005; Govaerts, van der Vleuten, Schuwirth, & Muijtjens, 2007) and evidence from a range of treatment modalities indicates that there is often low agreement between therapist competence ratings made by supervisors and independent judges, with supervisors generally providing higher ratings (Borders & Fong, 1992; Chevron & Rounsaville, 1983; Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012; Martino, Ball, Nich, Frankforter, & Carroll, 2009). The pros and cons of using supervisors to assess CBT competence were summarised by P1 and P7:

> "*If you know the person you know the context. In some ways that makes you much better at assessing competence. If you know because you've heard about it in supervision . . . what's happened in the run up to this session you are . . . in a much better position to judge whether it's appropriate work. On the other hand, you've also, you're also more likely to get lots of . . . social demands and influences on you in terms of, you know, if it's someone you have some level of relationship with or work with at some level, it's always difficult to give negative feedback.*" P1

> "*I think there are swings and roundabouts . . .* [using the therapist's supervisor] *has some advantages, ermm, in terms of having some knowledge of the therapist, ermm, some knowledge of the patient, some, therefore some possibly improved understanding of what's actually going on in the session that you're trying to rate . . . But clearly the potential disadvantage is that you're marking either out of sympathy or hostility, rather than strictly on what's actually happening in the session.*" P7

Although it is often more practical for therapists' own supervisors to act as assessors, the potential inaccuracy of supervisors' assessments was viewed by some participants as insurmountable. These participants therefore recommended that independent assessors should be used to assess CBT competence. Others participants, however, argued that this difficulty could be circumvented by double marking assessments or using independent assessors as a reliability check. Participants also noted that the decision to use supervisors or independent assessors may depend on the purpose of the assessment. For example, for formative assessments it may be advantageous to use supervisors who have knowledge of the broader picture and a relationship with the therapist, whilst for summative assessments it may be advantageous to use independent assessors. In sum, this theme highlights pros and cons to using both supervisors and independent assessors to assess therapists' ability to competently deliver CBT, thus illustrating the complexities involved in identifying who is best placed to assess CBT competence.

(4) Pitfalls: Identifying and Overcoming Assessment Biases

Participants raised a number of issues that could potentially bias CBT competence assessments. These can broadly be categorised as those which relate to (1) patients and (2) therapists, as outlined in the first and second sub-themes below. Discussion of these sub-themes focused on the observed impact of these factors as well as the ways in which the biases could be overcome.

4.1. Some patients will make stars of us all.

Participants noted that it is more difficult for therapists to deliver effective CBT and therefore to demonstrate competence with more complex cases. Conversely, straightforward cases were viewed as being less taxing for therapists. As P11 acknowledged, "*some patients will make stars of us all*". Although little research has been conducted in this area, evidence indicates that therapist competence differs depending on the particular patient. Imel and colleagues (Imel, Baer, Martino, Ball, & Carroll, 2011) found that a higher degree of baseline impairment in family/social functioning was associated with reduced competence in providing motivational enhancement therapy, although employment, alcohol, drug, legal, and psychiatric functioning at baseline did not predict competence. Similarly, Boswell et al (2013) reported that patient trait interpersonal aggression was

associated with lower competence in delivering CBT for panic disorder, although pre-session severity was unrelated to competence. Thus there is some evidence to suggest that patient complexity may indeed influence therapist competence.

Participants postulated a number of reasons why patient difficulty may impact on therapists' competence. For example, patients with only one 'straightforward' problem (e.g. panic disorder) would be easier to treat than patients with multiple comorbid problems or more complex difficulties (e.g. borderline personality disorder) which typically require a much higher level of general therapeutic skill as well as the skilful delivery of a broader range of complex CBT interventions. Participants also noted that patients do not always "*play the game*" (P8) in that they may talk too little or too much, not be psychologically minded, have low motivation, be reluctant to engage in treatment, or be experiencing social or environmental problems, all of which present challenges for the therapist attempting to implement a standard CBT protocol. The impact of patient difficulty on assessing therapist competence is illustrated by P6:

> "*If you have a really straightforward patient you try the first and obvious strategy it may work and you move on. And when you are listening to that therapy session it will sound like a competent therapy session. The therapist wasn't really tested very much. Whereas if you have a moderately complicated patient you might, you know, the strategy one and two may not work, you may have to try three or maybe even four, err, and to do that with some evidence and to keep going you know rather than bottling it. And that becomes a much higher level skill."* P6

Participants observed that the impact of patient difficulty and complexity can sometimes lead to unfair comparisons between therapists whose assessments are made on the basis of their performance with patients of varying difficulty. In particular, those working with more complex and challenging cases could be at a disadvantage, whilst those working with more straightforward cases are at an advantage. Thus participants highlighted the importance of assessors taking patient difficulty into account when making judgements about a therapists' competence. Participants also noted that this bias could be minimised by using standardised patients (e.g. in role plays) or by including multiple assessments of therapists' ability to deliver CBT across a range of different patient presentations and complexity.

4.2. Therapist self-presentation biases.

A further theme that emerged was that therapists may present themselves in a biased manner within competence assessments. Three possible self-presentation biases were discussed. First,

participants noted that therapists often put a positive spin on their performance in an effort to hide their mistakes:

> *"You know we've all got patients where we have deep shame about what it is we've failed to do, or indeed what we actually did, and then as a supervisee you try and hide these patients from your supervisor.*" P11

This bias particularly presented itself within self-assessments, verbal and written case discussion and also resulted in therapists selectively submitting recordings of their best treatment sessions for evaluation. It was thought that this bias was driven by a fear of exposing oneself as being incompetent, a fear which was especially apparent in 'high stakes' summative assessments. Participants recognised that such anxieties about being negatively evaluated not only affected trainees and junior therapists, but also senior therapists such as themselves:

> "*I think all of us are, you know, subject to what people sometimes call imposter syndrome- that one day they will all realise that actually I don't really know how to do this! Errmm and so I think exposing oneself in that way to scrutiny is anxiety provoking for many people. Almost everyone actually I would say.*" P7

Second, participants commented that a therapist's self-confidence often impacts on the way they present themselves during assessments. The degree of confidence therapists have in their own ability was thought to colour their own perception of their performance and thus any reports of their performance, especially within case discussion and self-assessment:

> "*If I look at the ratings my students give themselves, which they usually do as part of having an assessment, some of it just reflects confidence. There are some people who would score themselves really quite highly and I think 'what are you on?' and other people will score themselves very low . . . they are the self-critics.*" P6

Therapists were seen as largely falling into one of two groups: (1) those that are over confident and are not aware of their own flaws and (2) the "*worrying perfectionist group"* (P13) who are overly critical of themselves. This self-confidence bias may therefore provide an explanation as to why therapists are as likely to underestimate as to overestimate their own competence (Brosan et al., 2008; Carroll et al., 2000; Mathieson et al., 2010; McManus et al., 2011):

Third, participants observed that the anxiety therapists experience when their performance is being scrutinised (e.g. when recording treatment sessions, delivering role plays) can actually inhibit their performance. Although other participants commented that performance anxiety does not mask competence but demonstrates a limit to it, as a competent therapist should be mindful of and able to manage their own affect.

Participants highlighted a number of strategies for minimising the impact of therapist self-presentation biases. These included providing a warm and supportive assessment environment and making assessments after viewing at least some evidence of therapists' performance (e.g. in treatment sessions or role-plays) rather than relying on self-assessment, case-reports or verbal case discussion. Participants also suggested the routine use of session recordings so that this becomes the norm rather than the exception, thus making the process less anxiety provoking. This would also allow sessions recordings to be selected at random for evaluation so that assessors do not have to rely on therapist self-selection.

Discussion

Summary of Results

This study explored experts' conceptualisations of CBT competence and their experiences of assessing the competence of CBT therapists. Data analysis revealed four super-ordinate themes. First, *'what to assess: the complex and fuzzy concept of CBT competence'* reveals the challenges involved in conceptualising and defining the key competences assessors look for when assessing CBT competence. Second, *'how to assess CBT competence: selecting from the toolbox of assessment methods'* provides insight into the narrow range of methods used to assess therapists' knowledge and skills and draws attention to the discrepancy between the ideal multi-method approach and what is feasible in practice. Third, 'w*ho is best placed to assess CBT competence: expertise and independence'* highlights the complexities involved in identifying the optimal level of experience and independence needed to assess CBT competence. Fourth, *'pitfalls: identifying and overcoming assessment biases'* outlines issues which are thought to bias CBT competence assessments and suggests ways these could be overcome. Prioritise for future research and ways in which the assessment of CBT competence could be further improved are discussed in light of these findings.

Refining the Toolbox of Assessment Methods

A broad range of methods can be used to assess CBT competence, all of which have inherent strengths and weaknesses (Muse & McManus, 2013). However, participants in the current study viewed some methods as being more problematic than others. As a result, only a narrow selection of methods were implemented in practice, thus highlighting the importance of focusing future efforts on the development of methods which are perceived as credible, useful and valid by those involved in the assessment of CBT competence. When evaluating therapist's knowledge and practical understanding of CBT, assessors only employed structured measures within formal

educational settings, relying instead on case discussion within supervision. Standardised supervisor measures could therefore provide a useful means of assessing therapist knowledge in routine practice. Alternatively, broader assessments which assess both CBT knowledge and skills may be necessary. For example, case summaries could provide a useful adjunct to in-session performance ratings as they offer insight into whether therapists' clinical work is grounded in a sound understanding of underlying CBT theory, principles and research as well as providing essential therapeutic context.

Clinical skill was viewed as being at the heart of delivering competent CBT and was commonly assessed by viewing therapists' performance within recorded treatment sessions. Given the complexities involved in defining and specifying the skills involved in the competent delivery of CBT, rating scales which outline and behaviourally operationalise the skills involved in the competent delivery of CBT were viewed as providing a helpful framework for rating therapists' in-session performance. However, this study confirms the need for more comprehensive rating scales with improved validity, reliability and usability as outlined in Chapter one (Muse & McManus, 2013). Supervision was also frequently used to infer therapists' skills. However, this study supports Townend et al.'s (2002) finding that supervision is often based on second-hand reports, rather than performance-based observations. Given participants' strong preference for assessing clinical skill on the basis of direct performance-based observations, discussion-based supervisory assessments may be limited in their ability to serve as summative assessment measures. However, supervisory assessments still provide useful formative feedback and thus play an important role in training and on-going continuing professional development (Dochy & Moerkerke, 1997; Milne, 2007; Van der Vleuten et al., 2010). Although not widely used at present, it was suggested that role-play assessments warrant further exploration. Medical training settings routinely assess clinical skill using objective structured clinical examinations (OSCEs) comprising of a series of short encounters with standardised patients (Epstein, 2007; McNaughton, Ravitz, Wadell, & Hodges, 2008) and a similar approach has been advocated for assessing CBT skills (Fairburn & Cooper, 2011). However, research in this domain would need to address participants concerns that therapists' ability to perform within artificially constructed role-plays may not translate to an ability to perform within real clinical encounters.

Choosing the Right Tool for the Job

Selecting an assessment strategy is a complex process which will inevitably entail some compromises. The impact and relative importance of these compromises and thus the degree to which any given method is appropriate will, however, depend on the complex integration of a specific set of circumstances surrounding the assessment. Results from this study highlight a

number of factors which need to be taken into account when selecting an appropriate strategy for assessing CBT competence. It is important to weigh up the inherent strengths and weaknesses of each assessment method, as well as the validity, reliability, usability and feasibility of the specific tools available. The assessment context will impact on the utility of any given assessment method and will therefore also need to be considered. For example, protocol-specific measures may be required within the context of research trials, whereas transdiagnostic measures may be more appropriate within routine practice settings. The purpose of assessment should also guide the choice of assessment strategy. For example, the inclusion of formative feedback is not necessary for accreditation purposes but is essential for enhancing skills development within training settings. It will also be necessary to carefully consider the experience, expertise and independence of those who are available to conduct the assessments.

Multi-source Competency Assessment Programmes

The recognition that no single method in isolation is comprehensive or robust enough to measure the complex integration of knowledge and skills which constitute clinical competence has prompted a transition to multi-source competency assessment programmes within medical education (Van Der Vleuten & Schuwirth, 2005). Participants in the current study advocated a similar approach to assessing CBT competence involving multiple direct observations of therapist skill, knowledge-based assessments and examination of patient outcome. However, time and resource constraints proved substantial barriers to the implementation of such a multi-method approach. Thus, whilst further development of multi-source programmes for assessing CBT competence is warranted, it will be important to move beyond traditional psychometric evaluation to consider issues such as the cost, efficiency and feasibility of implementing such programmes (Baartman, Bastiaens, Kirschner, & van der Vleuten, 2006).

Combining Formative and Summative Assessments

Methods with high validity and reliability are the primary concern for 'high stakes' summative assessments which provide an overall judgment of competence for qualification or accreditation purposes. In contrast, formative assessments provide on-going corrective feedback in order to promote self-reflection and guide future learning and so require methods which provide in-depth feedback regarding specific aspects of competence. It has therefore been argued that different assessment methods may be required for formative and summative purposes (Epstein, 2007) and results from this study concur. For example, self-assessment was not viewed as robust or accurate enough for summative purposes. Nevertheless, comparing therapists' own ratings of their in-session performance with an assessor's rating provided a useful tool for enhancing therapist's

capacity for self-reflection (Bennett-Levy, 2006; Laireiter & Willutzki, 2003). That being said, participants did not typically employ independent formative and summative assessments, suggesting the need for assessment tools which can serve both functions. This could, for example, be achieved by including further narrative feedback in addition to numerical scoring within the scales used to rate therapist performance within treatment sessions or role-plays. The numerical rating would highlight areas of skill deficit, while more descriptive qualitative feedback would identify exactly where therapists fell short and what they could do to improve. The addition of such formative feedback plays an integral role in the development of competent, reflective practitioners, both within the context of training and on-going continuing professional development (Dochy & Moerkerke, 1997; Milne, 2007; Van der Vleuten et al., 2010), and is typically well received by those being assessed (Govaerts, Vleuten, Schuwirth, & Muijtjens, 2005; Shute, 2008; Van der Vleuten et al., 2010).

Recognising Assessment Consequences

Examining both the intended and unintended consequences of conducting competency assessments has been highlighted as a necessary and important aspect of test evaluation (Linn, Baker, & Dunbar, 1991; Messick, 1994). The results of this study draw attention to a number of unintended negative effects which can occur as a result of conducting CBT competence assessments. For example, the process of being assessed was recognised as an anxiety provoking process, which therefore requires a supportive assessment environment. Teaching therapists how to pass assessments, rather than how to deliver effective CBT, was also raised as a danger which should be avoided. As educational competence frameworks are much broader than the tools used to assess CBT competence (Roth & Pilling, 2008), focusing on the content of assessment tools could have a drastic impact on the breadth and quality of training. Of particular concern were participants' reports that striving to fulfil assessment criteria can sometimes lead to poorer treatment delivery. These issues do not appear to be restricted to the context of assessing CBT competence, as research within wider educational settings shows that assessments have a strong influence on learning and the provision of training, causing both students and instructors to focus on assessment requirements (Alderson & Wall, 1993; Frederiksen, 1984; Prodromou, 1995). It is therefore imperative that those involved in assessing the competence of CBT therapists are alert to and put measures in place to mitigate against such unintended consequences.

Limitations

Conclusions from this study must be interpreted in the light of limitations inherent in the qualitative method used. In line with the principles of IPA, this study offers in-depth insight into the

experiences of a limited group of individuals, rather than seeking to provide a representative sample (Brocki, & Wearden, 2006). However, it is important to recognise that, although participants were drawn from a range of academic, training and routine practice settings and included both men and women, all were from the UK, most were clinical psychologists and many had undertaken CBT training around the 1980's and 1990's. Hence the ability to generalise the findings beyond the current sample is limited. It is also important to recognise that, in accordance with IPA principles, the researchers played a dynamic role in generating and analysing the data. Hence it is possible that another researcher with different personal characteristics and theoretical beliefs would have facilitated a different interview and offered a different interpretation of the data. The retrospective nature of the study also means participants' accounts of their experiences may have been influenced by factors such as the social context of the interview and memory biases. Due to these limitations, caution must therefore be exercised in the interpretation and application of the findings from this study.

Future Directions

The issues raised by participants in the study highlight a number of areas that warrant further investigation. For example, the study raises questions about the impact of patient-related factors (e.g. patient complexity), therapist-related factors (e.g. self-presentation biases), and assessor-related factors (e.g. independence, experience, degree of training) on therapist competence. The findings also reaffirm the need for further development of the CBT competence assessment measures that are currently available. Several potentially useful avenues for exploration were identified, including examining the utility of structured role-play assessments, developing standardised supervisor measures for assessing both knowledge and skill, and further refinement of scales for rating therapists' performance within treatment sessions. In addition to examining reliability and validity, it will be important that any measures which are developed are also feasible to implement, easy to use and serve both formative and summative functions. However, it will also be important to recognise that no single measure of CBT competence will be able to effectively measure all aspects of CBT competence. Thus, tools will need to be developed, selected and implemented according to the context and purpose of the assessment, with a multi-method approach ultimately offering the best way to provide a well-rounded, comprehensive assessment of CBT competence.

References

Alderson, J. C., & Wall, D. (1993). Does Washback Exist? Applied Linguistics, 14(2), 115-129. doi: 10.1093/applin/14.2.115

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & van der Vleuten, C. P. M. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. Studies in Educational Evaluation, 32(2), 153-170. doi: 10.1016/j.stueduc.2006.04.006

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. Professional Psychology-Research and Practice, 38(5), 493-500. doi: 10.1037/0735-7028.38.5.493

Bennett-Levy, J. (2006). Therapist skills: A cognitive model of their acquisition and refinement. *Behavioural and Cognitive Psychotherapy, 34*(1), 57-78. doi: 10.1017/S1352465805002420

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. doi: 10.1017/S1352465801004040

Borders, L. D., & Fong, M. L. (1992). Evaluation of supervisees. *The Clinical Supervisor, 9*(2), 43-51. doi: 10.1300/J001v09n02_05

Brocki, J. M., & Wearden, A. J. (2006). A critical evaluation of the use of interpretative phenomenological analysis (IPA) in health psychology. *Psychology & Health, 21*(1), 87-108. doi: 10.1080/14768320500230185

Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy, 36*(5), 581-587. doi: 10.1017/S1352465808004438

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research, 22*(6), 1-11. doi: 10.1080/10503307.2012.716528

DeVellis, R. F. (2012). *Scale development: Theory and Applications*. Thousand Oask, CA: Sage.

Dobson, K. S., Shaw, B. F., & Vallis, T. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24*(4), 295-300. doi: 10.1111/j.2044-8260.1985.tb00662.x

Dochy, F., & Moerkerke, G. (1997). The present, the past and the future of achievement testing and performance assessment. *International Journal of Education Research, 27*(5), 415-432.

Dudek, N. L., Marks, M. B., & Regehr, G. (2005). Failure to fail: The perspectives of clinical supervisors. *Academic Medicine, 80*(10), 84-87. doi: 10.1097/00001888-200510001-00023

Epstein, R. M. (2007). Assessment in medical education. New England Journal of Medicine, 356(4), 387-396. doi: 10.1056/NEJMra054784.

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49*(6-7), 373-378. doi: 10.1016/j.brat.2011.03.005

Govaerts, M., van der Vleuten, C., Schuwirth, L., & Muijtjens, A. (2007). Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education, 12*(2), 239-260. doi: 10.1007/s10459-006-9043-1

Kaslow, N. J. (2004). Competencies in professional psychology. American Psychologist, 59(8), 774-781. doi: 10.1037/0003-066x.59.8.774

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. London: Sage Publications.

Laireiter, A. R., & Willutzki, U. (2003). Self-reflection and self-practice in training of cognitive behaviour therapy: An overview. *Clinical Psychology and Psychotherapy, 10*(1), 19-30. doi: 10.1002/cpp.348

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21. doi: 10.3102/0013189x020008015

Martino, S., Ball, S., Nich, C., Frankforter, T. L., & Carroll, K. M. (2009). Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research, 19*(2), 181-193. doi: 10.1080/10503300802688460.

Mathieson, F. M., Barnfield, T., & Beaumont, G. (2010). Are we as good as we think we are? Self-assessment versus other forms of assessment of competence in psychotherapy. *The Cognitive Behaviour Therapist, 2*(01), 43-50. doi: 10.1017/S1754470X08000081

Milne, D. (2007). Evaluation of staff development: The essential 'SCOPPE'. *Journal of Mental Health*, 16, 389 – 400. doi: 10.1080/09638230701382818

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*(2), 73-84. doi: 10.1037/a0018121

McManus, F., Rakovshik, S., Kennerley, H., Fennell, M., & Westbrook, D. (2011). An investigation of the accuracy of therapists' self-assessment of cognitive-behaviour therapy skills. *British Journal of Clinical Psychology*. doi: 10.1111/j.2044-8260.2011.02028.x

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance Assessments. Educational Researcher, 23(2), 13-23. doi: 10.3102/0013189x023002013

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

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Prodromou, L. (1995). The backwash effect: From testing to teaching. ELT Journal, 49(1), 13-25. doi: 10.1093/elt/49.1.13

Reid, K., Flowers, P., & larkin, M. (2005). Exploring lived experience. *Psychologist, 18*(1), 20-23.

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

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognititve and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153-189. doi: 10.3102/0034654307313795

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomonological analysis: Theory, method and research*. London: Sage.

Sofaer, S. (1999). Qualitative methods: What are they and why use them? *Health Services Research, 34*(5), 1101-1118. doi: 10.1093/intqhc/14.4.329

Townend, M., Iannetta, L., & Freeston, M. H. (2002). Clinical supervsion in practice: A survey of UK cognitive behavioural psychotherapists accredited by the BABCP. *Behavioural and Cognitive Psychotherapy, 30*(04), 485-500. doi: 10.1017/S1352465802004095

Van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: Building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology, 24*(6), 703-719. doi: 10.1016/j.bpobgyn.2010.04.001

McNaughton, N., Ravitz, P., Wadell, A., & Hodges, B. D. (2008). Psychiatric Education and Simulation: A Review of the Literature (Vol. 53). Ottawa, ON, CANADA: Canadian Psychiatric Association.

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), 620-630. doi: 10.1037/0022-006X.61.4.620

Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *Journal of Nervous and Mental Disease, 199*(4), 276-279. doi: 10.1097/NMD.0b013e3182124617

Appendices

Appendix 1: Ethical Approval

MEDICAL SCIENCES DIVISIONAL OFFICE
LEVEL 3, JOHN RADCLIFFE HOSPITAL
HEADINGTON, OXFORD, OX3 9DU

Tel: +44(0)1865 2289?4 Fax: +44(0)1865 ?50?50
www.medsci.ox.ac.uk

Mrs Kosie Mortimer
Senior Assistant Registrar (Planning and Resource Allocation)



CONFIDENTIAL

Ref: MSD/IDREC/C1/2012/26

Miss Kate Muse
Department of Psychiatry

28 March 2012

Dear Kate

CUREC checklist

I am writing to acknowledge receipt of your CUREC/1 form for your project: **A qualitative analysis of experts' experiences of how CBT competence is conceptualised and assessed.**

On the basis of the information you have provided this has now been approved by the Medical Sciences IDREC **subject to:**

a) **receipt of the hard copy of your application signed by your head of department (Please note that should this not appear or the head of department is unwilling to sign then this approval would not be valid);**

b) **your following the BPS guidelines for online research;**

c) **it is your responsibility to comply with the requirements for administering any tests or questionnaires and if in doubt to contact the publisher of those tests or questionnaires.**

The reference number for this project is **MSD/IDREC/C1/2012/2012/26** and may I remind you that your project may be reviewed at some stage during an annual audit of projects.

Amendments

Should you at some stage alter some of the techniques or procedures then you should first undertake a checklist (CUREC/1) to see whether these changes alter the ethics of the research. If these remain the same then the committee will require notification of the changes to lodge with the project. If they do not remain the same then you may need to complete a CUREC/2 form and undergo further scrutiny by the committee.

Please do not hesitate to contact me if you have any queries about this.

Yours sincerely

Rosie Mortimer

Senior Assistant Registrar

Appendix 2: Information Sheet

UNIVERSITY OF OXFORD: DEPARTMENT OF PSYCHIATRY

PARTICIPANT INFORMATION SHEET

A qualitative analysis of experts' experiences of how CBT competence is conceptualised and assessed.

Principal Researcher: Kate Muse, Supervisor: Dr Freda McManus

You are being invited to take part in a study exploring experts' experiences of assessing CBT competence. In order to decide whether you would like to take part, please read through the following information which will clarify why the study is being conducted and what your involvement would be. If you do decide that you would like to take part then please reply to the initial email you received inviting you to participate indicating this and we will contact you to arrange an interview time.

Why have I been contacted?

You have been invited to take part in the study because you have been identified as an expert in the field of CBT with experience of making judgments regarding the competence of CBT therapists. We would therefore like to ask you about your experiences of the way in which CBT competence is conceptualised and assessed. It is hoped that a number of experts in this field will choose to be involved in the study in order to build an understanding of experts' experiences of assessing CBT competence.

What is the purpose of the study?

Given the current expansion of CBT training and service provision in the UK and the emphasis on dissemination of CBT, the need to reliably measure CBT competence across a range of settings is paramount to the successful expansion and dissemination of CBT. There is however a lack of consensus regarding the definition of CBT competence and the way it should be assessed. Hence the purpose of the present study is to use qualitative methodology to examine expert assessors' experiences of making judgments regarding the competence of CBT therapists. It is hoped that this will provide greater insight into the way in which CBT competence is currently evidenced and differentiated as well as an understanding of how the assessment of competence could be further improved upon. The information gained could therefore potentially be used to enhance current methods of assessing CBT competence or may contribute to the future development of novel assessment methods. The study is being conducted by Kate Muse as a part of a DPhil at the Department of Psychiatry, University of Oxford and is funded by the British Association of Behavioural and Cognitive Psychotherapists (BABCP).

What does participation involve?

If you decide that you would like to take part, we will contact you to arrange a convenient time for an interview to be conducted. We have rooms available at the Department of Psychiatry, University of Oxford and are happy to conduct the interview there. However, in order to reduce any inconvenience caused, we are also happy to conduct the interview at a location which is convenient for you (e.g. at your place of work).

The interview will be carried out by a principal researcher (Kate Muse or Freda McManus) and is expected to last up to one hour. It will involve the researcher asking you a set of open-ended

questions so that you can describe your experiences of assessing the competence of CBT therapists. The interview will need to be audio recorded and will be transcribed verbatim by the researcher at a later date.

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 and can withdraw during the study by letting the researcher know. If this is the case, any data collected from you during the interview 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 interview (e.g. audiotapes) will contain a unique number rather than your actual name. Once the interview has taken place, the interview will be transcribed. Whilst doing this, any personal information related to you will be changed (e.g. names of people or places). Verbatim quotes from transcripts of your interview may be used in the write-up of the study, though we will ensure that this does not include any identifiable information. 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.

That being said, we appreciate you taking time from your busy schedule in order to take part in the study and recognise your valuable contribution to this research. Thus, although your name will not be linked to any results, we would like to thank you for your contribution to the research by including your name within the acknowledgements of any published papers which may occur as a result of this study. We have therefore included a tick box option within the consent form for you to indicate whether you would like to have your name included in the acknowledgements or not. Please note that this decision will not affect your ability to participate in the study.

What are the possible advantages of taking part?

There will not be any direct benefit to you for taking part in the study. However, the information you provide may be beneficial in helping to improve understanding of and methods for assessing CBT competence in the future.

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 interview happens to include any questions which, for whatever reason, you do not wish to answer the interview can be stopped or the question can be omitted.

What if there is a problem?

If there is a problem and you are harmed by participation in the study, you may have grounds for legal action for compensation against the University of Oxford. If you have a concern about any aspect of this study, you can ask to speak to the principal researcher or supervisor who will do their best to answer your questions (contact details on page 3).

Who has reviewed this study?

This study has been reviewed and given favourable opinion by the Research Ethics Committee of the University of Oxford (reference number MSD/IDREC/C1/2012/2012/26).

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:

Kate Muse
Department of Psychiatry
Warneford Hospital
Oxford
OX3 7JX
Tel: 078348 17388
Email: kate.muse@psych.ox.ac.uk

Supervisor:

Dr. Freda McManus
Department of Psychiatry
Warneford Hospital
Oxford
OX3 7JX
Tel: 01865 613145
Email: freda.mcmanus@psych.ox.ac.uk

Appendix 3: Consent Form

UNIVERSITY OF OXFORD

DEPARTMENT OF PSYCHIATRY

WARNEFORD HOSPITAL
OXFORD
OX3 7JX

TEL: (01865) 613 143
EMAIL: kate.muse@psych.ox.ac.uk

CONSENT FORM

A qualitative analysis of experts' experiences of how CBT competence is conceptualised and assessed.

Principal Researcher: Kate Muse
Supervisor: Dr Freda McManus

Please initial box

I confirm that I have read and understand the information sheet dated 27th January 2012 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.	☐
I agree to the interview session being audiotaped and transcribed.	☐
I agree to the use of anonymised quotes being used in the write up of the study and understand that I will not be linked to any responses I give.	☐
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 10 years.	☐
I understand that data collected during the study may be looked at by individuals from the University of Oxford and other regulatory authorities, for the purpose of 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 agree to take part in the above study.	☐

We would like to thank you for your contribution to the research by including your name within the acknowledgements of any published papers. Please indicate whether you would be happy to have your name included in the acknowledgements or not. This will not influence your ability to take part in the study.

☐ Yes, please acknowledge me ☐ No, I would rather remain anonymous

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Researcher	Date	Signature

Appendix 4: Template for Initial Email Invitation to Participate

Dear xxx,

Re: An invitation to participate in a qualitative study examining experts' experiences of how CBT competence is conceptualised and assessed.

We are carrying out a qualitative study examining expert assessors' experiences of how CBT competence is conceptualised and assessed. We have identified you as possibly having a great deal of experience in making judgments regarding the competence of CBT therapists. Hence we are writing to ask whether you would be able to participate in the study.

Taking part would essentially involve spending an hour discussing your experiences and opinions regarding the way in which the competence of CBT therapists' is assessed at a time and location which is convenient for you. I have attached an information sheet which provides further details about why the study is being conducted and what your involvement would be.

Please let us know if you are interested in taking part, or if you have any queries.

Also, if you know of anybody else who might be interested in participating then I would be most greatful if you would either let us know the individual(s) names so that we can contact them directly, or alternatively we are happy for you to forward this email and information sheet to them.

Thank you very much for taking the time to consider taking part,

Best wishes,

Dr Freda McManus & Kate Muse

Appendix 5: Demographic Form

Participant ID _____

Participant Demographics

1. Age: _____
2. Gender: ☐ Male ☐ Female
3. Profession: _____
4. Current job role: _____
5. Do you have any formal CBT training? ☐ No ☐ Yes

   If yes, please specify _____

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

   _____
7. How many years have you been practicing CBT? _____ years
8. What would you estimate your total CBT therapy experience as being? (please circle)

   ☐ < 50 cases ☐ 50 – 200 cases ☐ 200 + cases
9. Are you BABCP accredited:

   As a CBT therapist? ☐ No ☐ Yes

   As a CBT supervisor? ☐ No ☐ Yes

   As a CBT trainer? ☐ No ☐ Yes
10. Do / have you provided CBT supervision? ☐ No ☐ Yes
11. Do / have you provided CBT training? ☐ No ☐ Yes

Appendix 6: An Example of an Analysed Transcript: Participant no. 6

Content[1]	Name	Coding notes	Theme[2]
So first of all in order to, sort of, set the scene. I would like to ask you a bit about what your understanding of the term CBT competence is? What does it meant to be a competent therapist?	Interviewer		
Well I mean to be a competent. Well that's the first question is are we talking about CBT competence in therapy or in other areas? So are we focusing on therapy?	Participant		
We are focusing on therapy delivery yes	Interviewer		
We are focusing on therapy right. Ermm. Well there ermm well what is my understanding of that? Is to use ermmm the style of CBT ermmm thoroughly and heartfully with the patients that you work with. Err. The key features being ermm using errr a sound formulation that is empirically grounded and theoretically grounded. Ermmm and doing so with collaborative empiricism. Ermm. Using guided discovery as one dominant method of conversing with ones patients but not the only one but it will be something err and lots of hypothesis testing, shared hypothesis testing. If I saw that I would probably call is CBT. Ermmm. Then of course there, in a wider sense competency would be thinking about specific therapeutic interventions that might, one might use. For example design and setting up a behavioural experiment and reflecting on the behavioural experiment. These are very specific skills ermm that are informed by the style of CBT and how well they are looked at. And then of course a huge range of different, an ever increasing range of clinical skills one could use. So that would be	Participant	Need to adhere to CBT principles in order to be competent. The formulation is very important- must be formulation driven CBT. Knowledge of theory is important- i.e. must be theoretically driven treatment. Primarily/most important = general CBT techniques/style- collaborative empiricism/guided discovery etc. and secondarily specific interventions e.g. behavioural experiments.	More than just a technician: Responding flexibly in line with CBT theory and research

secondary really. So it's the techniques but first of all comes the whole style and the whole shape of CBT.			
And do you think those are things which as specific to CBT? Or more general therapeutic skills? Or is it a bit of both you are looking for?	Interviewer		
Well there's, there, there is, clearly there is a relationship. The, the developing and the maintaining a bond. And I think that is not specific to CBT I think you can't do anything without that. A lot of the skills from warmth, empathy, positive regard, all of that stuff I think is incredibly important and err in, in, but I have, I don't think it is sufficient to be calling it CBT. I think the tools that you use to do that ermmm are sometimes different from the tools you might use if you were doing any other kind of therapy. So for example the, the explicitness err that, the, the transparency about the process is something I would use for form a part of the relationship that I'm sure my psychodynamic colleagues would not. Errr so there are different ways of doing it. Ermmm so, so the question was... erm....	Participant	You need good general/common therapeutic skills +CBT skills to be a competent CBT therapist.	A solid foundation of general clinical skills
Whether you are looking at general therapeutic skills or, or, sort of, specific CBT skills	Interviewer		
Ok so, so some clearly are, are about, about you know the basic, the basic relationship which I think are general. We are clearly also magpies. There are things that we do and we call is CBT which has, you know, share work that I do in role plays for example we have nicked from somewhere else. I would call it CBT if it's done within the context of formulation driven errmm kind of piece of work that adheres to somewhat the structure, the Socratic nature, the collaborative empiricism that I referred to earlier. And then I would call is CBT. If it was done,	Participant	Need general therapeutic skills.	A solid foundation of general clinical skills

you know, within share work then it's the same, but the stuff all around it would be different.			
Ok so kind of, adhering to the CBT model and skilfully implementing the strategies	Interviewer		
And, yeh, and some of the key feature are inherent in that, in the style as well as the model. But exactly what you do could be you know, from other places. We have clearly grown in the current skills we use. Some of the imagery re-scripting we are doing these days is a far cry from what you might have started out with in 1980 or something.	Participant	CBT has changed a lot over the years.	
So do you think, what, a competent CBT therapist would or should be doing changes as CBT evolves?	Interviewer		
Well I think the range of skills, strategies have expanded enormously and the populations with which we work has made, kind of, what we need to understand, the empirical, the kind of theoretical grounding we need to have in the understanding of the evidence that we need to have has become much broader. What we are trying to conceptualise has become much broader and the environment in which we are trying to use CBT has become much more varied and will have an impact on the specific delivery of it. So, certainly where I work in a kind of very specialised service it is varied in the direction of working with vast amount of co-morbidity, life events, all kind of medical who-ha going on. Physical difficulties and often with people who are not ideal candidates for any kind of therapy because of their locus of control or because of their life circumstances or because of their difficulties in articulating their diff, their problems or indeed in maintaining a relationship. So I think there, we need wider skills in more areas as CBT has developed.	Participant	Need a sound CBT knowledge base. As CBT has developed, the knowledge/skills needed has expanded. Context of service.	More than just a technician: Responding flexibly in line with CBT theory and research

And could you tell me a bit about your experience of assessing competence of other CBT therapists?	Interviewer		
Okay. Well it ranges from the very informal, which is often what supervision is made of, you know, let's have a chat about the case, what is your specific question and then we have a reflection on that, what is happening and what skills you need, so we might do a cartoony role play which might be a few sentences about how you would go about working with something. And that would be my informal way of doing it. Ermmm as well as wider case discussion about conceptual ideas that might be relevant or, or the formulation. So that's, that's one level of competency assessment I would do.	Participant	Don't use the same assessment tools when in less formal settings where the structure is not imposed- use role plays and case discussion in less formal contexts vs. CTS/CTS-R in formal contexts.	Formative vs. summative. Q- or informal vs. formal- same or different issue?
Then going up ermmm I would do specific role-plays. So recently in assessing competence and indeed in training in competence I played the patient and the therapist ermm who wasn't feeling all that competent and indeed isn't thought to be that competent would treat me and we would pause very regularly and each segment of each chunk of therapy, so after the agenda setting, after covering the first bit of the agenda. Reflecting on what happened. I would give feedback and would model it the other way around. All of which is audio taped so that the person can go away and think about that. So that would be, that would be a much more advanced way of teaching competence and assessing competency so we could find where are your weaknesses and where are your strengths.	Participant	Role-plays are used informally as learning/development tools to identify strengths/weaknesses + give feedback. Use role-plays when less conducting formal assessments.	Formative vs. summative. Q- or informal vs. formal- same or different issue?
Ermm listening to, mostly it's audio tapes because it's difficult to video record sessions but I supervise lots of people not just xxx because I do that to but I also supervise a lot of my XXX colleagues across the trust where, ermm, and that's not just in CBT it's also in managing their jobs. But err as part of that job I would, once	Participant	Difference between summative (BABCP accreditation) which is pass/fail and formative which is much more in-depth feedback/a learning	Formative vs. summative. Q- or informal vs. formal-

a year or so listen to a therapy tape that they would choose. So errm.. as part of the BABCP accreditation process but also to give very detailed feedback on, on their skills within that context or that therapy with that patient at that moment in time. So I would do that. Ermmm and mostly that's audio tapes. I would also use snippets and of err live clinical material err in supervision with my clinical trainee as well as in the clinical supervision I would do. Ermmm I would sit in certainly with my clinical trainee I would sit in quite a bit and observe and then give feedback. Ermmm and sometimes err I would see the patients afterwards again with a clinical trainee or somebody I had been training they, often don't have a follow-up with the patient because, say the trainee has moved on, I would then meet with the patient for follow-up and get feedback from them about what's happened and their perception of ermmm the therapist. Ermmm. Do I do anything else? No I don't think so...		tool. Don't use the same assessment tools when in less formal settings where the structure is not imposed- CTS/R vs. sitting in on sessions ('live' observation), asking the patient. Can be useful to ask the patient for feedback.	same or different issue? Role-plays: A promising new avenue or a blind alley? Patient outcome fundamental or meaningless?
Ermm, just following up on a few things that you said there. Ermm you mentioned that you give feedback on recordings of sessions. Ermm how would you go about giving feedback for the sessions, sort of using rating scales and that kind of thing?	Interviewer		
Well there are two ways of doing it clearly and the CTS/CTS-R and I would comment on each of the items of the CTS-R but I would also have an overview comment but particularly when supervising my clinical colleagues here that becomes very difficult to use because very often the reason I am listening to a tape is because it is a very difficult therapeutic environment so the person is not engaged very well, they're dissociating or having major flashbacks in the middle of the thing and, you know, there's all this, you know, there's all this stuff going on, they are not ideal patient where you would expect to tick some of the boxes you will have to change what you do and the CTS-R is not very flexible about, about that I don't think so, sometimes err I would use ermm, keeping some of the ideas	Participant	It is difficult to use the CTS/R when difficulties arise in treatment. Don't use the same assessment tools when in less formal settings where the structure is not imposed- may use item from the CTS/R but may also focus on a specific question/issue the therapist has. Some is possibly more to do with giving supervision than assessing	Formative vs. summative. Q- or informal vs. formal- same or different issue?

in mind of the core style of CBT in my mind I would comment on that and, and on specific questions that the individual therapist would have asked me.. How did I manage, you know, how could I have managed the dissociation differently? Or, you know, did I bring them back in the room in the right way or have I made the best use of that? Or recently ermm somebody had the most chatty patient in the known universe. I mean seriously. Ermm this, you know, what, what I've tried all of the obvious, what other skills can I use to shut him off?! (laughs) So now I would then address myself very specifically to that ermm so it could be the CTS-R bits of the CTS-R or it could be fluid style or directly related to a very specific question that the individual is struggling with.		competence- mixed roles of giving supervision + assessing competence.	
Ahumm. So it sounds like a lot of that is to do with providing, sort of, detailed feedback on strengths and weaknesses and, and enabling progression as opposed to saying "yes" or "no" you are or are not competent in a more formal assessment manner?	Interviewer		
I mean I guess I think clearly for the Diploma course and for some of clinical training I have to yes make a, that kind of 'yes or no' you are competent but it would probably, this is why and this is how you can move that further. Or this is why this is absolutely brilliant...	Participant	Even in more formal contexts it is important to give detailed feedback/points to move forward with.	
So things you could do to improve or build on that	Interviewer		
Yeh.	Participant		
And you said, you mentioned you used snippets of sessions, so would that be, sort of, what, what do you mean by that?	Interviewer		

So it depends on the, on the, all of this also relates to the very specific goals that therapists might have. So if they get clinical training or some of the clinical psychologists who want to learn a specific thing. So I supervise some colleagues who are not that confident in trauma focused intervention. And we would work out some kind of very specific goals. I would like to learn how to do discrimination training. How to interweave working on appraisals with reliving or more specifically and more elegantly. So you'd have very specific goals and err quiet, in some, in some cases very technically focussed. You would take maybe one of those or at most two err and if they bring snippets of the therapy session they would bring snippets that relate directly to that specific goal. So I've now tried to do discrimination training here is the bit of me doing the discrimination training. Or here's the bit where it worked really well and here's the bit where it didn't work so well. Err and we listen to that together and we reflect on it. Ermmm. Yeh and within, within, within the therapy, the supervision hour that we do.	Participant	Again, about providing supervision rather than assessing competence.	
And do you, and when you look at ermm the more broadly at the whole session would you be looking at any particular err stage of therapy that you would be wanting to draw the session from? Or any particular number of sessions that you really feel you would need to be seeing?	Interviewer		
In order to assess competency?	Participant		
Yeh. In order to, sort of, be confident that the person you are either supervising or assessing is, is competent?	Interviewer		
Well I think it's easier to assess the presence of competence then the absence of competence. So if I have got, if somebody gave me a session and it was pretty lousy, I wouldn't think on the basis of that I could say you were not competent. I	Participant	It is not ideal to be drawing inferences about somebodies competence on the basis of isolated treatment	Patient difficulty. Easier to recognise incompetence?

mean it might be that they are not, it could be so lousy that you think "Jesus" (laughs). But it might be that with this patient, in this situation and they had a bad day and they were just clearly not on top form. But on another day with another patient or in another circumstance I would be. However, that probably doesn't apply the other way around. If you have somebody who has a reasonably difficult patient, not, not you know to simple, not completely out there, and they can and they structure the session well, they use guided discovery well, they are focused, they are specific, they get lots of feedback and do homework well I would probably be more confident in saying if they can do it in that session they can probably do it in others too.		sessions. Feels more confident concluding that somebody is competent than somebody is not competent on the basis of a small number of sessions. Influence of patient difficulty.	
Right. Ahum..	Interviewer		
So I would be slightly happier that way around than the other way around. Ermm but err what I would want to see is I would ermm I think ideally in order to sort of competency, complete competency, you would need to have an idea of how they do CBT and use the style of CBT but also how they apply specific strategies and that I suppose in, going back to the definition of competence, are you saying you know how to formulate and use all that style stuff errmm and you can identify you know a negative thought and work with it. Or are you looking for a whole range of specific strategies that you can do well. So I know that Paul Salkovskis is developing a kind of, serious of competences for, I don't know is it OCD I think it is OCD. Ermm. And how to work in various ways. Now you can't really do that if you, if you listen to a therapy sessions of a single treatment because you're not going to have all therapy skills embedded in that. So if the competency is each and every possible skill that you might use as a therapist then that becomes much more difficult doesn't it.	Participant	Difficult to assess all of the possible techniques you could use in CBT- especially when viewing isolated sessions. Need to see a range of clinical skills across patient presentations to assess competence fully.	Gold standard vs. feasibility
Ahmmm	Interviewer		Gold standard vs. feasibility
And that is generally what we would go for.	Participant		Gold standard vs. feasibility

MMM. You would more be looking at, with your actual patients ermm as they come up	Interviewer		Gold standard vs. feasibility
As they come up and as they are appropriate rather than thinking, you know, you now have to treat all anxiety disorders and all depression and you have to be competent for every single aspect of therapy that you might want to use. I don't think we ever do that.	Participant	Looking at: Do therapists know where/when it is appropriate to apply the skills/techniques rather than can they do all of the techniques in isolation. Any disorder or session is only a snapshot of an aspect of therapist's competence.	Gold standard vs. feasibility
Yeh. It would probably be quite a long assessment wouldn't it! (laughs)	Interviewer		Gold standard vs. feasibility
It would be and I think all of us would fail! (laughs) so I think there's kind of core competence skills vs. kind of being competent in everything.	Participant	Boundary of competence? Basics vs. everything!?	Gold standard vs. feasibility
Yeh. Ahmmm.	Interviewer		
But there are, there are traps that, that certainly await you. The traps being that ermm therapists selectively show bits of therapy tape to show competence which inhibits learning I think err or they don't, they, they ermmm don't use live discussion and retreat into case discussions of the safe cases. And don't bring to supervision in whatever form the things where they are really struggling either interpersonally or in terms of a specific skill	Participant	Therapist self-selection bias inhibits learning and anxiety about showing session recordings often stops them from showing this.	Therapist anxiety and self-presentation biases
Because of the fear of, of being failed, or being seen as not being competent?	Interviewer		Therapist anxiety and self-presentation biases

Oh being seen in a negative perspective yeh. Ermm. And indeed because it feels slightly uncomfortable doesn't it to say well that was incompetent (laughs)	Participant	Fear of being seen negatively encourages self-presentation bias.	Therapist anxiety and self-presentation biases
Absolutely.	Interviewer		
Sometimes because they are worried about, if we are actually formally assessing them, about being seen negatively or even as a supervisor if you aren't being formally assessed, just thinking, you know, what she thinks of me. You know.	Participant	Fear of being seen negatively encourages self-presentation bias.	Therapist anxiety and self-presentation biases
Yeh. Yeh. Absolutely. And ermm. You also mentioned about the patient feedback. So ermm this would be ermm the therapist has done the whole treatment sessions, the whole course of treatment, and then you would be seeing them for a follow-up session is that right?	Interviewer		
That would be one way that happens. And that would be mostly with clinical trainees where I do very often the follow-up because they are not around. And it is also quite nice to see what's come out at the end just to double check. So that would be a comment on how did that go, how did you get on, what was useful, ermm, which is not , you know, the, it's not saying well what did you think of the therapist, it's not as blunt as that, although clearly it has an element of that. Ermm. Sometimes that also happens by way of the review of the therapist who is sort of treating the patient so that we might get together on key decision points. Do you go from a stabilising phase to a trauma focused phased would be very often where I might come into the therapist as the known supervisor and have a conversation with the patient which then also gives feedback halfway through.	Participant	It is useful to get feedback from the patient about how therapy went/their perception of therapist's competence- especially in less formal contexts.	Formative vs. summative. Q- or informal vs. formal- same or different issue?
And that would be with the therapist in the room as well?	Interviewer		

With the therapist, that would be with the therapist in the room yeh.	Participant		
MMM. That's interesting. And so, presumably you would be doing that in a relatively informal way. So asking the patient what seems relevant. You are not going in with a sort of check list	Interviewer		
No absolutely not, yeh.	Participant		
Yeh.	Interviewer		
It's about, yeh it's about, it's about the patients perception of how well they feel their needs have been met by a therapist and the strategies that the therapist and the patient have used and what else needs to be done and how comfortable they feel and how good the therapeutic alliance is.	Participant		
Ahmmm. Yeh. Yeh. And do you every look at, sort of, the patient outcomes as a, as a measure of competence? So, sort of, how well the patient has progressed or err questionnaire outcome measures, that kind of thing?	Interviewer		Patient outcome fundamental or meaningless?
Yes I do but I wouldn't think that in itself and particularly in the more straightforward cases it wouldn't be a sign of competence. I've seen people who've done incredibly well with not very good therapists.	Participant	Patient outcome is a useful indicator of competence but is not necessarily accurate and must therefore be used in combination with other indicators.	Patient outcome fundamental or meaningless?
Right, yes.	Interviewer		Patient outcome fundamental or meaningless?

Because they happen to have a lot of internal resources and you point them roughly in the right direction even though lots of the therapy wasn't that fabulous and they still got better.	Participant	Patient outcomes could improve when the therapist is not competent	Patient outcome fundamental or meaningless?
Ahmm	Interviewer		Patient outcome fundamental or meaningless?
And nor does the absence of progress necessarily an indicator of therapist incompetence. Particularly if you are working with very significant co-morbidities and life adversity often you don't get much movement even if you do your best. So I, I don't think, I think it's in a, I think what we need is multiple indicators really.	Participant	Equally, patient outcome could show no improvement- even though the therapist is competently delivering therapy. No one method is ideal so we need multiple methods of assessing of competence.	Patient outcome fundamental or meaningless? Gold standard vs. feasibility
So would, what would your, kind of gold standard approach be to assessing competence, which indicators would you want to be using? All of the things that you have mentioned?	Interviewer		
Well I mean there are the bounds of practicality. I mean if you talk about gold standard and how you would you really assess it I think it would be a mixture of live observation by a recording or some sort or ermmm and case discussion because we need to think about the case formulation, whether or how well they are theory grounded and, and look at their reasoning which is not necessarily absolutely apparent in there, in the live, in the, in the live session that we would have. Get patient feedback and look at patient outcome ermmm. I think all of those, a combination of all of those together would be the best way of assessing it	Participant	Practicalities/limited resources mean it is not possible to do what I would ideally like to do (i.e. to use multiple different indicators/ combination of methods + sources).	Gold standard vs. feasibility
Yeh. You mentioned practicalities maybe being a hindrance. What kind of things would get in the way of, sort of, applying that gold standard approach?	Interviewer		

Well it takes an enormous amount of time and energy and effort. And time energy and effort is not readily available. Ermm. And indeed ermm getting patients to do it, to give feedback and to do various things can also be quite difficult. It can be unethical to bully them into it. But it is just time and money that get in the way of doing it really.	Participant	Practicalities mean it is not possible to do what I would ideally like to do (i.e. to use multiple different indicators).	Gold standard vs. feasibility
Yeh.	Interviewer		
So I think ermmm the kind of combination of case discussion with role-play, with a bit of live assessment and looking at patient outcome data, that, that probably is the most practical way of doing it.	Participant	Multi-method approach to assessment.	
Ahmm. Yeh. And would you want to be looking at the kind of, case formulation and the discussion ermm alongside rating sessions- ermm so where you looking at the same, the same patient ermm or would you want to be, sort of, doing that separately?	Interviewer		
Does that make sense?	Interviewer		
Well I would want to see, particularly in developing competence, I would want to see err treatment plans and ermmm based on the formulations and the, kind of treatment targets and patients goals. And I would want to see that for pretty much every patient.	Participant	Use of treatment plans + written formulations in assessing goals.	
Right. Yes.	Interviewer		

Ermm. I would not , wouldn't necessarily expect myself to listen to every tape of every session for all patients. I would wanted to have listened to a few or snippets of a few in order to form an opinion really. Or specific skills if we are, if the contract in terms of the competency is to learn skill x.	Participant	Need some samples of observation to form an opinion.	Gold standard vs. feasibility
Ahmmm. Okay. It sounds like there's, it's quite important to have an indication of what the therapist can do actually in real clinical practice?	Interviewer		
I've met, I don't know whether you've found this, I've met people who can, who have a wonderful declarative knowledge. They can talk beautifully and persuasively. You sit them in a room and they are not that good. There are other people who struggle ermmm with the declarative kind of, stuff, and are procedurally really good. And I mean, just I mean surely you will have noticed this yourself and my colleagues and I comment on this frequently, is to rate therapy tapes is just so painful. And they often talk about this because I actually sitting in the therapy room with the therapist watching them doing it is not that awful but rating tapes seriously kills me and kills most of my colleagues! (laughs) I think partly it can just get really dull kind of thing, but I think actually I think the real problem is because the skill required to assess competency is enormous because you have to take procedural skill and then translate it into declarative knowledge so that you can describe what the person is doing in a way that you can put into very clear words and then use the words to guide what becomes another procedural skill.	Participant	Difference between talking about vs. doing therapy (i.e. procedural and declarative knowledge)- rating means moving back and forth between the two. Rating sessions is very effortful / difficult because it involves a lot of skill- very aversive process.	Gold standard vs. feasibility. Expert or novice assessors?
Yeh. It's tough isn't it- yeh.	Interviewer		
Actually it's incredibly, it's incredibly exhausting. And it's this constant translating, it's almost from one language to the other.	Participant		Expert or novice assessors?

It's really, really quiet, technically quiet difficult to do and, yep it's just challenging. And I think that's what we don't like. (laughs). It's even why, why do I like what I've just heard? It can be a very simple snippet and it will take quite a lot of thinking. Now. I can then say I like that because XXX, then next time to make it even better then XX and then you have to translate it back into procedural it's just hard.	Participant		Expert or novice assessors?
Do rating scales help with that?	Interviewer		Expert or novice assessors?
Well they help in being, they help you to jump straight from procedural to a rating is easier than to describe why you think you liked it or why you didn't think you liked it. So I can say now that was a 5. Err for identifying negative thought without having to put it into words very much.	Participant	Making a rating on a scale is easier than giving in depth feedback.	Expert or novice assessors?
Do you think you miss something then?	Interviewer		Expert or novice assessors?
Oh god you miss all of the feedback that is involved around that. So there is clearly if you have a pass/fail system at the end of the day that would be fine. But if you are trying to teach skills you can't do that.	Participant	Making a rating on a scale is easier than giving in depth feedback.	Expert or novice assessors?
Yes. It's the developmental feedback that's useful	Interviewer		Expert or novice assessors?
Yeh it doesn't actually advance skill ermm at all. Well yeh. So if I, I listened to a tape and simply had to give, you know, 1-2-3 you know ratings and had to make no other comment on it that would be, I could do that really easily. Much easier.	Participant	Making a rating on a scale is easier than giving in depth feedback- which prevents skills development.	Expert or novice assessors?
But less helpful for the therapist you are rating.	Interviewer		

Yes. And it doesn't require that constant translation from the procedural to the declarative and back again but I think that's what makes expert therapists. I've seen ...	Participant		
..The ability to do that...	Interviewer		
The ability to do that. And I think, I think ermm Tracy Eells (?) has done a lot of work on therapeutic relationships and formulation actually. Did a wonderful study looking at novice therapists formulating, experienced therapists formulating, and expert therapists formulating. And what he found is that the novices and the experienced therapists did no better than each other. And one of, one of his arguments was, and I think it's a very valid one is that, you know, if you just do and then don't think about doing, you can't very easily refine your skill. It's a very highly complex skill. Because you can only reflect on it if you put it into words. Now if you were riding a bike or a horse or whatever then that would be different but it if it's a higher order skill I think you need to be able to say- the reason I didn't like that bit of the way the session went is because x y and z and that's where the problem was. Because then you can start thinking "well how can I do that differently?". But if you don't think about it you can't do it. And I think experts have a lot more cycles of procedure, thinking, reflecting, you know, the learning cycle.	Participant	Expert therapists are more able to self-reflect.	
Yeh. Mmmm. Do you think that self-rating is a useful way of assessing one's own competence ermm in that respect?	Interviewer		
I think it is really useful to help the supervisor understand their supervisee's perspective but it does not necessarily accurately reflect the capacity that the person has. I think ideally they should think about what it is that they liked about	Participant	self-assessment is useful learning tool but has limitations in that it is often	Formative vs. summative.

what they did and why or why so and so. Just giving a 2 I don't think does very much. Except you know if somebody actually thinks I didn't really set an agenda that was rubbish, you know as a jolt to remind you that actually you should be doing it. That's fine, that's the way it works. Other than that I don't think it teaches very much.		not accurate- so used more for formative than summative asst.	Therapist anxiety and self-presentation biases
Ahmmmm.	Interviewer		
If I look at the ratings my students would have given themselves, they usually do as part of having an assessment and some of it just reflects confidence. There are some people who would score themselves really quite highly and I think "what are you on?" And other people will score themselves very low and there seems to be characteristic, there are the self-critics and then there are the "oh that's alright". We have more self-critics than anything else.	Participant	Self-assessment is not accurate- more reflective of self-confidence than competence. Confidence colours self-assessment in both directions.	Formative vs. summative. Therapist anxiety and self-presentation biases
Yeh. So kind of, a useful supervision tool but not really useful in terms of determining the competence of, of the therapist per se?	Interviewer		
Yeh. It's too prone to bias.	Participant		
Is there anything else which you, ermm, have experienced or tried or which you think "mmm no that doesn't really work very well, that doesn't really give me a clear indication of the, the competence of the therapist"?	Interviewer		
Well I think case discussions alone don't. Ermm I think ermmm. I would say I, I've become increasingly sceptical about what a person says they do and what they actually do. Ermm. But if you get them to, to play it out in the room then that becomes immediately a lot better I think. Errr. But no I don't think so.	Participant	Case discussions alone are not enough because of self-presentation bias. Observation of therapy/role play is much better than discussion.	Therapist anxiety and self-presentation biases.

			Gold standard vs. feasibility.
Okay. Ermm. And...who do you think is really best placed to assess competence? Ermm who do you think is the best person for the job so to speak?	Interviewer		
Ermm that is a really difficult question. I mean clearly it is something that most people are doing. There is a certain degree of skill is ermm important. But I'm slightly nervous of expertise so ermm I'm perfectly well aware that I'm toughest at assessing competence in the area of my expertise.	Participant	Assessing competence is a skill. But assessors are tougher in their own area of expertise.	Expert or novice assessors?
Ahmm.	Interviewer		
You know, so I am, I love, I like OCD but I mean I'm no expert in OCD so the way I would assess competency there would be much more benign than if somebody comes along and has a fairly complicated multiple trauma case. I would give them a hell of a time. Now I try not to (laughs) but it's hard, it's an effort. So I do, I do sometimes also wonder about a certain level of competency that, that that is kind of matched to what you are trying to teach. I think if you are at a very high level of competence and you are teaching a novice that gap can be really difficult to bridge.	Participant	But assessors are tougher in their own area of expertise.	Expert or novice assessors?
Mmmm. Yeh.	Interviewer		
But if you are very, very good at training you can do it, but you have to be very good at training.	Participant		
So would you have different standards that you would be wanting people at different levels of training and experience to be achieving?	Interviewer		

Yeh.	Participant		
So different guidelines of what you... well different things you would be looking for or different level of ability of the same things?	Interviewer		
Well I would be, I think it's both actually. There's the core competence that you would want to get very early on in CBT ermm. Working collaboratively, collaborative empiricism essentially. Guided discovery, err formulating and the centrality and specificity of appraisal. I would have thought that might be, and off the top of my head, I would think that those are some of the core things that you would want somebody to get, to find at a level a minimum level of competency. But then you would want to both up that skill but also broaden that skill in a range of disorders with a different levels of types of presentations and in greater complexity in terms of how the patient presents and the challenges they face.	Participant	Difference between novice vs. experienced- same basic skill set but for experts = wider disorders, presentation and increased complexity.	
And do you think that there is anything which impacts on how easy it is to appear competent. So I'm thinking you mentioned complexity and that kind of thing?	Interviewer		
Oh yeh I mean if you have a really straightforward patient you try the first and obvious strategy it may work. And you move on. And when you are listening to that therapy session it will sound like a competent therapy session. The, the therapist wasn't really tested very much. Whereas if you have a moderately complicated patient you might, you know, the strategy one and two may not work, you may have to try three or maybe even four err and to do that with some evidence and to keep going you know rather than battling. And that becomes a much higher level skill. And of course then you get really difficult people ermm where that becomes harder still but you may, anybody no matter how competent they are, they could do whatever they like, really competently and they still aren't going to get anywhere. So it's that, it's that kind of, in the CTS-R at the moment	Participant	It's easier to appear competent with less complex patients.	Patient presentation

has in some sense has a response item inherent in them. It's how the patient responds is what defines how good it was to a certain extent. Ermm. So then it becomes very difficult. So I think you need to have at least an opportunity to exhibit getting stuck and becoming unstuck.			
Right mmm. And do you think there is anything else which impact on, sort of, how easy it is to appear competent? Or to rate competency?	Interviewer		
Well if you just use case discussions there's your capacity to bull***t! Ermm having a very simple patient makes it easier. If you stay I suppose within a narrow range of disorders or presentations.	Participant	It's easier to appear competent with less complex patients. Self-presentation bias is evidence in case discussions.	Therapist anxiety and self-presentation biases.
mmm okay.	Interviewer		
So if you only ever treat a particular problem and nothing else that could, that could come across as being competent, very competent but may actually reflect general clinical competency	Participant	Need to view competent behaviour with a range of disorders- limited domain competence vs. broader competence.	Gold standard vs. feasibility.
So would you want to be seeing a variety of different patients with different disorders?	Interviewer		
Well at least if that's, if that's the training model yeh.	Participant		
Otherwise you could say oh well you are a competent PTSD therapist but not more broadly competent?	Interviewer		
Yes. You can't treat psychosis for love nor money you know, absolutely yeh.	Participant		

Ahmm. And do you think, I mean obviously you are taking the disorder into account when you are assessing competence, do you think it's a very different process err assessing competence with different disorders? Do you think there needs to be different ways of assessing it or are there broad things which are more transdiagnostic?	Interviewer		The challenge of diversity within the CBT family
Oh well I think the vast majority is transdiagnostic	Participant		The challenge of diversity within the CBT family
Okay	Interviewer		The challenge of diversity within the CBT family
Ermmm. But there are some things that become very specific. So flashbacks being, multi-sensory flashbacks in session in PTSD and particularly in severe PTSD that is when you will see it and you will see it almost nowhere else. And if you have somebody in the room is has got, who is in the middle of a flashback, all of a sudden bang (clicks fingers) you, if you're working with PTSD you've really got to get competent at dealing with that. But you won't have any chance to learn how to do that unless you are in the room with that kind of presentation. Errr. Or working err with somebody who is hearing voices. You could only do that in that presentation. So I think the vast majority, there is huge overlap in what we do but there are some specifics. Ermm.	Participant	Most aspects of competence are transdiagnostic but there are disorder/symptom/protocol-specific aspects of competence too.	The challenge of diversity within the CBT family
Now the other thing I am wondering about is whether we assess and indeed should assess I mean we probably shouldn't is inhibitory beliefs that the therapist has that interfere with competence with the competent use of something. I'm thinking, have you read Glen Waller's paper on therapist drift for example?	Participant	beliefs therapists hold can inhibit competence	

I don't know. No I don't think I have actually	Interviewer		
Well it's a very good kick up the backside! BRAT 2010 where he, he goes through a whole series of things which make therapists drift and become sub-optimally useful to their patients. Anything from ermm the working environment and having to see 8 patients a day five days a week, that, that probably would do it! (laughs) But he then moves on in this kind of increasingly scorching critique of us therapists and just thinking about some other things which make us work sub optimally of which avoidance of key therapeutic strategies that are quiet hot being one of them.	Participant	Therapists' competence tends to drift over time.	
You know behavioural experiments, exposure, re-living. You know the stuff that in the moment can cause significant distress lots of people back off from for a number of reasons and when you identify some of the core beliefs the therapists have you tend to fear the competent use of particular therapeutic strategies because, you know, as soon as it gets hot, you act to cool it down. Not necessarily because you don't have the skill to keep it hot but because you think you shouldn't. And that becomes more, so I'm wondering, and I'm not sure we actually should assess that, but I'm wondering whether we should ever look at therapists' attitudes and beliefs about ermmm avoiding therapeutic strategies.	Participant	Therapists' competence tends to drift over time- possibly due to therapists beliefs.	
It sounds like that would be quite helpful in terms of personal development err as a therapist in developing your practice would be addressing those kinds of issues	Interviewer		
Yes. And there's of course the other beliefs that err that we tend to carry around which are yeh well the evidence base might say but ought to be disregarded because.. so there's a whole set of belief that we have or sometimes stumble across in supervision which means that people are not utilising their capacity to be competent	Participant		

And I suppose in order for those beliefs to sort of come out and be looked at then you would need to be doing the assessment within a supervisory capacity where you knew the therapist	Interviewer		
Yeh	Participant		
Ermm then you could discuss that	Interviewer		
Yes you could. I'm not sure you could actually put it in writing...	Participant		
It's a bit like trying to get into the US, you know "have you ever been convicted of..." (laughs). Whoever ticks the box? Yeh.	Participant		
Yeh absolutely! Well.. I am trying to think if there is anything that we haven't covered which I was hoping to ask about...ermm.. I think that is most of the things... actually one thing I was going to ask you is, you mentioned about the different, err slightly different, slight differences between different disorders err and some skills being specific only to that disorder. Would you be wanting to use the same kind of rating scale across the board or would you be wanting to use different kind of err rating scales depending on the disorder?	Interviewer		
Well it depends whether you are strategy specific or not. Errrm. Whether there is kind of you know, here are 50,000 strategies we can use- how competent are you on all of them? That could be disorder independent actually. Discrimination training or bringing somebody back form a flashback or talking about voices you know, that kind of thing, where you do that. Whereas at the moment the CTS-R doesn't do that. It, it focuses on, on, on the specificity of appraisal ermmm and on behavioural techniques. Errr it doesn't do and it has kind of a general how well	Participant	Being a competent therapist is more than being competent at specific techniques.	More than just a technician Responding flexibly in line with CBT theory and research.

were therapeutic strategies utilised but it doesn't, it doesn't do anything more than that, it doesn't think about the process of thinking for example and how we manage the process of thinking. Ermm and I would have thought Paul and to some extent David Clark would say that is missing from our competence assessments. And certainly it would be used by in, in physical health in surgery for example where you have to demonstrate competency on a particular procedure which and I'm not saying, clearly you can have people who are very good at procedures but can't hold the whole together. Ermm and you mustn't lose sight of that.			
So then you go back to that idea of multi-assessment, well multiple methods needed (participant nods).	Interviewer		
Yeh. Mmm. That is really interesting. Is there anything which you think is err, sort of, important in this area or any experiences you have had which you think are relevant which we haven't touched on?	Interviewer		
I don't think so.	Participant		
Thank you/ goodbye etc.	Interviewer		

[1] Highlighted text has been included within a theme/category.

[2] The theme which the text was assigned to is indicated in the far right column headed "theme". Please note that the referenced theme is that which was coded at time of transcription. The themes have therefore changed in subsequent analysis and write up.

Appendix 7: An Example of a Theme and Associated Interview Extracts

More than just a technician (17 sources)

Participant 1
"I assess it on their talking about CBT- whether they sound like they know what they are doing whether their ideas about what CBT is fit with mine. Ermm and if I do observe their practice I assess it on things like are they able to structure sessions and I assess, I put quiet a heavy weighting on their conceptual integration. Ermm do they seem to understand why they are doing what they are doing? So do they seem to be able to formulate this person's problem in a cognitive-behavioural model and then use that to decide which CBT interventions to carry out? Erm. So sort of knowledge about CBT practice from what they talk about. Conceptual understanding of CBT and being able to use that to decide what to do."

"And then I think also interpersonal skills. I probably assess it on because anyone can do CBT anyone can say well you know your problem is your belief that you're inadequate and let me tell you that you're not. Or you know, now list some evidence that you are inadequate now list some evidence that you are not ok re-rate your original belief. It's not really gunna get you anywhere. So what you need is a level of interpersonal skill to get the patient to engage with the CBT tasks."

"I think I also I'm influenced by their ability to skill up the patient to do CBT. 'Cuz if you can teach someone else to do it you can do it yourself. You know you can't really teach someone else to do it without doing it yourself. And CBT is not something that is done to the patient. It is essentially my model of it is that your job is to train the patient to be able to carry out CBT interventions for themselves."

"One of the difficulties with them, as far as separating out basic, clinical, therapeutic, clinical therapy skills and CBT skills. 'Cuz you never really know quite what you're assessing. 'Cuz you can't do good CBT without basic good clinical skills. But a lot of them seem to be non CBT, not really that specific to CBT, there's only a few of the items that are specific to CBT really."

Participant 2
"You can adhere to a protocol and just do it very badly. And, when people sound like a computer programme and awful words "and what was going through your mind right now?" because they, they, they're regurgitating what they have been taught but they're not thinking. So I think a competent therapist is thoughtful therapist and can understand the principles of what they are doing is adhering to the protocol but is able to use it flexibly."

"I'm looking for understanding of the principles and being able to adapt them to the client in front of you. So I'm looking for that flexibility to be able to help the client within your protocol."

"When I'm not using the cognitive therapy rating scale? Well I think it is very difficult (laughs). Ermmm. I sort of, I ask myself the question "do they understand what they are trying to, does the therapist understand what they are trying to do"? Are they able to, to do it in a way that I think is ethical and appropriate? And are they able to personalise it."

"So do they seem to show understanding of the principles and carry out the intervention in a way that is ethical and personalised with the patient?"

"And appropriate for the stage of therapy."

“Or are they doing something that I don't, if I don't, I suppose if I don't know why they are doing something that's a problem for me. So my, kind of general rule of thumb is, if somebody stopped a tape I would like to think that I've asked a question for a reason. If I stop a tape in supervision and I say "why did you ask that question" and the therapist goes "I don't know" I don't think they are doing therapy competently ‘cuz there directionless, they don't know what they're doing.”

“I think having a rationale for it, and the rationale either being relating to the protocol or relating to something that's gunna help effect change. You know or, they have to have, if somebody doesn't have a reason for opening their mouths I've got a problem with it.”

“So I mean all of the other things on the cognitive therapy scale, like, you know interpersonal effectiveness, and empathy and therapeutic alliance. I mean I think those are all important. So I think those, those are sort of the generic kind of competences. But I think, so I think I, the absence of those is very notable. In that, I don't think the person is delivering therapy competently if, for example, somebody says something like, I can just see the specific example of saying "every time I go to pick my kids up, I think they won't be there and something terrible will have happened to them". And the therapist said something else, I can't remember exactly what, you know "what was the, what else were you thinking" or something like that but he didn't say "my god, how awful for you". I think, so they'd lost their humanity. So I think when you, when the therapist has lost their humanity, then they are not doing any therapy competently. So that's the general part of it. The specifics of it. I think you have to operationalize it and you have to, and it will be, might be different for each disorder. Or each problem that you are trying to deal with. So treating mental contamination competently is gunna be different from treating depression competently.”

Participant 3

“I guess it's actually a mixture of general therapeutic competence and specific CBT knowledge and ability to apply techniques to apply techniques appropriately.”

“But I would actually be concerned about- do they understand what the CBT is for the, that's relevant to the client they're working with and are they doing something that would recognisably be CBT for that problem with that person.”

“Ermm. Yes, so it would be a mixture of discussion and seeing what they are or hearing what they are doing.”

“Well CBT competence or therapeutic competence has to relate to the client and what works for the client and that's different for different clients it's not just that we should be looking at what effects outcomes for clients but we should also be recognising that clients themselves are different from each other, and what helps one person doesn't help another.”

“Ohhh. Ermmm. I guess it's actually a mixture of general therapeutic competence and specific CBT knowledge and ability to apply techniques to apply techniques appropriately.”

“And the reason I say that is for instance in the CTS or CTS-R or any of the other instruments, there will be items that are calling on general therapeutic skills like interpersonal effectiveness, ermm. And in fact in the original CTS there was both an empathy rating and an interpersonal rating and err a professional rating. And the research suggests that unless you have general therapeutic abilities, you won't be, you know you are less likely to be effective however good your CBT techniques might be, so it's a combination.

“Errrm. Within the general therapeutic ones, ermm, well there's, you know ability to create a safe and trusting relationship so the client can say talk and, and, you know be honest. Ermm. And that calls upon, erm , you know warmth and genuineness and errr, the Rogerian variables.

“And one of the disadvantages of the CTS-R I think is that it's reduced the number of boxes that relate to the therapy relationship relative to the what the CTS had, whereas the therapy relationship actually is very important, so the, the weighting has changed. And they did that because the items correlated, the therapy relationship items correlated. But that doesn't necessarily mean that you want to reduce the weighting right down.”

Participant 4
I also see it as an art. Because for competency I am looking for people who are using the techniques appropriately.

“My understanding is that competence reflects being aware of the technical components of CBT. Ermm embedded in a background of understanding the theory that dictates the best use of those techniques. Ermm and that's going to vary from disorder to disorder as well as being a generic thing. So I see it as ermm a science, i.e. it's got some technical elements.”

“Errmm. Well I'd want to know that, let's check the things that I'd want to know, I'd want to know that someone’s got the technical skills, but that's not enough. I'd want to know that they are aware of the theory, that's going to inform the use of the technical stuff, but that's not enough. I want, ermm, evidence that they can marry theory with real life clinical practice, but that's not enough. Because I was to see that they're doing that within a quality therapeutic alliance that is characterised by collaborative empiricism.”

Participant 5
“Knowledge and skills. Skills growing out of an understanding of ermm theory and principle. So it’s not just a bunch of techniques in other words. We are talking here about a coherent errmm understanding of the nature of difficulties. And coherent treatment programmes directly related to that.”

“Ermm. You, set, well, in terms of assessing competence. Ermm. Theoretical knowledge and knowledge of clinical principles is demonstrated, should be demonstrated through what the person does with a particular patient.”

Interviewer- “Right, in the session?”

“Yeh. Ermmm in addition to that you can assess. Ermm. Both of those things through for example essays and written case presentations. What you don't, what you get then is whether they actually know in theory what's what and how to do it. But you don't, get, you know, a real accurate indication of whether they can actually do it in practice. And to me that is the essence of competence. It's not just knowing the theory it's actually knowing how to use it.”

“Ermmm. And I think practical training on the job or IAPT type stuff. It's probably still valuable to have written case presentations but I don't know that you need essays really. If you have case presentations which require people to summarise relevant theory and research in principle then that's enough I think. Ermm or you can do it by having a session recording with an accompanying commentary which explains that sort of contextual information.”

“But to me being able to apply something effectively means you have to know the context. In terms of theory, research and principle.”

"..(pause).. Ermm. What do I think? I think that ermm we need to be aware of, I think, you know, for me when it comes to assessing competence babies and bathwater comes to mind. Ditto with training. That we know there are evidence-based protocols which are very powerful. And I know that from, you know, using them as much as anything. Ermm. We also know that in routine clinical practice life is more complicated than that. Ermmm there's ermm the protocol itself and then with these very specific and ermm means of dealing with maintenance factors for example that are a part of that. But there is also your capacity to be a good therapist beyond that ermm and I think that there is a risk that in homing in on ermm very tightly focused techniques ermm we may lose things which are actually quite important. For example the capacity to generalise to other problem areas. The capacity to deal with co-morbidity, complexity ermmm different populations, different settings, all that kind of thing. So I don't think, I'm not saying either or, it seems to me both and, you know, and if you are, if you're using an evidence-based protocol with a patient for who it is actually 100% spot on then that is what needs to be assessed. But if you're going to go out in the real world and operate in a community mental health team somewhere I think you need more than that. So I have a concern arising from the current focus on tightly defined models and protocols. And that's absolutely not to underplay their value, 'cuz I think they're terrific but it's just to say, they're great but you know what, there may be more to life than that."

"You need something more specific probably. Or some elements need to be specific. Not all in my view. I think that it's still helpful to assess generic competences like the capacity to relate to patients for example."

"I think it is important that relevant elements are. So in other words if you want to see whether the person is doing good CBT for that particular problem area as defined through a particular established protocol- yes. But actually I think that doesn't cover everything. So, ermm, the model specific bits, it doesn't, don't cover session structure, ermmm, overall strategy, formulation skills, therapeutic relationship skills."

Participant 6

"The, the developing and the maintaining a bond. And I think that is not specific to CBT I think you can't do anything without that. A lot of the skills from warmth, empathy, positive regard, all of that stuff I think is incredibly important and err in, in, but I have, I don't think it is sufficient to be calling it CBT."

"Ok so, so some clearly are, are about, about you know the basic, the basic relationship which I think are general."

"Then of course there, in a wider sense competency would be thinking about specific therapeutic interventions that might, one might use. For example design and setting up a behavioural experiment and reflecting on the behavioural experiment. These are very specific skills ermm that are informed by the style of CBT and how well they are looked at. And then of course a huge range of different, an ever increasing range of clinical skills one could use. So that would be secondary really. So it's the techniques but first of all comes the whole style and the whole shape of CBT."

"Well I think the range of skills, strategies have expanded enormously and the populations with which we work has made, kind of, what we need to understand, the empirical, the kind of theoretical grounding we need to have in the understanding of the evidence that we need to have has become much broader."

"Well it depends whether you are strategy specific or not. Errrm. Whether there is kind of you know, here are 50,000 strategies we can use- how competent are you on all of them? That could

be disorder independent actually. Discrimination training or bringing somebody back form a flashback or talking about voices you know, that kind of thing, where you do that. Whereas at the moment the CTS-R doesn't do that. It it focuses on, on, on the specificity of appraisal ermmm and on behavioural techniques. Errr it doesn't do and it has kind of a general how well were therapeutic strategies utilised but it doesn't, it doesn't do anything more than that, it doesn't think about the process of thinking for example and how we manage the process of thinking. Ermm and I would have thought Paul and to some extent David Clark would say that is missing from our competence assessments. And certainly it would be used by in, in physical health in surgery for example where you have to demonstrate competency on a particular procedure which and I'm not saying, clearly you can have people who are very good at procedures but can't hold the whole together. Ermm and you mustn't lose sight of that."

Participant 7

"You know I think one of the first things one is looking for is nothing to do with CBT specifically but do people have the ability to form a therapeutic, helpful relationship with other human beings kind of thing? You know can they be warm and empathic and supportive and all of those kind of things is certainly one of the first things. Certainly that if it is not there makes you think eek you know."

Participant 8

"And then I guess in other settings it's been less formal than that because perhaps there, I think of the xxx service for example or other similar situations like that there's been some kind of training delivered and then followed-up by regular supervision so you're much more reliant upon what people report in supervision about their practice. Ermm but I guess there, you're thinking more about often the way which people discuss their cases, you know evidences that they are able to formulate problems using the CBT model, that their discussion and thinking about the case and the way they are working with the person is informed by clear CBT model and that the kinds of things they are doing and planning to do fit with that, they make sense within the model and, and within the formulation for the client.

"I think usually its transdiagnostic things that we're looking for. Ermm. Because people on the whole that aren't I don't know maybe it's just people I supervise but they tend not to be applying CBT in a very protocol driven fashion for particular presenting problems. So there'll be adaptation on what they do according to the kind of problem they're working with but it's mainly those transdiagnostic things that we're concerned about. And I would think would be the most important things in assessing competence really because it allows, if you've got those building blocks in place it allows for flexibility and when you meet a complex case or you meet somebody who doesn't fit in a particular box you've still got your back to basics core skills that you can apply to try to tackle the problem."

Participant 9

"If, if you don't have the interpersonal skills ermmm you can't do therapy. And if you can't do therapy you can't do a particular breed or brand of therapy. Ermm. So I wouldn't, I would ermmm differentiate somebody who's a technician, who performs certain tasks err and somebody who can actually perform the full round of therapeutic tasks."

Interviewer- "Could you tell me a bit more what you mean by that, perhaps give me an example of what a technician versus a competent therapist would be?"

"Ermmmm yeh. Ermmm. I'll give you a errr errr a for instance in particular to the field of CBT, actually just yesterday we were, we were listening to err somebody errmm doing a session of CBT

for a depressed patient. Errrmm. Most of the boxes got ticked. Ermm. The, there were Socratic questions, there was an agenda, there was homework, ermm. So a whole range of, of boxed got ticked. Err. The degree to which the therapist was truly attentive to the patient was questionable ermmm. The degree to which the therapist truly empathised with the position of the patient was questionable. So although they performed a number of technical behaviours which were appropriate to that context. The degree to which they were actually therapeutic behaviours in that context was called into question."

"Yes and no, yes and no. Ermm. Obviously one would have a different set of expectations as to the experienced practitioner going beyond competence as, err, a level of performance. ermm so again if you think of the, the Dreyfus scale you are looking for proficiency or expertise then that is above and beyond competence. At the same time it's very easy for experienced practitioners to claim expertise when they are actually departing from adherence to err say err normative standards. Err so, where I'd question some of the err of what errr Dreyfus seems to be saying about proficiency and expertise which can be read as saying you're actually ditching protocol and adherence and doing something supremely beyond that errrm you're, I think the expert is still in touch with all of the underpinning principles of competence and is applying them to a higher standard and is adapting but always staying in touch with theory and evidence.

Well, some of us, ermm, who view the ermmm, we draw the parallel with musicians. Ermm. Particularly improvisational musicians. And the great Charley Park (?) said you learn all of the changes then forget them. Ermm I don't know if you play an instrument?

Interviewer- I don't unfortunately! (laughs)

"If you don't know what the scales are you can't improvise"

"And you are still improvising on and around the scale or if you have gone off scale you know you've gone off scale and you know how to get back on. Ermm. So you are not, you're not ditching the rudiments of music, you're incorporating all of the principles and rudiments of music but you have moved on from simply reading the score. Errr, the, the, the novice and the advance beginner is constantly reading from the score but isn't interpreting the score, isn't applying the principles and rudiments of music to change due to circumstances.

"There's, there's a, there's a flip side to that as well as something that err myself and my colleagues XXX and XXX talk about and ermmm, it's one of the great Jazz pianists errr whose name escapes me at the moment, ermm, who, one of the famous pieces he played was tea for two, a famous jazz piece. He played this variation on tea for two and he got to a point where people came to the ermm to another concert and instead of there being fantastic new variations he was still playing the same one. And they said "well how come you aren't playing new variations" and he said "well that's as good as I can get" having got that good, he stuck to that. And so there's, ermm, there's, I think this is interesting it's not that people who've gone beyond the, the competence must then inevitably be doing things which look weird and wonderful which is different from somebody who is a trainee- there'll be certain circumstances in which they say well doing it this way is the way which works, therefore I will replicate that practice ermm, you know, you don't want your brain surgeon to have thought of a very clever new variation of opening your head. There are certain ways of cutting your head open and you want them to stick to that practice. Landing a plane generally speaking is the same behaviour."

"Well I think because therapy is an interaction your constantly in the business of balancing a set of procedures against the requirements of an individual. There's ermm that's, that's a difficult task. And even treating one individual. If all you assessed was across the course of therapy with one

patient the state of the patient would change and you can't really claim to be competent in practice if all you know to do is to repeat a particular behaviour, then a particular behaviour, then a particular behaviour."

"Well, we, well because of the national guidelines of IAPT we would want something that comes under the banner of change. So that ermm, their, the, using the CTS-R there's an item which assess change methods. Now ermm going back to the earlier point about the overly procedural novice, we, we're seeing too many where the novice who's failed, I don't know why I did the inverted commas thing (laughs) and it's not even true because they've just failed you know (laughs) ermm is nonplussed because they say look I did a thought record ala Christine Padesky, I did all seven columns and we got the thought. Right yes but that's clearly not where that session was going, that wan an inappropriate use, your, your patient was already learning what they needed to learn in terms of reviewing the homework, getting more learning out of it through your dialogue and increasing their motivation for their next homework. So the key thing here is have they learnt something new which is going to alter their affective state or their beliefs or their motivation? Ermm so we're much more interested in ermm what the therapist has done to promote patient learning than whether or not they have used a particular technique, like the thought record, or, ermm, a err pie chart, err so on. Does that, does that answer your question?

"What might be argued is that by simply seeing that they can do a set of procedures you don't know that they understand the underpinning principles. You are inferring that they understand them, the place that they show that they know the underlying principles is in err essay and a case study or a dissertation."

Participant 10
"But then also beyond that whether they are using particular techniques as well. So a combination of both. Ermm and not only just using them but using them err appropriately as well. So I'm thinking I suppose a mixture of adherence and competence as going together."

"Yes I think it would include a judgement of whether they are doing it right. Because I think it's easy to adhere to a manual by doing a particular technique but using it at the wrong moment or the wrong time."

"So there's a whole range of techniques that you can select from. There's two which I think are really the drivers which we want and then there's other ones which you can bring in as appropriate depending on the person."

"Guess I'm thinking about the obviously the basics about are they engaging the person. Ermm. And treating them well and listening to them. Ermm. So there's that basic ermm building a therapeutic relationship but then also beyond that whether they are using particular techniques as well."

"I mean certainly some of the basics is what you want. You do want an ability to allow somebody to engage in therapy and for them to talk and you know, to treat them well in therapy, certainly in psychosis there's a long history of people not being listed to so, you know, I think that's very important whatever level that's something, whatever they should be able to do that. So it's always about that. Having the chance for the patient to see what they are thinking and feeling and being listened to and the therapist to at least do that."

Participant 11
"And, and the therapist very much trying to take a stand that really, really accommodates the clients perspective or world view."

"Is guided by it and doesn't in any way attempt to impose ermmm a sort of arbitrary or pre-formed set of ideas onto the patient."

"I actually do think that there's an aspect of CBT that often does get dropped in favour of it being seen as a series of techniques and that is this sense of a fundamental philosophy. So it's not so I mean Steve has this nice word which ermm I think is fairly obvious in terms of what it means is err this idea of instantiation you know as in basically how he puts something into practice so from his point of view the critical competence is actually helping patients to spot what's going on in their own heads to start to be able to interrogate their own cognitions. Ermm how you do it of course could be done in a variety of ways. The tragedy is ermm and this comes back to I suppose another angle on competence more in terms of sort of a sense of incompetence you know the tragedy is that people have started to see a lot of CBT as a series of activities do a thought record errmmm err do this do that so it's all very over technique bound and people lose sight of the fact that errmm well look why are you doing this, you know, what's your aim in doing this. Ermmm. Errr and and, and at that point I think CBT can become problematic because as were all too aware ermmm you know it, it can be caricatured err in lots of ways and at that point you know when, when it does look like you're just implementing something but not really knowing why or you're doing it for the sake of doing it and it can become a bit pointless and that's when I think patients also start to, you lose your alliance with the patient.

"And I'm also quite struck that you know certainly when I've been talking quite a bit to more skilled CBT therapists and again this may be a philosophical difference among CBT therapists but you know it's the first years tend to be very technique orientated there's no, they, they find it hard to stay with in a broader sense of a generic therapeutic skills and seeing people as people and they're seeing them located in their interpersonal world and having thoughts and feelings and being able to tell a story about the person really you know it becomes much more a matter of what you can do with them. Whereas I. Which is a bad thing I think."

"The other thing by the way just to add in ermmm is ermmmm that the other the at the at one side of our map are our generic therapeutic competences and they're in our conceptualisation. I mean they're generic because they seem to work quite well through most other modalities of therapy, so psychodynamic, systemic, humanistic all have the same generic underpinning competences. But there are some critical bits of it, well they're all critical in a way, but there's certainly that sort of sense of being able to develop and maintain a good therapeutic alliance and engage the client and deal with emotions. We do actually see it as pretty critical too and, you know, there's good research evidence which shoes actually that if you try using, to implement CBT in the context of a poor alliance you're not going to do very well. Errmmm. And, and that's the other little bit of neglected competence potentially, not for anyone with expertise but if you were looking into where people sometimes go wrong it's when they sort of plough on, they have a sense of what they'd like to do but they leave the patient behind."

Participant 12

"I do think there's something about Jazz though. What I mean by that is that it's about the way in which certain sets of skills need to come together for certain situations. So a particular client so a client who is very erratic, who's relationships are all over the place, who pushes boundaries, for a client like that it would be structure and the empathy may be particularly important. For a client who is very OCPD that structure might just actually feed in and exacerbate the OCPD. Ermmm so I think there is something about Jazz and how things come together. And I'm not sure how CTS-R ratings really take that into account. You know you've got, clients need different things at different times. Ermmm and even that isn't absolute. You know there are lots of different ways to get to the top of the mountain. Ermmm. But there are probably some ways which won't get you to the top of the mountain. You know a client who is very all over the place with their boundaries and their

relationships and is up and down and some might use terminology like borderline. You don't wanna be buying into that stuff if you want to get anywhere. Because that won't get you to the top of the mountain. So there probably are some ways of things not to do."

"So it's interesting isn't it, it's like a metaphor that's just come to mind is errmm you know that shoal of fish that you sometimes, you know you see a shoal of fish and if you think about it every fish in there is like an aspect of competence. But it kinda needs to come apart and come into a different shape for different situations and different clients. And competency is actually the ability to do that. It's not like you treat every client, every presentation, every session exactly the same."

"But also some metacompetences about being able to build a relationship and ermm errrrmmm progress a therapy from start through to finish."

Participant 13
"Now some of those things that help people change are protocol-specific and some are probably generic. Like being engaging. Ermm. You know, if you can't engage and retain clients then that's, that sort of seems to be necessary for the treatment to work."

"I think there's a sort of generic training you need for holding sessions, beginning, stopping sessions. How to deal with crisis. What to do if somebody doesn't turn up. How to keep somebody on topic if you like. There are some, quite a lot of generic skills. But I wouldn't see it, some of which are CBT specific but some of which are pretty common to any short term psychological treatment."

Participant 14
"Well I suppose I I'd want them to have a thorough understanding of the cognitive-behavioural model behind a particular treatment that they were offering. Ermmm. And ermmm that they would fully understand all of the procedures that are involved in the treatment and how they should be deployed"

"Ermm. And I think also that they are able to modify their own behaviour ermmm to take into account the particular presentation of the patient, not only as an individual but also in terms of the disorder that they're presenting with. So you really have to behave systematically differently when you are treating somebody with say social anxiety than somebody with PTSD which happen to be the two things that we do on a daily basis and I would not consider somebody competent unless they knew how to adjust their behaviour for those two disorders, which are very different. And the transference issues involved are different. And so you know, that would apply to lots of other disorders. So I'm thinking of borderline obviously there is a lot of knowledge about the condition and the way it presents in therapy in terms of the interpersonal interactions with the therapist. So the person should have mastered as well as the particular CBT techniques."

"At the same time that they would have considerable sort of interpersonal skills ermmm so they are able to pick up on ermmm err any, well firstly so they are able to convey all the key ideas in a language which makes sense to the patient that they can pick up on ermmm any ermmm err any moments when err the patient may become unduly distressed or doesn't seem to be following what's happening."

"I mean there are obviously some ermm general skills, which I don't know if everyone would agree with"

"Ermmm and I think if you're doing cognitive therapy there is a particular skill which you simply have to have or you don't get anywhere which is your ability to get inside somebody else's head

and to see the world form the perspective of their beliefs. And there are quite a lot of good CBT therapists who are more behaviourally driven say, maybe behavioural activation would be a good example of a current one for which you don't necessarily need to have that skill to do it well. But you simply can't do cognitive therapy well without that and some people can't do it."

"To see whether they can essentially decentre, get out of their head and into somebody else's head. And ermm some conditions are very good test cases for it. So PTSD is a very good one. Because ermm when people describe their own traumas, they are always horrible and so errmm most therapists you will immediately have an emotional reaction to the story and will also have a quite good idea of why they think it's horrible. But a good cognitive therapist won't assume in any way that the reason that the patient finds the events horrible is what they've just noticed from their own emotional reaction. And so they will, you know, try to ignore that and probe really you know, what it is that so bad about it for this particular patient and won't do lots of reassurance about oh of course I understand how terrible it is anyone would feel that way because ermm that's sort of a pretty serious misunderstanding of cognitive therapy which doesn't assume that anyone feels that way they have very individual reactions."

"So general interpersonal skills. Getting on well with people."

"I'm just thinking that there are some disorders where, you know, I think you do have to. It's a bit, I think a good therapist is a bit like a chameleon, you know, depending on what environment they're in or what disorder they've got in front of them, they adjust their behaviour ermm and its and it's not really to do with what they say it's to do with the way they behave."

Participant 15

"And this is the high order principles I think it's not just the application of technique but it's the application of technique in a skilful manner based on a knowledge and understanding of the objectives of the intervention itself."

"How do you hold an overall view of it and how do you make decisions about when to intervene and in precisely what way within the general framework err how much to stick very closely to err a particular manual approach and sometimes when you need to vary that slightly."

"Err. Our framework typically will also have disorder specific knowledge. You know something about the nature of social anxiety disorder or depression ermmm and ermmm err you know there's something about the specific treatment and rationale for that treatment in relation to say depression so the difference between say a CBT based approach for depression or a behavioural activation approach, there is as you know a different underpinning theoretical idea about the cause of depression in those two models and an understanding of that I think it's pretty important for the effective delivery of them. So we try and start with a broad conceptualisation or a broad understanding rather of a problem and then it narrows down but at the same time you've already got the sorry if you like the metacompetences, the overview so you are exercising judgement you are developing if you like therapist or clinicians and not just technicians"

"Something from what we know generally about what we know about the effective delivery of psychological treatments and this goes beyond CBT but is err there needs to be an errr effective relationship that err provides the basis on which a err any treatment is delivered. I think it's hard to do without that"

"But certainly I think the view that Tony and I would take is unlike some people that the alliance that the relationship is not sufficient. It's not the primary driver of change. Ermm

“But if you look at the overall framework you would think that some are just necessary that without an alliance, without a collaboration err without a structure it would be difficult to see how you could effectively do CBT without a therapist being quite active at that point it would be difficult to see.”

“Quite a bit of that err I think is you know issues about building an alliance, maintaining that kind of secure frame is going to be quite important and that’s probably common to many treatment approaches, whether it’s DBT or metallisation based therapy or schema focused CBT I suspect that some of that will be common

Participant 17
“Well I think the thing that people struggle with actually is and will continue to struggle with is, is actually what’s in the framework down as like metacompetences. Yeh. Which is it’s all very well knowing, kinda like actually knowing what to do but when things don’t go sort of to plan, go kind of awry you need to be able to be flexible. And developing flexibility as a therapist err and being able to adapt to things is something that's is always an on-going kind of learning experience”

“But actually, what am I for when I’m actually either observing somebody or listening to their tape of anything like that is firstly can they like, you know, form a relationship with somebody. Errr. Are they able to kind of have rapport ermm are they able to do sort of the general therapeutic stuff, because if they can't do that then they've got lots of trouble. And, and then err. I guess err and there's lots of aspects of that, like listening carefully and summarising back and err accurate reflection and empathy of their emotional experience and that kind of thing.”

Participant 19
“A bit of both really. I mean a bit of both I mean I suppose on the course on the xxx on the xxx we do it more formally I suppose but I do this in other supervision is actually when people are presenting a problem and we’re identifying disorders I’ll ask the person what model they’re using and I’ll ask what reading they’ve done why they chose that approach vs. another approach and that’s much more first of all obviously I’m, I’m checking out whether they have an evidence-based approach, whether they’re referring to literature that’s up to date and whether they’re aware of what’s out there. Ermmm but I’m mostly just interested in why people would choose one above another and you know I think sometimes the fact that people don’t even think about why they’re using a particular approach is part of what’s interesting. Ermm. So yeh I, I, I and the challenge I always feel for all therapists is the challenge is how they then link that with the individual sitting in front of them. So yeh.”

Appendix 8: Invitation to Provide Respondent Feedback.

Dear xx,

Re: A qualitative exploration of experts' understandings and experiences of conceptualising and assessing competence in cognitive behavioural therapy

You will recall that last year you kindly gave your valuable time to participate in a qualitative research study exploring experts' experiences of assessing CBT competence, which was conducted as part of my DPhil research under the supervision of Freda McManus. We would like to take the opportunity to thank you very much for taking part in this study as, without your valuable contribution, this research would not have been possible.

We would like to give you the opportunity to see the results of the study. As well forming a chapter within my DPhil thesis, an abridged version of the study has been written up for publication. Please find attached a draft version of this manuscript for your information. As you stated that you would like to remain anonymous, we have not included your name in the acknowledgments of the paper /OR/ As you stated that you would be happy for us to recognise your contribution to the research, we have including your name within the acknowledgements of the paper. We hope to submit the paper for publication in two weeks' time (14th June 2013). However, we wanted to give you the opportunity to review and reflect on the manuscript prior to submission. By requesting participant feedback, we hope to enhance the accuracy, credibility and validity of our research findings. Thus, we would welcome any comments, observations, criticisms or questions you may have.

Thank you again for your participation,

Kind regards,

Kate Muse and Freda McManus

CHAPTER THREE:

The Development of a Novel Scale for Assessing Competence in the Delivery of Cognitive Behavioural Therapy: The Assessment of Core CBT Skills (ACCS)

Chapter Overview

It is essential for training providers, clinical services and researchers to have access to reliable, valid and usable methods of assessing the competence of therapists delivering Cognitive Behavioural Therapy (CBT). However, the measures for assessing CBT competence that are currently available have been extensively criticised. Hence there is an urgent need for improved measures for assessing the competence with which CBT is delivered, particularly those which assess therapists' clinical skills. This thesis aims to address this issue by developing a novel rating scale (the Assessment of Core CBT Skills: ACCS). The ACCS provides an assessment framework for assessors to provide formative and summative feedback regarding therapists' performance within observed CBT treatment sessions and for therapists to rate and reflect on their own performance. The central aim of the ACCS is to build upon existing rating scales (particularly the Cognitive Therapy Scale- Revised) in order to provide a tool with improved validity, reliability and usability. Development of the ACCS scale followed a classical measurement theory approach involving three key stages: (1) scale development, (2) expert review, and (3) scale evaluation. This chapter focuses on the initial development of the scale (stage one) and outlines the four key steps that were involved in developing the ACCS: step one- a definition of the scale's scope (i.e. a description of the construct the ACCS aims to measure), step two- the creation of an item pool (i.e. the development of theory-driven scale items), step three- the creation of a response scale, and step four- the creation of a manual (i.e. developing rating guidelines).

Contents

Introduction 155

Step 1: Define Scale Scope 158

1.1. Identify current scale problems. 158

1.1. Define target construct. 162

Step 2: Create an Item Pool 163

Domain 1 - agenda setting. 166

Domain 2 - effective use of time. 166

Domain 3 - homework. 167

Domain 4 - delivering CBT interventions. 168

Domain 5 - measuring change. 169

Domain 6 - conceptualisation. 170

Domain 7 - managing therapeutic relationship. 171

Domain 8 - effective two-way communication. 172

Step 3: Create a Response Scale 173

3.1. Select an appropriate response format. 173

3.2. Identify optimum response categories. 173

3.3. Scale anchors 174

3.4. Item scoring 176

3.5. Feedback form. 177

Step 4: Create a Manual 177

Discussion 178

References 180

Appendix 1. Feedback Form and Scale Manual (version 1) 189

Introduction

It is essential for training providers, clinical services and researchers to have access to reliable, valid and usable methods of assessing the competence of therapists delivering Cognitive Behavioural Therapy (CBT) (McHugh & Barlow, 2010; Rakovshik & McManus, 2010; Weck, Bohn, Ginzburg, & Ulrich, 2011). This is especially important given the current drive to expand CBT training and increase the dissemination of CBT into routine clinical practice settings (D. M. Clark, 2011; Karlin et al., 2012; McHugh & Barlow, 2010). As outlined in Chapters one and two, the measures for assessing CBT competence that are currently available have been extensively criticised, there is an urgent need for improved measures (Barber, Sharpless, Klostermann, & McCarthy, 2007; Fairburn & Cooper, 2011; Kazantzis, 2003; McGlinchey & Dobson, 2003; Muse & McManus, 2013; Rakovshik & McManus, 2010; Sharpless & Barber, 2009; Waltz, Addis, Koerner, & Jacobson, 1993). Although clinical skill is widely viewed as being at the heart of delivering competent CBT (see Chapter two), it is also the most complex aspect of CBT competence to assess. It is, therefore, perhaps not surprising that limited progress has been made in developing valid, reliable and usable measures for assessing therapists' CBT skills. Despite the complexity involved in operationalising and assessing clinical skill, it is important to avoid oversimplification of CBT assessments by focusing on aspects of competence which are more easily measured (i.e. therapists' knowledge of CBT). Thus, the current research will focus on developing an improved measure for assessing whether therapists can demonstrate the skills necessary to effectively deliver CBT.

Chapter one identified six methods which can be used to assess CBT skills. However, many have significant limitations: patient outcome provides a limited and confounded picture, patient surveys have limited validity, and feasibility and self-assessments are often inaccurate. Given the major flaws in using these methods to assess CBT skills they are not examined further within this thesis. This leaves three methods of assessing CBT skills which warrant further investigation: role-plays, assessments based on supervision sessions, and assessor-rated treatment sessions. Assessments based on supervision sessions which include direct observation of treatment sessions (e.g. by viewing session recordings) may provide a useful means of assessing CBT skills, especially as this method is already frequently used within training settings and routine clinical practice (Chapter two). Research in this domain could usefully focus on improving and further evaluating the scales currently used by supervisors to rate competence. However, although it is considered best practice to view session material within supervision, in reality the use of direct observational techniques is rare, with supervisors often relying on case discussions and therapists' second-hand reports of what happened in CBT sessions rather than

direct observation of therapists' performance (Townend, Iannetta, & Freeston, 2002, Chapter two). Given that, in practice, supervisory assessments are likely to be limited in their ability to provide accurate assessments of therapists' CBT skills, the current research will not focus on the further development of assessments based on supervision sessions.

Both role-plays and assessor-rated treatment sessions provide assessments of clinical skill based on direct observation of therapists' performance and thus warrant further investigation. Whilst these methods both have their own inherent advantages and disadvantages, the current thesis will focus on developing improved measures which can be used by assessors to rate therapists' performance within treatment sessions. This decision was made for two key reasons. First, there is a danger that using role-plays may lead to the simplification of clinical scenarios and create a less authentic subjective experience, thereby reducing the degree to which the assessment is representative of clinical practice (Sharpless & Barber, 2009). The qualitative interview outlined in Chapter two identified this as a significant concern amongst experts in the field. This concern therefore requires addressing before this method can be used in practice. Indeed it is essential that the measures used to assess therapists' CBT skills provide an accurate indication of their ability to competently deliver CBT to real patients in real clinical settings. This issue is circumvented by using assessor-rated treatment sessions because the assessment is based on therapists' performance within real treatment sessions. Second, it is important to focus on developing a measure that could be feasibly implemented in a range of training and routine practice settings. The development and implementation of role-plays is time consuming, expensive and resource intensive (Kaslow et al., 2009), therefore undermining the feasibility of this assessment method. Whilst there are also time and resource implications involved in viewing and rating recordings of therapy sessions, these are unlikely to be as substantial as those involved in using role-plays. Furthermore, rating treatment sessions is a method which is already widely accepted and used in practice (Chapter two), meaning that an improved rating scale could be quickly and easily implemented within a range of clinical and training contexts.

When assessing therapists' performance within treatment sessions, assessors typically rate the skill with which CBT is delivered using a standardised rating scale. Rating scales outline and behaviourally operationalise the skills involved in the competent delivery of CBT and are therefore viewed by assessors as providing a helpful framework for rating therapists' CBT skills (Chapter two). Chapter one identified a number of scales that can be used to rate the skill with which CBT is delivered. These scale can be divided into those which are transdiagnostic (i.e. focus on competences which are not specific to any one diagnosis) and those which are disorder-specific (i.e. focus on competences required to deliver disorder-specific treatment protocols).

Disorder-specific scales are highly relevant for use within controlled research settings where patients are selected on the basis of diagnosis. However, they are less appropriate within wider clinical practice, particularly training settings in which clinicians learn how to deliver a variety of CBT protocols or clinical practices where therapists work with patients experiencing a wide range of mental health problems (Barber et al., 2007; Chapter two). The use of disorder-specific scales in these contexts would make it difficult to draw comparisons across therapists and would require assessors to be familiar with and possibly even trained in how to use a range of different scales, thus undermining the feasibility of this method of assessment. It was, therefore, decided to focus on the development of a transdiagnostic measure of CBT competence which would be applicable across a range of treatment and training settings.

Chapter one identified three transdiagnostic rating scales. The Cognitive Therapy Adherence and Competence Scale (CTACS: Barber, Liese, & Abrams, 2003) is problematic because the competence rating is confounded by the inclusion of an 'overall competence' item and the amalgamation of quality and appropriateness ratings, which appear to measure the same construct. More significantly, the CTAS was designed to assess competence in the provision of CBT for cocaine abuse and the utility of the scale to assess CBT more broadly remains questionable. The Cognitive Therapy Scale (CTS, or Cognitive Therapy Rating Scale: CTRS, www.beckinstitute.org) and the Cognitive Therapy Scale-Revised: CTS-R, Blackburn et al.'s (2001) are more widely used to assess therapists' competence (Chapter two).

The CTS has been criticised for being overly focused on cognitive therapy in the context of depression, inadequate definition of scale points, requiring a high level of assessor inference, item overlap, multiple concepts being addressed by single items, neglecting key competences, and difficulties attaining adequate inter-rater reliability (Chapters one and two). The concurrent use of multiple different versions of the CTS is also problematic. As outlined in Chapter one, the CTS-R sought to overcome some of the limitations of the CTS and does seem to provide an improvement. However, the CTS-R has also been criticised for limited usability due to an unclear instruction manual and rating scale, failing to take into account key contextual variables (e.g. stage of therapy, severity of patients' problems), an under emphasis of basic therapeutic skills, limited sensitivity, ambiguity and lack of behavioural specificity in scale item descriptors, a high level of assessor inference, poor inter-rater reliability and a lack of opportunity to provide detailed formative feedback (Chapters one and two). The theory, evidence-base and practice of CBT has also developed and evolved since the CTS and CTS-R were initially developed.

In sum, further refinement and improvement upon current CBT competence scales is necessary. Hence this thesis sought to build upon existing rating scales (particularly the CTS and

CTS-R) in order to provide a tool with improved validity, reliability and usability. Specifically, this thesis focuses on the development of a novel CBT competence rating scale (the Assessment of Core CBT Skills: ACCS) which can be used to provide formative and summative feedback regarding therapists' performance within observed treatment sessions.

Psychometrics is a methodological approach that had been widely adopted for developing scales measuring a range of psychological and social phenomena (DeVellis, 2012). Item response theory (IRT) and classical test (or 'measurement') theories (CTT) are the most dominant psychometric approaches. CTT methods are based on the assumption that an observed score is the result of a respondent's true score in addition to a single error term, whilst IRT methods differentiate between different types of error, particularly error relating to individual item characteristics. Thus, CTT is more concerned with a collection of items that are combined into a composite score, whereas IRT focuses on individual items. IRT is a far more complex approach to scale development that CTT, involves a much higher degree of expert judgement and is a relatively novel approach that is still in the active stage of development. As IRT aims to develop items that are insensitive to other respondent characteristics, it also required repeated testing of large and heterogeneous samples. Thus, in most cases, CTT is considered the favourable approach (the exceptions being the development of scales with items that are inherently hierarchical [i.e. where items are not parallel indicators of the underlying latent variable, as is assumed within CTT] and scales which are concerned with differential item functioning [i.e. where items are expected to perform differently across groups]). Thus, development of the ACCS follows the CTT approach outlined by DeVellis (2012). The approach involves three key stages: (1) scale development, (2) expert review, and (3) scale evaluation (DeVellis, 2012). The current chapter focuses on the initial development of the scale (stage one), whilst the expert review (stage two) is outlined in Chapter four and scale evaluation (stage three) is outlined in Chapters five and six. The initial process of scale development can be further broken down into four steps: 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, and step four- create a manual (i.e. develop rating guidelines). Each of these steps is outlined in more detail below.

Step 1: Define Scale Scope

1.1. Identify current scale problems.

The first step in developing a new measure is to outline the problems with existing rating scales, accompanying manuals and scale implementation in order to highlight potential pitfalls and problems that we would aim to overcome in the development of a novel scale. This was

achieved through reviewing findings from the literature review (Chapter one) and revisiting the interview transcripts obtained in the qualitative study (Chapter two). As outlined in Chapters one and two, at present the most widely used transdiagnostic measures of CBT competence are the CTS or CTS-R, hence the criticisms of current scales particularly focus on these scales. Problems raised with the CTS/CTS-R fall within three broad categories: scale items, response scale, and rating manual. The specific problems within these three categories and ways in which the ACCS aims to overcome these are outlined in Table one. Overall the difficulties with current problems indicate that the ACCS will need to ensure increased capacity for formative feedback, improve definitional clarity, provide clearer and more detailed rating guidelines, reduce item overlap, reduce the number of concepts addressed by a single item, account for recent advancements in CBT, ensure applicability across axis one disorders and allow applicability for a range of cognitive and behaviourally focused therapies.

Table 1: Criticisms of current rating scales (particularly the Cognitive Therapy Scale [CTS] and Revised Cognitive Therapy Scale [CTS-R] and ways the ACCS aims to overcome these difficulties.

Problem area	Problem description	Ways the ACCS aims to overcome these difficulties
1. Scale Items	• Failure to assess how well people use outcome monitoring (e.g. selecting appropriate outcome measures and using them to inform formulations / treatment decisions).	• Addition of an item assessing the use of outcome monitoring
	• Underrepresentation of basic therapeutic skills (e.g. therapeutic alliance / relationship, listening skills etc.).	• Greater inclusion of basic therapeutic skills within the scale items
	• Overly focused on CBT within the context of depression.	• Assessment of core skills necessary when delivering a range of CBT protocols to patients with a range of different problems (though still focusing largely on Axis 1 disorders).
	• Overly focused on cognitive elements of CBT at the detriment of behavioural elements of CBT (e.g. behavioural interventions).	• Incorporation of behavioral aspects of CBT interventions.
	• Multiple constructs assessed within individual items.	• Improved separation and demarcation of items.
	• Overlap between items.	• Also improved separation and demarcation of items.
	• Inclusion of items that are unlikely to be demonstrated within every session.	• Only include core skills which are likely to be demonstrated in every session.
2. Response scale	• Use of ambiguous / subjective anchor points on the response scales (e.g. idea of excellence and degrees of excellence).	• Greater specificity / more clearly defined anchor descriptions.
	• Lack of operationally defined / behaviorally specific, item-specific anchors / descriptors.	• Greater behavioural specificity in anchor descriptions. Anchor descriptions given for each item individually.
	• Inadequate definition of scale points.	• More detailed definitions for different scale points.

2. Response scale	• Confusing instructions (e.g. separate illustrative descriptions for items but instructions say that assessors are not meant to base rating on the illustrative descriptions). • Poor inter-rater reliability due to individual interpretation/use of scale. • Response scale takes limited account of key contextual issues (e.g. stage of therapy, patient difficulty etc.). • Lacks the opportunity to provide in-depth formative feedback (e.g. space to highlight key strengths/weaknesses, space for identifying how 'problem' areas can be improved upon to aid skill development / progression).	• Clearer instructions about how to use the scale in the user manual. • As above (clearer demarcation of items, behaviorally specific item descriptors, and clearer instruction manual). • Incorporate contextual issues within the item descriptors. • Provide more opportunity to formative feedback.
3. Rating manual	A large amount of time consuming training is required before the manual can be used. • Lack of detailed implementation protocol outlining exactly what the scale is designed to measure, how to rate a therapist's performance using the scale, what assessment materials ratings should be based upon and who should complete the ratings. • No guidelines are given as to what contextual information should be taken into account (e.g. stage of therapy, rationale / reasons why therapist chose certain strategies, formulation, patient outcome data, treatment plans based on formulations and treatment targets / patients goals, materials completed in session or for homework).	• Clearer instruction manual / scale which requires little or no training to use. • Clearer instructions outlining the purpose of the scale and the way it should be used to assess competence. • Provide guidance as to which contextual information needs to be viewed in order to use the ACCS.

1.1. Define target construct.

It is essential that scale development begins with clearly articulating the scope and specificity of the construct the scale aims to measure (L. A. Clark & Watson, 1995; Cronbach & Meehl, 1955; Loewenthal, 2001). Hence the following broad definition of what aspects of CBT competence the scale aimed to measure was constructed:

The scale assesses core general therapeutic and CBT-specific skills required to appropriately deliver individual cognitive-behavioural therapy to adults experiencing mental health problems.

This definition can be broken down into a number of constituent parts. First, the scale is designed to assess 'limited-domain intervention competence': that is, the general therapeutic and treatment-specific knowledge and skills required to appropriately deliver CBT interventions that reflect the current evidence base for treatment of the patient's presenting problem (Barber et al., 2007; Kaslow, 2004). Thus broader professional skills (e.g. ethical practice, effective use of supervision) are beyond the scope of the scale. Second, the scale is designed to assess therapists' performance within observed treatment sessions. As outlined in Chapter one, this method of assessment falls within level four on Miller's (1990) hierarchical assessment framework and as such assesses whether a therapist can use their CBT skills in routine clinical practice. Thus, the scale is not designed to assess therapists' skills within artificial clinical environments (level three), which can be assessed using role-plays. Similarly, the scale does not examine therapists' practical understanding of CBT (level two) or their CBT knowledge (level one), which are better assessed using academic assignments such as multiple choice questionnaires, essays, case reports or clinical vignettes.

Third, it was decided that the scale would focus on core CBT competences that are transdiagnostic (i.e. not specific to any one diagnosis) so that it could be used to assess competence across different disorder-specific treatment protocols. It has been argued that such generic scales fail to assess the disorder-specific strategies and procedures thought to be central mechanisms of change in CBT (Fairburn & Cooper, 2011). Hence a number of disorders-specific scales have been developed in recent years (e.g. the Cognitive Therapy Scale-Psychosis Psychosis [Haddock et al., 2001], Yale Adherence and Competence Scale [Carroll et al., 2000], the Cognitive Therapy Competence Scale for Social Phobia [von Consbruch, Clark, & Stangier, 2012]. Such disorder-specific scales may be highly relevant for use within controlled research settings where patients are selected and treated on the basis of diagnosis. However, such diagnosis-specific scales may be less applicable within wider clinical practice, particularly training settings in which clinicians learn how to deliver a variety of CBT protocols or clinical practices where therapists

work with patients experiencing a range of mental health problems and often encounter co-morbidity (Barber et al., 2007). Furthermore, there has been a recent proliferation of manuals targeting single disorders. Thus, focusing on competences within the context of disorder-specific treatment protocols would require the use of multiple different scales, thereby drastically increasing the financial and resource implications of this method of assessment. Due to concerns such as these, many experts in the field have argued that both training and routine practice assessments should focus on core transdiagnostic competences (see Chapter two). It was therefore decided that the scale would focus on core transdiagnostic CBT competences, rather than attempting to develop one of many disorder-specific scales.

Fourth, the scale is designed for use within an 'active' mid-treatment session, rather than assessment sessions, very early treatment sessions or end-phase treatment sessions. Thus, the scale will not assess those skills evident only at the very beginning or end of treatment (e.g. assessment, goal setting, relapse prevention etc.). This decision was made because mid-treatment sessions typically involve the active selection and implementation of strategies designed to target and alter processes maintaining the patient's difficulties. Finally, the scale focuses on the competences required to deliver CBT for adults. This is because adaptation is required when working with other population groups (e.g. children, adolescents, individuals with intellectual disabilities) rather than an adult population. As a result, there are a number of differences in the competences required to deliver effective CBT to these populations (Hupp, Jewell, Reitman, & LeBlanc, 2010; Roth, Calder, & Pilling, 2013; Roth & Pilling, 2007), with rating scales being developed specifically to assess competence in CBT with children and young people (Stallard, Myles, & Branson, 2014).

Step 2: Create an Item Pool

Following the deductive approach (Burisch, 1984), a theoretical definition of the construct under examination was used to guide the development of scale items. This was achieved by examining the competences outlined in Roth and Pilling's (2007) competence framework, revisiting interview transcripts obtained in the qualitative study (see Chapter two), and reviewing items included in other relevant scales identified in the literature review (e.g. the CTS, CTS-R, CTACS- see Chapter one) to create an item pool. The initial review resulted in a very large potential item set, which was refined by removing redundant items, items too specific to a treatment protocol or disorder (e.g. skilful use of a specific technique such as activity scheduling or re-living), and items unlikely to be observable within the majority of mid-treatment sessions (e.g. assessment skills, relapse prevention skills, addressing alliance ruptures, managing risk etc.). Items that tapped into overlapping constructs were also amalgamated. Finally, items assessing

similar aspects of CBT competence were grouped together within overarching domains. This was an iterative refinement process, which was carried out by Kate Muse under the supervision of Dr Freda McManus.

There are no clear guidelines as to how many items a scale should include. Keeping a measure short reduces the time taken to develop and administer the scale and minimises the impact of response biases caused by boredom or fatigue, whilst increasing the number of items in a scale increases its internal consistency and reliability (Carmines & Zeller, 1979; Streiner & Norman, 2003). There is also a balance to be struck between including too few items and thus failing to adequately capture the multi-faceted nature of competence, and including too many items, resulting in a high degree of item overlap. It has been suggested that a minimum of four items are needed to test the homogeneity of items within each construct and establish adequate reliability (Carmines & Zeller, 1979; Harvey, Billings, & Nilan, 1985). The current scale is much larger than this minimum to enable formative assessment and summative feedback on of a range of CBT competences, to ensure that each aspect of the construct was adequately represented, and to allow for further scale refinement (L. A. Clark & Watson, 1995; DeVellis, 2012).

The final item pool consisted of eight core competence domains, each of which includes two to four items (see Figure one). Once the final item structure was outlined, a definition of each domain was created. Domain definitions were written following questionnaire design principles to avoid the following issues: overly long items, double-barrelled items (i.e. a single item which assesses multiple aspects of CBT competence), multiple negatives, vague or ambiguous terminology, hypothetical questions, leading questions, positive or negatively loaded words and hidden assumptions (Fife-Schaw, 2006; Fowler, 1995). A brief overview of each domain is provided below.

Before outlining the domains included within the scale, it is worth noting that there is not a domain dedicated to Socratic questioning. Socratic questioning is a teaching method that uses a progressive series of questions designed to facilitate independent thinking and promote active learning (Kennerley, 2007; Overholser, 2010). This questioning should 'guide the patient's discovery' to reveal new possibilities and help patients gain new perspectives. There is widespread agreement that Socratic questioning within the context of collaborative empiricism and guided discovery is the 'cornerstone' of CBT (J. S. Beck, 1995; Overholser, 2010; Padesky, 1993; Roth & Pilling, 2007; Tee & Kazantzis, 2011) and should underpin all aspects of CBT, from conceptualising patient's problems through to devising and reviewing CBT interventions. Thus, Socratic Questioning was embedded within the rating framework of other specific competences (e.g. conceptualisation, agenda setting, delivering CBT interventions etc.).

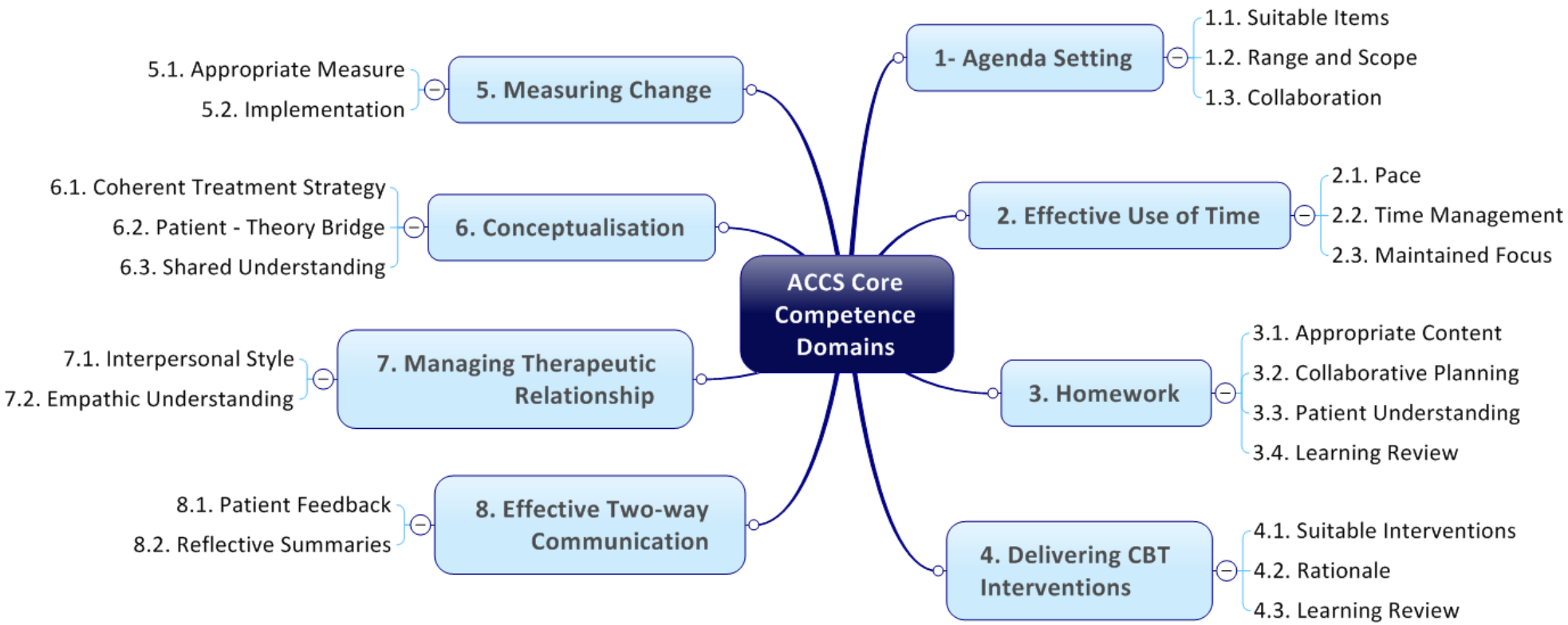


Figure 1: Final item structure of the rating scale

It is also worth noting that patient difficulty has not been incorporated within the skills domains as this may unduly confound the ratings provided within the domains. Furthermore, reserving the highest ratings for performance demonstrated in the face of patient difficulty may unfairly prevent those working with more straightforward cases from being awarded a high performance rating. However, assessors are asked to provide a patient difficulty rating within the scale. This is to identify the context within which the therapist's performance has been evaluated.

Domain 1 - agenda setting.

Ability to collaboratively set a realistic, mutually agreed agenda which prioritises suitable, specific target problems and includes all appropriate routine items.

CBT is a relatively short-term treatment, typically involving between six and 15 one-hour sessions (Westbrook, Kennerley, & Kirk, 2007). To make optimal use of this limited time, it is recommended that the therapist and patient spend a short amount of time at the beginning of each session to agree an agenda (J. S. Beck, 1995; Newman, 2013; Roth & Pilling, 2007; Westbrook et al., 2007; J. H. Wright, 2006). Setting an agenda is thought to facilitate the completion of therapeutic tasks and reduce unhelpful drift by helping the therapist and patient identify the most pertinent issues that need to be addressed within each treatment session. Three items were included within this domain. First, 'suitable items' assesses whether the therapist prioritised suitable, specific target problems and included all appropriate routine items. Second, 'range and scope' examines whether the therapist prioritised an appropriate number of items with suitable scope and significance and thus the degree to which the agenda can be considered realistic. Third, 'collaboration' assesses whether the therapist used appropriate collaboration to develop a mutually agreed agenda.

Domain 2 - effective use of time.

Ability to structure, pace and manage time within the session in a manner which is appropriate to the therapeutic context and patient's capacity for learning and enables maintained focus on important issues.

As outlined in domain one, CBT is a relatively brief, time limited treatment. Thus it is important to make optimal use of the time available. Making effective use of the time involves three key components, which are outlined as separate items within this domain (J. S. Beck, 1995; Newman, 2013; Westbrook et al., 2007). First, 'pace' examines whether the

pace of the session was appropriate to the therapeutic context and patient's capacity for learning, thereby examining the degree to which the therapist struck an appropriate balance between moving at a speed which enabled a sufficient amount to be accomplished during the session and moving so fast that the patient was overwhelmed. Second, 'time management' assesses whether the therapist was able to efficiently balance the allocation of time within the session in order to enable important issues to be covered within the session. Finally, 'maintained focus' examines the degree to which the therapist helped the patient stay focused on the tasks and goals of the session whilst maintaining appropriate flexibility and sensitivity in response to unanticipated issues.

Domain 3 - homework.

Ability to plan achievable homework which has clear and specific goals, is tailored to the individual patient's therapeutic context and is drawn directly from session material. To plan the homework in a collaborative manner which ensures the patient understands the purpose and content of the assignment and has the opportunity to discuss their reaction to it. To conduct a detailed review of previous homework in order to help the patient identify what they learned from the experience.

A number of research studies have found that therapists' assignment of homework, and patients' compliance to homework tasks, leads to better treatment outcomes (Bryant, Simons, & Thase, 1999; Burns & Nolen-Hoeksema, 1991, 1992; Fennell & Teasdale, 1987; Kazantzis, Deane, & Ronan, 2000; Kazantzis & Lampropoulos, 2002; Neimeyer & Feixas, 1990; Persons, Burns, & Perloff, 1988; Rees, McEvoy, & Nathan, 2005; Ryum, Stiles, Svartberg, & McCullough, 2010), indicating that homework plays an important role in the effective delivery of CBT. Homework is considered an integral part of CBT for a number of reasons (A. T. Beck, Rush, Shaw, & Emery, 1979; Kazantzis, Deane, Ronen, & Luciano, 2005). It is hypothesised that homework enables patients to begin to generalise and transfer in-session learning to situations outside the therapeutic setting, thus helping them to make meaningful changes in their day to day lives. By allowing patients to gain direct experience of completing therapeutic tasks in problematic situations, homework is also thought to promote experiential learning at emotional and cognitive levels. Finally, homework encourages patient autonomy and independence, thus supporting the maintenance of improvements after treatment has ended and reducing the risk of relapse.

Four items were created to capture the therapist behaviours believed to facilitate homework compliance (A. T. Beck et al., 1979; Bryant et al., 1999; Kazantzis et al., 2005; Kazantzis & Lampropoulos, 2002). The first three items refer to the way in which the therapist planned homework tasks. First, 'appropriate content' examines whether the therapist planned achievable

homework, which had clear and specific goals, was tailored to the individual patient's therapeutic context and was drawn directly from session material. Second, 'collaborative planning' assesses the degree to which the therapist planned the homework in a collaborative manner. Third, 'patient understanding' assesses whether the therapist facilitated the patient's understanding of the purpose and content of the assignment and elicited their reaction to it. The fourth item moves on to examine the process of reviewing homework. Specifically, the 'learning review' item assesses whether the therapist conducted a detailed review of previous homework in order to help the patient identify what they learned from the experience.

Domain 4 - delivering CBT interventions.

Ability to select suitable evidence-based cognitive-behavioural interventions to target clearly specified key cognition(s) or maintaining factors, skilfully apply and review the impact of interventions in a manner which ensures that the aim of the intervention is clear to, and understood by, the patient and that helps the patient identify what they learned from the experience.

This domain refers to the selection and application of cognitive and/or behavioural interventions, also referred to as CBT 'techniques' or 'change strategies'. As cognitive and behavioural interventions are a key defining feature of CBT, it is important to examine whether therapists have selected and skilfully implemented suitable interventions (J. S. Beck, 1995). CBT interventions are designed to promote therapeutic change by targeting cognitions, behaviours or emotions that both relevant CBT theory and the patient's individual formulation suggest are likely to be maintaining the patient's problems. A range of cognitive and behavioural interventions can be used to promote change. Examples include cognitive restructuring, exposure/response prevention, addressing safety behaviours, imagery re-scripting, examining pros and cons, thought records, activity monitoring/scheduling, behavioural experiments, role play, graded task assignment, problem solving, assertiveness or other skills training, and behaviour modification. However, this list is by no means exhaustive and is likely to change over time as new intervention techniques are developed.

Three items are used to assess the skilful selection and delivery of CBT interventions. First, 'suitable interventions' determines whether the therapist selected suitable evidence-based cognitive-behavioural interventions to target clearly specified cognitions, behaviours or other maintaining factors. Second, 'rationale' assesses whether the therapist made sure that the aim of the intervention(s) was clear to, and understood by, the patient. Finally, 'learning review' assesses whether the therapist reviewed the impact of interventions to help the patient identify

what they learned from the experience. It is also worth noting that, as a number of factors other than therapist competence can influence whether an intervention is successful in bringing about change, these items focus on the quality of the cognitive-behavioural interventions and not whether change actually occurred.

Domain 5 - measuring change.

Ability to select appropriate, clinically relevant methods for measuring change in symptoms and associated features (beliefs, behaviours, feelings) and movement towards goals. Ability to administer these methods at suitable time points across and within session and to skilfully review, interpret and respond to the information gleaned.

The regular use of valid and reliable measures serves a number of useful functions within CBT treatment (J. S. Beck, 1995; Kirk, 1998; Westbrook et al., 2007). Using measures regularly allows the patient and therapist to establish a baseline of important aspects of the problem (e.g. symptoms, beliefs, behaviours, feelings etc.) and then to use this to evaluate the impact of interventions. Regular measurement also allows the patient and therapist to objectively, accurately and consistently assess treatment progress and movement towards goals. An additional benefit to the regular use of measures is that observations of behaviours, thoughts and feelings recorded at the time of occurrence are typically more accurate and reliable than retrospective accounts. A range of methods can usefully be employed to measure change, including standardised questionnaires and idiosyncratic measures (e.g. diaries, frequency counts, ratings of duration of event or experience, self-ratings of emotions and cognitions etc.). It is recommended that relevant measures be employed across sessions to measure change in symptoms and movement towards goals over the course of treatment as a whole. However, these methods should also be employed within treatment sessions to measure the impact of cognitive-behavioural interventions completed within session and for homework (e.g. belief ratings pre/post intervention). Two items were included within the domain of measuring change. The first, ‘appropriate measure’, is used to determine whether the therapist selected appropriate, clinically relevant methods for measuring change. The second, ‘implementation’ assesses the degree to which the therapist administered these methods at suitable time points across and within the session and skilfully reviewed, interpreted and responded to the information gleaned.

Domain 6 - conceptualisation.

Ability to develop a mutually agreed conceptualisation (or 'formulation') which provides a bridge between CBT theory and the individual patient and offers a coherent, meaningful and relevant account of the key features of the patient's problems. Ability to draw upon the conceptualisation in order to facilitate understanding for both the patient and therapist and to develop a coherent treatment strategy.

Case conceptualisation (otherwise known as case formulation) involves the therapist working with the patient to derive and then explain their presenting issues (Kuyken, Padesky, & Dudley, 2011). It has been widely claimed that case conceptualisation is an underpinning principle of CBT and that a comprehensive case conceptualisation is needed for successful treatment (J. S. Beck, 1995; Bieling & Kuyken, 2003; I. Blackburn & Twaddle, 1996; Butler, 1998; Kuyken et al., 2011; Pearsons, 1993). Kuyken et al. (2011) argue that case conceptualisation is an essential competent of CBT because it syntheses patient experience, CBT theory and research, normalises patient's presenting issues, promotes patient engagement, makes complex problems more manageable, guides the selection of interventions, and highlights potential problems.

An initial conceptualisation is typically developed in the early phase of treatment as part of the assessment process. However, a conceptualisation should not simply precede treatment but should evolve throughout therapy as further clinical information arises and recurrent patterns and themes emerge. This information will need to be added to the conceptualisation or may require the conceptualisation to be modified or revised. Additionally, the therapist should draw upon and refer to the conceptualisation throughout treatment to help the patient understand their problems within a CBT framework and to highlight potential treatment strategies. Thus a therapist's ability to conceptualise a patient's problems should be evident throughout the course of treatment (Butler, 1998; Kuyken et al., 2011). For this reason, it was considered appropriate to include conceptualisation as a core domain within the current scale.

There are many components involved in competently developing and using a case conceptualisation. Indeed, an entire scale has been developed to measure this construct alone (Collaborative Case Conceptualisation - Rating Scale: Kuyken et al., 2011). However, for the purposes of the current scale, these components were distilled within three items. First, 'coherent strategy' assess whether the therapists drew upon the conceptualisation to develop a coherent treatment strategy. Second, 'patient theory bridge' determines whether the conceptualisation provided a bridge between CBT theory and the individual patient in order to offer a coherent, meaningful and relevant account of the key features of the patient's problems.

Finally, 'shared understanding' examines the degree to which the therapist developed a mutually agreed conceptualisation and drew upon to facilitate understanding for both the patient and therapist.

Domain 7 - managing therapeutic relationship.

Ability to foster a positive therapeutic alliance by (i) accurately grasping the content and emotional tone of the patient's understanding of themselves and the world around them and sensitively and appropriately conveying this understanding and (ii) embodying an excellent interpersonal style which is congruent with the therapeutic context and patient's needs.

The therapeutic relationship can be simply defined as "the personal qualities of the patient, personal qualities of the therapist, and the interactions between them" (Jesse H. Wright & Davis, 1994, p.27). Cognitive behavioural therapists are not alone in highlighting the importance of the therapeutic relationship within treatment. Indeed, it is a longstanding and widely held belief that a sound therapeutic alliance is an essential component for success in psychotherapy (Barrett-Lennard, 1962; A. T. Beck et al., 1979; J. S. Beck, 1995; Gilbert & Leahy, 2007; Greenson, 2008; Horvath & Luborsky, 1993; Raue & Goldfried, 1994; Rogers, 1957). Research has also provided supported for this claim, with a number of studies reporting that the quality of the therapeutic relationship is predictive of treatment outcomes, although patients' and observers' judgments of the alliance appear to be more predictive of outcome than therapists' reports (Gaston, 1990; Horvath & Symonds, 1991; Keijsers, Schaap, & Hoogduin, 2000; Klein et al., 2003; luborsky, 1994; Martin, Garske, & Davis, 2000; Orlinsky, Ronnestad, & Willutzki, 2004; Trepka, Rees, Shapiro, Hardy, & Barkham, 2004; Waddington, 2002). Although it has been argued that the development and maintenance of the therapeutic relationship is a primary curative component of psychotherapy, CBT theorists largely advocate the idea that a sound working alliance is a necessary context for intervention, but is not sufficient to bring about change (A. T. Beck et al., 1979; Gilbert & Leahy, 2007; Waddington, 2002).

A number of patient and therapist factors are thought to influence the development and maintenance of a positive therapeutic alliance (Gilbert & Leahy, 2007; Horvath & Luborsky, 1993; Waddington, 2002). As the current scale focuses on the competence of the therapist, two key therapist factors are included. The first item, 'interpersonal style', focuses on the therapist's ability to embody an excellent interpersonal style congruent with the therapeutic context and patient's needs. The second item, 'empathic understanding', examines whether the therapist accurately grasps the content and emotional tone of the patient's understanding of themselves and the world around them, and convey this understanding sensitively and appropriately. These

items reflect the Rogerian therapist variables of empathic understanding and interpersonal skills such as genuineness, warmth and unconditional positive regard toward the patient (A. T. Beck et al., 1979; Rogers, 1957) as therapists who demonstrate these characteristics have been shown to achieve better outcomes (Burns & Nolen-Hoeksema, 1992; Orlinsky, Klaus, & Parks, 1994). However, given the collaborative, reciprocal nature of the relationship in CBT, these items also assess the degree to which these behaviours are matched to the patient's therapeutic context (e.g. interpersonal style, history etc.) and the therapist's ability to effectively communicate these qualities to the patient.

Domain 8 - effective two-way communication.

Ability to elicit, explore and respond to in-depth feedback about the patient's reaction to all aspects of session and to work with the patient to reflect upon and summarise salient session content in order to facilitate joint understanding of crucial therapeutic material and consolidate key learning.

The first item in this domain, 'patient feedback', relates to the therapist's ability to elicit, explore and respond to the patient's reaction to all aspects of session. Patient feedback is typically sought at the beginning and end of every treatment session, as well as at regular intervals throughout the session. Eliciting feedback can strengthen the therapeutic alliance by illustrating to the patient that the therapist genuinely cares about the patient's viewpoint (J. S. Beck, 1995). Actively encouraging patients to express negative reactions to aspects of the session, or to the therapist themselves, also allows the therapists to identify and respond appropriately to potential barriers to treatment or problems in the therapeutic relationships (J. S. Beck, 1995). The second item, 'reflective summaries', refers to the therapist's ability to work with the patient to reflect upon and summarise salient session content in order to facilitate joint understanding of crucial therapeutic material and consolidate key learning. Reflective summaries can be helpful in structuring sessions and maintaining focus on important issues (Roth & Pilling, 2007). Eliciting in-depth patient reflections on and summaries of session content also provides therapists with the opportunity to identify and resolve any confusion or misunderstandings (J. S. Beck, 1995). Finally, both eliciting and providing capsule summaries of salient information discussed in the session can facilitate the patient's understanding of crucial therapeutic material by highlighting and consolidating key learning points (J. S. Beck, 1995).

Step 3: Create a Response Scale

The response scale was created by Kate Muse under the supervision of Dr Freda McManus. Expert feedback was also sought at each stage of development. This expert group consisted of eight researchers from the University of Oxford with a range of experience in developing and evaluating questionnaire measures (i.e. psychometric expertise). The following five issues were addressed in creating a response scale;

3.1. Select an appropriate response format.

A number of possible response formats were considered. Some, such as open questions, ranking items, and binary responses are not appropriate for a multi-item competence rating scale. Semantic differential scales and visual analogue scales (VASs) are also not well suited to the current purposes. Semantic differential scales contain a series of bipolar adjectives for the various properties of the object under study and respondents indicate their impressions of each property by indicating locations along its continuum (e.g. *powerful* 3 2 1 0 1 2 3 *weak*). However, semantic differential scales are prone to central tendency error and demonstrate poor sensitivity (DeVellis, 2012; Fowler, 1995). A VAS asks respondents to mark a continuous line between pairs of descriptors representing opposite ends of a continuum. A VAS scale is difficult and time consuming to administer. Furthermore, although they allow sensitive measurement, respondents often struggle to differentiate between the huge range of scoring possibilities available in a VAS, meaning that many points on the scale remain unused (DeVellis, 2012; Fowler, 1995). The most appropriate response format for the current purposes is a rating scale. Likert scales are a popular form of rating scale, which provide a declarative sentence with response options with varying degrees of agreement or endorsement of the statement (e.g. *agree* to *disagree*). Likert scales are cognitively complex and are less reliable and more prone to acquiescence bias than questions with item-specific response options (Converse & Presser, 1986; Fowler, 1995; Saris, Revilla, Krosnick, & Shaeffer, 2010). Thus, item-specific response options appear to be the best choice of response format for the current scale.

3.2. Identify optimum response categories.

Deciding how many scale points to include is an important aspect of scale construction. Using a large range of response categories enables a scale to discriminate between differences in the underlying construct and limits the degree to which the tendency to avoid using extreme response categories deflates variance (Streiner & Norman, 2003; Van Vaerenbergh & Thomas, 2012). However, too large a range may not be beneficial. Masters (1974) found that scale reliability increases up to four scale points and levels off thereafter, and Matell and Jacoby (1972)

report that scales longer than seven points do not convey additional substantive information. Furthermore, respondents prefer scales with fewer categories, finding scales with more than ten options more time consuming and less easy to use (Krosnick & Fabrigar, 1997; Preston & Colman, 2000). The optimal length of a rating scale is therefore widely considered to be between five and seven points (Krosnick & Fabrigar, 1997). It was therefore decided that five response options would be used in this scale in order to allow adequate discrimination between levels of competence without making the scale unwieldy or difficult to use.

3.3. Scale anchors.

It was necessary to consider whether to label all scale points with written descriptions or with numbers only. There is little difference between scales with adjectives under each box and end-anchored scales (Dixon, Bobo, & Stevick, 1984; Newstead & Arnold, 1989) and research indicates that respondents are more influenced by the adjectives on the end of scales than those in the intermediate positions (Frisbie & Brandenburg, 1979; Wildt & Mazis, 1978). However, respondents are more satisfied when many or all of the points on the scale are labelled (Dickinson & Zellinger, 1980; Wallsten, Budescu, Zwick, & Kemp, 1993). Furthermore, using behaviourally anchored scale points can help to improve agreement between raters, reduce the impact of the halo effect and improve measurement validity (Krosnick & Fabrigar, 1997; Streiner & Norman, 2003). The current scale, therefore, provides a generic anchor for each scale point as well as item-specific behavioural anchors for each scale point. In addition, numbers are provided to help the respondent select an appropriate response option. However, as respondents rarely endorse negative scale points (Schwarz, Knauper, Hippler, Noelle-Neumann, & Clark, 1991), only values above zero were used.

After deciding on the response format, response categories and scale anchors, two response scales were created. The first scale is a generic five-point scale (see Figure two). This is used to provide an overarching framework for the subsequent item-specific scales. It is also used to provide a ‘global rating’ reflecting the assessor’s overall impression of the therapist’s performance within the session as a whole.

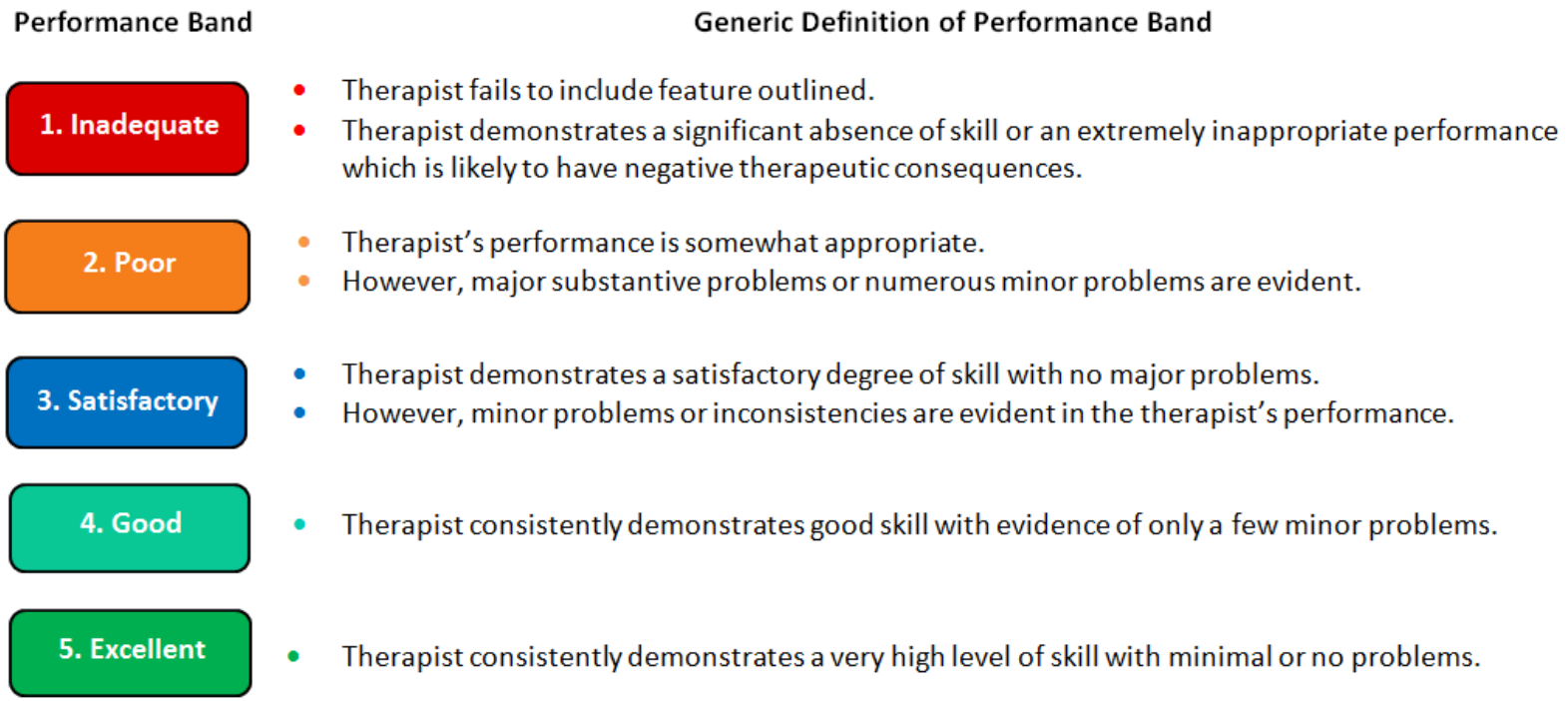


Figure 2: Overarching definitions for the five-point scale

The second scale is item-specific and provides concrete examples of behaviours consistent with each scale point (see an example in Figure three). These 'exemplar therapist behaviours' are intended to provide guidance as to the type of performance which is consistent with each point on the rating scale for that particular item.

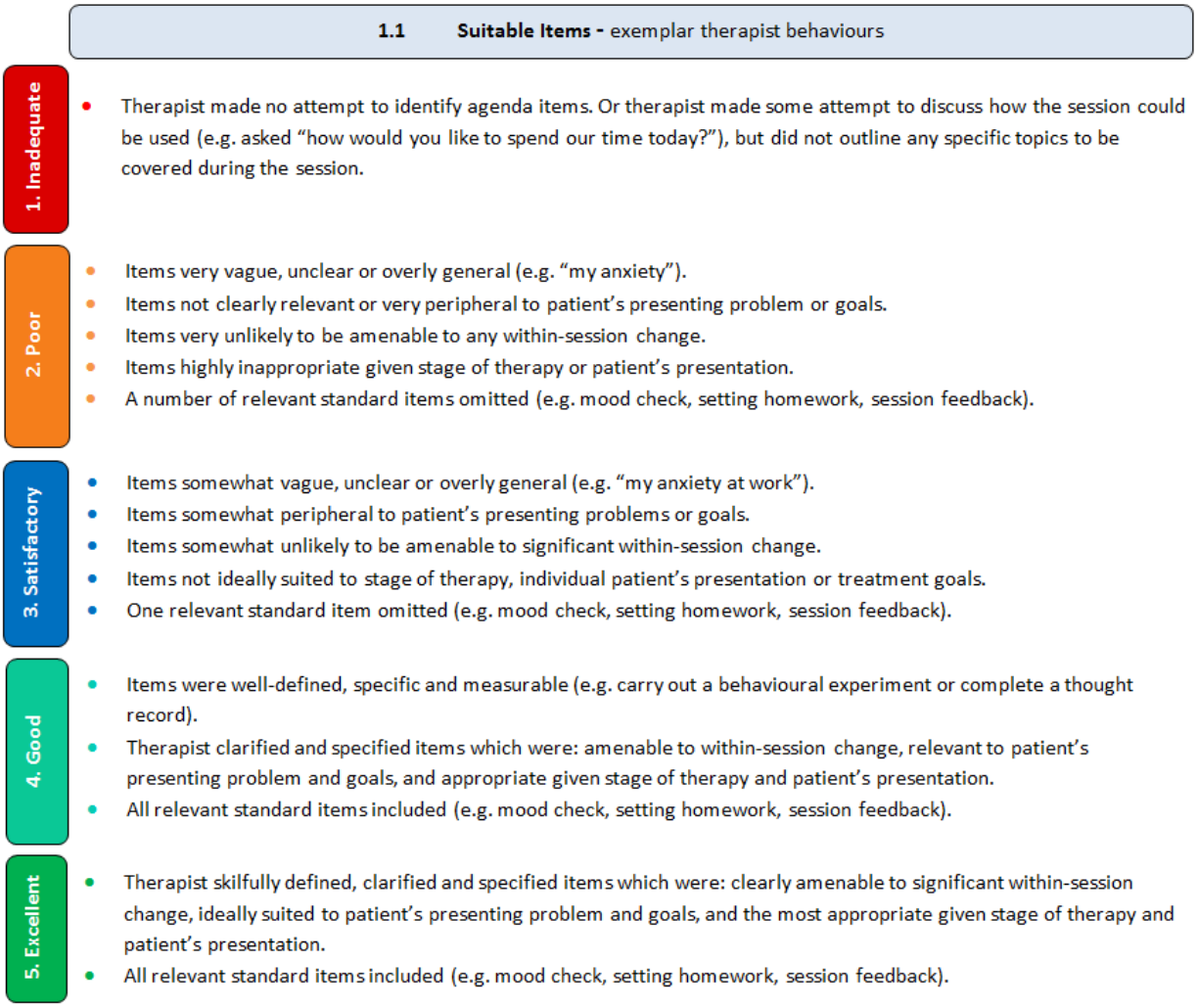


Figure 3: An example of item-specific concrete examples of behaviours consistent with each scale point for item 1.1.

3.4. Item scoring.

All items were included because expert opinion, relevant theory or relevant research indicates that the specific skill is an important and relevant aspect of CBT competence. However, at this stage little is known about whether some skills are of more importance than others. This is because there are very few studies examining the relationship between CBT competence and patient outcome, and those which do tend to focus on 'packages' of skills (Ginzburg et al., 2012;

Karlin et al., 2012; Kuyken & Tsivrikos, 2009; Shaw et al., 1999; Strunk, Brotman, DeRubeis, & Hollon, 2010; Trepka et al., 2004). Therefore, the decision was made to give each item equal weight in the scoring of the scale. Thus, individual items can be added together to provide a total domain score. These total domain scores can then be added to create a total scale score. A benefit of this simple scoring format is that is easy to use, therefore reducing the burden on assessors using the scale.

3.5. Feedback form.

The scale aims to provide both formative and summative feedback about a therapist's performance within an observed treatment session: that is, it aims to provide assessors with the space to provide in-depth narrative feedback in addition to a numerical rating of competence. This is because assessments that serve both formative and summative functions are viewed by experts in the field as being the most effective way of promoting therapists' development of CBT skills (see Chapter two). Furthermore, formative feedback plays an integral role in the development of competent, reflective practitioners, both within the context of training and on-going continuing professional development (Dochy & Moerkerke, 1997; Milne, 2007; Van der Vleuten et al., 2010), and is typically well received by those being assessed (Govaerts, Vleuten, Schuwirth, & Muijtjens, 2005; Shute, 2008; Van der Vleuten et al., 2010). Thus, within the feedback form (Appendix one) assessors are not only given the space to record their numerical ratings, but are also asked to specific highlight areas of strength and learning needs. Assessor can draw on the exemplar therapist behaviours provided and the session material being assessed to provide examples of what made some aspects of the therapist's performance more successful than others. Where areas of weakness have been identified, assessors can suggest alternative, more skilful way of working within this domain and highlight ways in which the therapist could further develop their skills in this area.

Step 4: Create a Manual

The accompanying rating manual (see Appendix one) had two key aims. First, to provide guidance for assessors as to how to make judgements about the quality or skilfulness of therapists' performance using the rating scale. As well as general guidelines, the manual provides clarification about the constructs measured by each scale item by detailing item-specific examples of behaviours consistent with each scale point. Second, to specify implementation guidelines outlining what materials the assessments should be based upon and who is best placed to assess competence using the scale.

Discussion

This chapter focuses on the process of developing the Assessment of Core CBT Skills (ACCS) rating scale. In developing the scale a number of factors were considered to ensure the quality and applicability of the scale, particularly the issues of validity, reliability, usability, and feasibility. The main difficulty encountered was incorporating and navigating all of these various issues in order to ensure the scale was not only psychometrically sound but also fit for purpose and likely to be used in practice. In particular, four key areas of difficulty arose. First, whilst it was important to ensure the measure adequately sampled the different aspects of CBT competence (i.e. had good content validity), a number of decisions were made which restricted the range of the aspects of CBT competence included in the scale. This included the decision to focus on transdiagnostic aspects of competence evident in 'active' mid-treatment sessions, a decision that was made in order to ensure the scale was applicable across a range of training and routine practice settings. Similarly, Roth and Pilling (2007) outline over fifty specific competences required to deliver CBT. However, the ACCS item pool focuses on a subset of core competences in order to ensure that the scale has adequate feasibility and usability, whilst still trying to ensure that each aspect of the construct of CBT competence was adequately represented.

Second, Roth and Pilling's (2007) framework suggests that there is likely to be a great deal of overlap between the competences required to deliver CBT. However, it was important to attempt to separate CBT competence into relatively specific, independent items in order to achieve adequate inter-rater reliability and to ensure the scale was easy for both novice and experienced therapists to use. Third, a detailed manual was included providing information about how to use the scale as well as item-specific anchor descriptions. Although this means that the ACCS is more complex and time consuming to use at first, the manual was included in order to achieve adequate inter-rater reliability, ensure the provision of effective formative feedback, encourage detailed self-reflection, and preclude the need for assessor training. Fourth, the decision to use five scale points involved a careful balance between enabling sensitivity to change, whilst maintaining usability. It was also decided to use performance-based headings (e.g. good, excellent) as titles for the generic five-point scale. Sharpless and Barber's (2009) competence framework suggests that therapists progress through five key developmental stages: novice, advance beginner, competence, proficiency and finally expertise. However, the stage-specific benchmarks outlined in this framework were not used as titles for the ACCS scale bandings. Competence is considered to be a fluctuating construct, rather than stable construct. Indeed, experts in the field have warned of the potential for 'therapist drift' if a therapist does not sufficiently monitor and maintain their clinical knowledge and skills (Waller, 2009). Viewing

competence as a fluctuating construct undermines the use of a scale with concrete developmental banding titles. This is particularly the case within routine practice settings where therapists should already have reached the developmental milestone of competence, but may well still fall down on their level of skill in one or more of the number of items included in the ACCS at some points in time. For example, a number of experts in the qualitative study outlined in chapter two commented that over time they had noticed that they sometimes either forgot to implement or rushed some of the 'basics' such as setting an agenda. In an example such as this, the individual would be considered to be a proficient or expert therapist. However, on a rating scale using developmental bandings, they would have been labelled as 'novice' or 'advanced beginner' in the domain of agenda setting.

The numerous decisions made in developing a first draft of the ACCS scale were based on careful examination of the wider literature and evidence-base and gave consideration to a wide range of psychometric and practical issues. However, this can only be considered to be the first stage of the process of scale development. In order for the scale to be considered fit for purpose, a number of further steps will need to be taken, including seeking feedback from a range of novices and experts in the field and examining the psychometric properties of the scale. These evaluative processes will be required in order to confirm that the scale has adequate validity, reliability, usability and feasibility. However, they will also serve a vital role in highlighting further developments and refinements that are required in order to provide a coherent and robust CBT competence rating scale.

References

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

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology-Research and Practice, 38*(5), 493-500. doi: 10.1037/0735-7028.38.5.493

Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs: General and Applied, 76*(43), 1-36. doi: 10.1037/h0093918

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New-York: Guilford Press.

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. New York: Guildford Publications.

Bieling, P. J., & Kuyken, W. (2003). Is cognitive case formulation science or science fiction? *Clinical Psychology: Science and Practice, 10*(1), 52-69. doi: 10.1093/clipsy.10.1.52

Blackburn, I., & Twaddle, V. (1996). *Cognitive therapy in action*. London: Souvenir Press.

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. doi: 10.1017/S1352465801004040

Bryant, M. J., Simons, A. D., & Thase, M. E. (1999). Therapist skill and patient variables in homework compliance: Controlling an uncontrolled variable in cognitive therapy outcome research. *Cognitive Therapy and Research, 23*(4), 381-399. doi: 10.1023/A:1018703901116

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist, 39*(3), 214-227. doi: 10.1037/0003-066x.39.3.214

Burns, D. D., & Nolen-Hoeksema, S. (1991). Coping styles, homework compliance, and the effectiveness of cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 59*(2), 305-311. doi: 10.1037/0022-006x.59.2.305

Burns, D. D., & Nolen-Hoeksema, S. (1992). Therapeutic empathy and recovery from depression in cognitive-behavioral therapy: A structural equation model. *Journal of Consulting and Clinical Psychology, 60*(3), 441-449. doi: 10.1037/0022-006x.60.3.441

Butler, G. (1998). Clinical formulation. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology* (Vol. 6). Oxford: Elsevier Science Pergamon Press.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. London, UK: Sage Publications.

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

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23*(4), 318-327. doi: 10.3109/09540261.2011.606803

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319. doi: 10.1037/1040-3590.7.3.309

Converse, J., & Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Thousand Oaks, CA: Sage.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281 - 302. doi: 10.1037/h0040957

DeVellis, R. F. (2012). *Scale development: Theory and Applications*. Thousand Oask, CA: Sage.

Dickinson, T. L., & Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of Applied Psychology, 65*(2), 147-154. doi: 10.1037/0021-9010.65.2.147

Dixon, P. N., Bobo, M., & Stevick, R. A. (1984). Response Differences and Preferences for All-Category-Defined and End-Defined Likert Formats. *Educational and Psychological Measurement, 44*(1), 61-66. doi: 10.1177/0013164484441006

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49*(6-7), 373-378. doi: 10.1016/j.brat.2011.03.005

Fennell, M. V., & Teasdale, J. (1987). Cognitive therapy for depression: Individual differences and the process of change. *Cognitive Therapy and Research, 11*(2), 253-271. doi: 10.1007/bf01183269

Fife-Schaw, C. (2006). Questionnaire design. In G. M. Breakwell, S. Hammond, C. Fife-Schaw & J. A. Smith (Eds.), *Research methods in psyhcology* (3rd ed., pp. 210-232). London: Sage Publications Ltd.

Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks, CA: Sage.

Frisbie, D. A., & Brandenburg, D. C. (1979). Equivalence of questionnaire items with varying response formats. *Journal of Educational Measurement, 16*(1), 43-48. doi: 10.1111/j.1745-3984.1979.tb00085.x

Gaston, L. (1990). The concept of the alliance and its role in psychotherapy: Theoretical and empirical considerations. *Psychotherapy: Theory, Research, Practice, Training, 27*(2), 143-153. doi: 10.1037/0033-3204.27.2.143

Gilbert, P., & Leahy, R. L. (2007). *The therapeutic relationship in the cognitive behavioral psychotherapies*. New York: Routledge.

Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy, 50*(12), 747-752. doi: 10.1016/j.brat.2012.09.001

Greenson, R. R. (2008). The working alliance and the transference neurosis. *The Psychoanalytic Quarterly, LXXVII*(1), 77-102. doi: 10.1002/j.2167-4086.2008.tb00334.x

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. doi: 10.1017/S1352465801002089

Harvey, R. J., Billings, R. S., & Nilan, K. J. (1985). Confirmatory factor analysis of the Job Diagnostic Survey: Good news and bad news. *Journal of Applied Psychology, 70*(3), 461-468. doi: 10.1037/0021-9010.70.3.461

Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology, 61*(4), 561-573. doi: 10.1037/0022-006x.61.4.561

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*(2), 139-149. doi: 10.1037/0022-0167.38.2.139

Hupp, S. D. A., Jewell, J. D., Reitman, D., & LeBlanc, M. (2010). Competencies in Child Clinical Psychology. In J. C. Thomas & M. Hersen (Eds.), *Handbook of Clinical Psychology Competencies* (pp. 43-72): Springer New York.

Karlin, B. E., Brown, G. K., Trockel, M., Cunning, D., Zeiss, A. M., & Taylor, C. B. (2012). National dissemination of cognitive behavioral therapy for depression in the department of veterans affairs health care system: Therapist and patient-level outcomes. *Journal of Consulting and Clinical Psychology, 80*(5), 707-718. doi: 10.1037/a0029328

Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist, 59*(8), 774-781. doi: 10.1037/0003-066x.59.8.774

Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency assessment toolkit for professional psychology. *Training and Education in Professional Psychology, 3*(4), 27 - 45. doi: 10.1037/a0015833

Kazantzis, N. (2003). Therapist competence in cognitive-behavioural therapies: Review of the contemporary empirical evidence. *Behaviour Change, 20*(1), 1-12. doi: 10.1375/bech.20.1.1.24845

Kazantzis, N., Deane, F. P., & Ronan, K. R. (2000). Homework Assignments in Cognitive and Behavioral Therapy: A Meta-Analysis. *Clinical Psychology: Science and Practice, 7*(2), 189-202.

Kazantzis, N., Deane, F. P., Ronen, K. R., & Luciano, L. (2005). *Using homework assignments in cogitive behavior therapy*. Abgindon: Routledge.

Kazantzis, N., & Lampropoulos, G. K. (2002). Reflecting on homework in psychotherapy: What can we conclude from research and experience? *Journal of Clinical Psychology, 58*(5), 577-585. doi: 10.1002/jclp.10034

Keijsers, G. P. J., Schaap, C. P. D. R., & Hoogduin, C. A. L. (2000). The impact of interpersonal patient and therapist behavior on outcome in cognitive-behavior therapy: A review of empirical studies. *Behavior Modification, 24*(2), 264-297. doi: 10.1177/0145445500242006

Kennerley, H. (2007). *Socratic questioning: A guide for therapists*. Oxford: Oxford Cognitive Therapy Centre.

Kirk, J. (1998). Cognitive-behavioural assessment. In K. Hawton, P. M. Salkovskis, J. Kirk & D. M. Clark (Eds.), *Cognitive behaviour therapy for psychiatric problems: A practical guide*. Oxford: Oxford University Press.

Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., . . . Keller, M. B. (2003). Therapeutic alliance in depression treatment: Controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology, 71*(6), 997-1006. doi: 10.1037/0022-006x.71.6.997

Krosnick, J., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 141-164). New York: Wiley.

Kuyken, W., Padesky, C. A., & Dudley, R. (2011). *Collaborative case conceptualisation: Working effectively with clients in cognitive-behavioral therapy*. New York: Guildford Press.

Kuyken, W., & Tsivrikos, D. (2009). Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychotherapy and Psychosomatics, 78*(1), 42-48. doi: 10.1159/000172619

Loewenthal, K. M. (2001). *An introduction to psychological tests and scales (2nd ed.)*. Philadelphia: Taylor & Francis Inc.

luborsky, L. (1994). Therapeutic alliance measures as predictors of future benefits of psychotherapy: Factors explaining the predictive success. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research and practice*. New York: Wiley.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*(3), 438-450. doi: 10.1037/0022-006x.68.3.438

Masters, E. R. (1974). The relationship between number of response categories and reliability of likert-type questionnaires. *Journal of Educational Measurement, 11*(1), 49-53. doi: 10.1111/j.1745-3984.1974.tb00970.x

Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology, 56*(6), 506-509. doi: 10.1037/h0033601

McGlinchey, J. B., & Dobson, K. S. (2003). Treatment integrity concerns in cognitive therapy for depression. *Journal of Cognitive Psychotherapy Winter, 17*(4), 299-318. doi: 10.1891/jcop.17.4.299.52543

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*(2), 73-84. doi: 10.1037/a0018121

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*, 63-67.

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

Neimeyer, R. A., & Feixas, G. (1990). The role of homework and skill acquisition in the outcome of group cognitive therapy for depression. *Behavior Therapy, 21*(3), 281-292. doi: 10.1016/S0005-7894(05)80331-4

Newman, C. F. (2013). *Core competencies in cogitive-behavioral therapy*. Sussex: Routledge.

Newstead, S. E., & Arnold, J. (1989). The effect of response format on ratings of teaching. *Educational and Psychological Measurement, 49*, 33-43.

Orlinsky, D. E., Klaus, G., & Parks, B. K. (1994). Process and outcome in psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 270 - 376). Oxford: John Wiley & Sons.

Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavioral Change* (pp. 307 - 390). New York: Wiley.

Overholser, J. C. (2010). Psychotherapy according to the socratic method: Integrating ancient philosophy with contemporary cognitive therapy. *Journal of Cognitive Psychotherapy, 24*(4), 354-363. doi: 10.1891/0889-8391.24.4.354

Padesky, C. A. (1993). *Socratic questioning: Changing minds of guided discovery?* Paper presented at the European Congress of Behavioural and Cognitive Therapies, London.

Pearsons, J. B. (1993). Case conceptualisation in cognitive-behavior therapy. In K. T. Kuehlwein & H. Rosen (Eds.), *Cognitive therapies in action: Evolving innovative practice* (pp. 33 - 53). San Fransisco: Jossey-Bass.

Persons, J., Burns, D., & Perloff, J. (1988). Predictors of dropout and outcome in cognitive therapy for depression in a private practice setting. *Cognitive Therapy and Research, 12*(6), 557-575. doi: 10.1007/bf01205010

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discrimination power and respondent preferences. *Acta Psychologica, 104*, 1-15. doi: 10.1016/S0001-6918(99)00050-5

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. doi: 10.1016/j.cpr.2010.03.004

Raue, P. J., & Goldfried, M. R. (1994). The therapeutic alliance in cogitive-behavior therapy. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research and practic*. New York: Wiley.

Rees, C. S., McEvoy, P., & Nathan, P. R. (2005). Relationship Between Homework Completion and Outcome in Cognitive Behaviour Therapy. *Cognitive Behaviour Therapy, 34*(4), 242-247. doi: 10.1080/16506070510011548

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95-103. doi: 10.1037/h0045357

Roth, A. D., Calder, F., & Pilling, S. (2013). *A competence framework for Child and Asolescent Mental Health Services*. London: Department of Health.

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognititve and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

Ryum, T., Stiles, T. C., Svartberg, M., & McCullough, L. (2010). The effects of therapist competence in assigning homework in cognitive therapy with cluster C personality disorders: Results from a randomized controlled trial. *Cognitive and Behavioral Practice, 17*(3), 283-289. doi: 10.1016/j.cbpra.2009.10.005

Saris, W., Revilla, M., Krosnick, J., & Shaeffer, E. (2010). Comparing questions with agree / disagree response options to questions with item-specific response options. *Survey research methods, 4*(1), 61-79.

Schwarz, N., Knauper, B., Hippler, H.-J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales numberic values may change the meaning of scale labels. *Public Opinion Quarterly, 55*(4), 570-582. doi: 10.1086/269282

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. doi: 10.1016/j.cpr.2008.09.008

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), 837-846. doi: 10.1037/0022-006x.67.6.837

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

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use (3rd ed.)*. Oxford: Oxford University Press.

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. doi: 10.1037/a0019631

Tee, J., & Kazantzis, N. (2011). Collaborative empiricism in cognitive therapy: A definition and theory for the relationship construct. *Clinical Psychology: Science and Practice, 18*(1), 47-61. doi: 10.1111/j.1468-2850.2010.01234.x

Townend, M., Iannetta, L., & Freeston, M. H. (2002). Clinical supervsion in practice: A survey of UK cognitive behavioural psychotherapists accredited by the BABCP. *Behavioural and Cognitive Psychotherapy, 30*(04), 485-500. doi: 10.1017/S1352465802004095

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. doi: 10.1023/B:COTR.0000021536.39173.66

Van Vaerenbergh, Y., & Thomas, T. D. (2012). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal of Public Opinion Research*. doi: 10.1093/ijpor/eds021

von Consbruch, K., Clark, D. M., & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: Psychometric properties of the cognitive therapy competence scale for social phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy, 40*(2), 149-161. doi: 10.1017/S1352465811000622

Waddington, L. (2002). The therapy relationship in cognitive therapy: A review. *Behavioural and Cognitive Psychotherapy, 30*(02), 179-191. doi: 10.1017/S1352465802002059

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

Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society, 31*(2), 135-138.

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), 620-630. doi: 10.1037/0022-006X.61.4.620

Weck, F., Bohn, C., Ginzburg, D. M., & Ulrich, S. (2011). Treatment integrity: Implementation, assessment, evaluation and correlations with outcome. *Verhaltenstherapie, 21*, 99-107. doi: 10.1159/000328840

Westbrook, D., Kennerley, H., & Kirk, J. (2007). *An introduction to Cognitive Behaviour Therapy: Skills and Applications*. London: Sage Publications Ltd.

Wildt, A. R., & Mazis, M. B. (1978). Determinants of scale response: Label versus position. *Journal of Marketing Research, 15*(2), 261-267. doi: 10.2307/3151256

Wright, J. H. (2006). *Learning cognitive-behavior therapy: An illustrated guide*. Arlington: American Psychiatric Publishing.

Wright, J. H., & Davis, D. (1994). The therapeutic relationship in cognitive-behavioral therapy: Patient perceptions and therapist responses. *Cognitive and Behavioral Practice, 1*(1), 25-45. doi: 10.1016/s1077-7229(05)80085-9

Appendices

Appendix 1: Feedback Form and Scale Manual (version 1)

Assessment of Core CBT Skills (ACCS)

USER MANUAL

Version 1- July 2013

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus and was conducted in affiliation with the University of Oxford's Department of Psychiatry and the Oxford Cognitive Therapy Centre (OCTC).

Contents Page

Domain 1 - Agenda Setting 190
1.1 Suitable Items 190
1.2. Range and Scope 190
1.3. Collaboration 190
Domain 2 - Effective Use of Time 190
2.1. Pace 190
2.2. Time Management 190
2.3. Maintained Focus 190
Domain 3 - Homework 190
3.1. Appropriate Content 190
3.2. Collaborative Planning 190
3.3. Patient Understanding 190
3.4. Learning Review 190
Domain 4 - Delivering CBT Interventions 190
4.1. Suitable Interventions 190
4.2. Rationale 190
4.3. Learning Review 190
Domain 5 - Measuring Change 190
5.1. Appropriate Measure 190
5.2. Implementation 190
Domain 6 - Conceptualisation 190
6.1. Coherent Strategy 190
6.2. Patient Theory Bridge 190
6.3. Shared Understanding 190
Domain 7 - Managing Therapeutic Relationship 190
7.1. Interpersonal style 190
7.2. Empathic Understanding 190
Domain 8 - Effective Two-way Communication 190
8.1. Patient Feedback 190
8.2. Reflective Summaries 190
Patient Difficulty Rating 190
Global Performance Rating 190
References 190

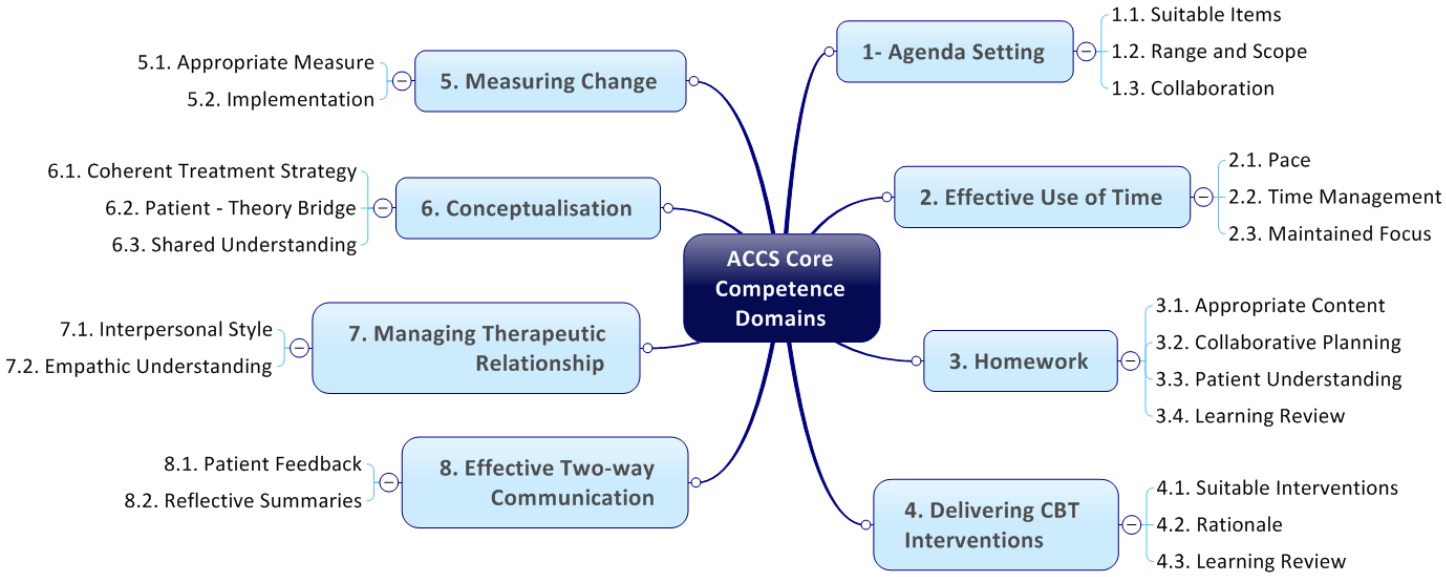
General Guidelines

Scale Scope

The Assessment of Core CBT Skills (ACCS) provides a framework for assessors to deliver formative and summative feedback about a therapist's performance within an observed treatment session. The scale assesses core general therapeutic and CBT-specific skills required to appropriately deliver individual cognitive-behavioural therapy to adults experiencing mental health problems. As the scale focuses on competences which are transdiagnostic (i.e. not specific to any one diagnosis) it can be used to assess competence across different disorder-specific treatment protocols. It is also worth noting that, as the scale is intended to assess a mid-treatment session, it does not include skills which, although important, do not occur in most active treatment sessions (e.g. goal setting, blue printing etc.).

Core competences were identified for inclusion in the scale on the basis of a review of relevant literature. In particular, the authors drew upon the Cognitive Therapy Scale (CTS, or Cognitive Therapy Rating Scale: CTRS, www.beckinstitute.org), the Cognitive Therapy Scale-Revised (CTS-R: Blackburn et al., 2001) and Roth and Pilling's (2007) competence framework.

As demonstrated in the figure below, the scale features eight competence domains, each of which includes two to four items:



There is widespread agreement across the field that collaborative empiricism and guided discovery are essential components of CBT which should underpin all aspects of CBT in order to promote active learning and help patients gain new perspectives. Given the pervasive and recurrent nature of these aspects of competence, they have not been included as independent competences within this scale. Instead, these aspects of competence have been embedded within the rating framework of other specific competences (e.g. conceptualisation, agenda setting, delivering CBT interventions etc.).

The scale provides a numerical total score which is designed to identify the therapist's competence across the session overall. Numerical scores are also provided for each competence domain and their respective sub-components in order to provide more detailed feedback as to the therapist's specific areas of strength and weakness. In addition, the feedback form includes space for assessors to include further qualitative feedback and to indicate what steps could be taken to further develop competence. Thus the scale can be used both to establish whether therapists have reached an adequate standard of competence (useful for accreditation, evaluating the impact of training, selecting and monitoring research trial therapists) and to provide detailed feedback to aid a therapist's development and progression (useful within supervision and training settings, for on-going personal development, and examining skill acquisition).

How to use the ACCS

This manual is intended to provide guidance for assessors as to how to make judgements about the quality or skilfulness of therapists' performance using the ACCS rating scale. It is recommended that assessors spend some time reading and familiarising themselves with the manual before conducting an assessment. Initially, the manual should be consulted on an item by item basis alongside completing the feedback form. Once the assessor becomes familiar with the scale, the manual can then be used as a frame of reference which can be consulted whenever questions or uncertainties arise. It is recommended that assessors periodically re-review the manual to prevent assessor-drift.

Step 1: Provide a numerical rating for each item

Each of the eight domains is broken down into two to four items, each of which is scored on a five-point scale. The overarching definitions of the five-point scale are as follows:

Performance Band	Generic Definition of Performance Band
1. Inadequate	• Therapist fails to include feature outlined. • Therapist demonstrates a significant absence of skill or an extremely inappropriate performance which is likely to have negative therapeutic consequences.
2. Poor	• Therapist's performance is somewhat appropriate. • However, major substantive problems or numerous minor problems are evident.
3. Satisfactory	• Therapist demonstrates a satisfactory degree of skill with no major problems. • However, minor problems or inconsistencies are evident in the therapist's performance.
4. Good	• Therapist consistently demonstrates good skill with evidence of only a few minor problems.
5. Excellent	• Therapist consistently demonstrates a very high level of skill with minimal or no problems.

Concrete examples of behaviours consistent with each scale point are provided for each sub-component individually. These 'exemplar therapist behaviours' are intended to provide guidance as to the type of performance which is consistent with each point on the rating scale.

The feedback form should be used to assign a rating for each item on the scale. All items within the domain can then be added to provide a total domain score and each total domain score can also be added to provide an overall session score.

Issues to consider-

- The exemplar behaviours are not able to provide an exhaustive list of possible performance indicators. Furthermore, ratings of competence will inevitably include a degree of subjectivity. However, in order to promote consensus between different assessors, it is important that assessors base their ratings on the item-specific exemplar therapist behaviours wherever possible.
- Assessors may find it helpful to make notes of about the therapist's performance that correspond with each item as they view the session. However, assessors should only score items after they have listened to the entire session.
- The 'halo effect' indicates that, when an assessor rates individual aspects of a therapist's competence, they are often unduly influenced by their overall impression of the therapists' performance (Cooper, 1981). As a result the therapist is provided with a uniformly high or low rating, rather than specific feedback as to their strengths and weaknesses. It is important for assessors to bear in mind that therapist's performance may well vary across domains- indeed a therapist can perform poorly in one competence domain and then do very well in another. It is therefore important to score each item independently of other items, and to avoid relying on an overall global impression.

Step 2: Provide a global performance rating

The scale also includes a global competence item which is rated using the generic five-point scale outlined above. This global rating should reflect the assessor's overall impression of the therapist's performance within the session as a whole. This global rating will normally correspond to domain ratings. However, this will not always be the case. The assessor may, for example, wish to provide a global rating which reflects the presence or absence of specific aspects of competence which, although important, have not been included within the core competence domains of this scale because they are not relevant across patients and/or treatment sessions.

Step 3: Rate patient difficulty:

Patient difficulty has not been incorporated within the scoring system as this may unduly confound the ratings. Furthermore, reserving the highest ratings for performance demonstrated in the face of patient difficulty may unfairly prevent those working with more straightforward cases from being awarded a high performance rating. However, assessors are asked to provide a patient difficulty rating. This is to identify the context within which the therapist's performance has been evaluated.

Step 4: Provide qualitative feedback:

After completing the numerical ratings, assessors are given the space to highlight areas of strength as well as learning needs. Assessor can draw on the exemplar therapist behaviours provided and the session material being assessed to provide examples of what made some aspects of the therapist's performance more successful than others. Where areas of weakness have been identified, assessors can suggest alternative, more skilful way of working within this domain and highlight ways in which the therapist could further develop their skills in this area. Such in-depth feedback is invaluable in aiding further development and progression.

Minimum Competence Score

A minimum score above which a therapist can be considered competent has not been provided for the ACCSS. This is because further research is required to establish an empirically proven 'tipping point' for competence using the ACCS. Furthermore, the level of competence and thus the score on the ACCS which is required may vary according to the context and purpose of the assessment. For example, when determining whether candidates are suitable for inclusion on a CBT training course, a lower level of performance may be acceptable, whereas those seeking accreditation may be required to demonstrate a higher level of performance.

Assessment Materials

The ACCS is designed to assess therapists' performance within an 'active' CBT treatment session. Assessments should therefore be made on the basis of a mid-treatment session which involves CBT intervention strategies, rather than an initial assessment session or end of treatment / blue-print session.

Ratings should be based upon source material which retains the most information possible, as this is more likely to provide the most accurate representation of the treatment session and thus the most accurate assessment of the therapists' competence (Waltz, Addis, Koerner, & Jacobson, 1993). As it is typically not possible to physically sit in on treatment sessions, rating are usually made by viewing recordings of a treatment session. Visual recordings provide the richest source of information and can therefore be viewed as the optimal source material, followed by audio recordings. Ratings should not be made on the basis of session transcripts or therapists' verbal or written reports of the session.

It is recommended that assessors base their ratings on a full treatment session, rather than segments of a session. This is because the reliability of individual item ratings is poorer when assessors view session segments rather than full sessions, with a number of aspects of competence not being assessed at all using this method (Weck, Bohn, Ginzburg, & Ulrich, 2011).

In addition to submitting a session recording, it is recommended that therapists provide a short written case summary. This should briefly outline key contextual information such as the stage of therapy, the patient's presenting problem, the session agenda, an outline of homework assignments, the main focus of the session, a brief summary of interventions carried out so far, a rationale for chosen treatment interventions, a treatment plan based on formulations and treatment targets/patients goals, and any relevant patient outcome data. Any materials completed in the session or for homework (e.g. questionnaires, diaries, thought records etc.) should also be submitted alongside the recording. Viewing this additional contextual information will help assessors make an informed rating of the therapist's performance.

Assessor Selection

It is unlikely that assessors who are themselves novice therapists would be able to provide detailed formative feedback about a therapist's specific areas of strengths and weaknesses. Indeed research suggests that therapist competence cannot be reliably rated by trained novices (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011). However, at present it is unclear what expertise are necessary to assess competence. As a minimum, it is recommended that assessments are carried out by therapists who have themselves received formal training or accreditation as a CBT therapist and thus can be considered competent therapists.

There is some evidence that assessors rate competence differently depending on their level of independence (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012a), although it is not known whose ratings are more accurate. There are a number of pros and cons involved with using both supervisors who have a relationship with the therapist and independent assessors and to rate a therapist's performance using the ACCS. Independent assessors are less likely to be influenced by social demands, relationship dynamics or information other than that obtained through viewing treatment sessions (e.g. prior competence, ability in other domains). On the other hand, supervisors may have access to a greater wealth of contextual information (e.g. treatment context, therapist's broader work) and be better placed to provide appropriate formative feedback. Those selecting assessors will therefore need to weigh up these pros and cons and make a decision based on pragmatics and the purpose of the assessment. If it is decided to use supervisors, one possibility for overcoming potential biases is to implement reliability checks by independent assessors.

Therapist Self-ratings

The scale can also be used by therapists to rate their own competence. As the ability to accurately rate one's own competence has been called into question, it is not recommended that self-ratings be used as a formal measure of therapist competence (Muse & McManus, 2013). However, self-ratings can be a useful learning tool. This is particularly the case when they are used in combination with assessor-ratings as this process of comparison can help to foster self-reflection and highlight on-going learning needs.

Note- self-ratings on the ACCS should <u>not be viewed by assessors prior to completing the ACCS</u> as this could bias the assessor's ratings.

Cautions

The ACCS is designed to assess whether a therapist has demonstrated the core generic and CBT specific skills required to deliver effective CBT within a given treatment session. Thus a number of aspects of competence are not assessed by the ACCS. First, the ACCS does not assess therapist's knowledge or understanding of CBT, aspects of competence which can instead be assessed using multiple choice questionnaires, essays, case reports or clinical vignettes (Muse & McManus, 2013). Second, as the ACCS focuses on 'intervention competence', broader professional skills (e.g. ethical practice, effective use of supervision) are not covered. Third, as the ACCS assesses core CBT skills evident during active, mid-treatment therapy sessions, it does not assess therapist's assessments or relapse-prevention skills. Hence it is recommended that, where the scale is used for summative assessment purposes, the ACCS should not be used as a stand-alone measure of competence. Instead the ACCS should form part of a multi-method competence assessment programme.

Although competence is often inferred from rating one or two isolated treatment sessions per therapist, recent studies suggest that a much higher sample of clinical work needs to be assessed in order to reliably assess therapist competence (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012b; Keen & Freeston, 2008). Furthermore, collecting multiple assessors' ratings of the same treatment session can increase reliability (Vallis, Shaw, & Dobson, 1986) and reduce the impact of the halo effect (Streiner & Norman, 2003). Thus in order to be confident of a reliable judgement as to a therapist's competence, multiple assessors would be used to assess a number of therapy sessions drawn across different patients (between 15 and 24 sessions rated by a single assessor, or 19 sessions rated by two assessors: Keen & Freeston, 2008). Although financial and time constraints often make this unfeasible, it is important to recognise the limitations involved in drawing conclusions about a therapist's competence and thus fitness to practice on the basis of a limited number of session ratings.

Domain 1 - Agenda Setting

Ability to collaboratively set a realistic, mutually agreed agenda which prioritises suitable, specific target problems and includes all appropriate routine items.

NB This item refers to **planning and setting** an appropriate agenda **not adhering to it** which should be rated under Item 2.3 'Maintained Focus'.

1.1 Suitable Items - exemplar therapist behaviours

1. Inadequate

- Therapist made no attempt to identify agenda items. Or therapist made some attempt to discuss how the session could be used (e.g. asked "how would you like to spend our time today?"), but did not outline any specific topics to be covered during the session.

2. Poor

- Items very vague, unclear or overly general (e.g. "my anxiety").
- Items not clearly relevant or very peripheral to patient's presenting problem or goals.
- Items very unlikely to be amenable to any within-session change.
- Items highly inappropriate given stage of therapy or patient's presentation.
- A number of relevant standard items omitted (e.g. mood check, setting homework, session feedback).

3. Satisfactory

- Items somewhat vague, unclear or overly general (e.g. "my anxiety at work").
- Items somewhat peripheral to patient's presenting problems or goals.
- Items somewhat unlikely to be amenable to significant within-session change.
- Items not ideally suited to stage of therapy, individual patient's presentation or treatment goals.
- One relevant standard item omitted (e.g. mood check, setting homework, session feedback).

4. Good

- Items were well-defined, specific and measurable (e.g. carry out a behavioural experiment or complete a thought record).
- Therapist clarified and specified items which were: amenable to within-session change, relevant to patient's presenting problem and goals, and appropriate given stage of therapy and patient's presentation.
- All relevant standard items included (e.g. mood check, setting homework, session feedback).

5. Excellent

- Therapist skilfully defined, clarified and specified items which were: clearly amenable to significant within-session change, ideally suited to patient's presenting problem and goals, and the most appropriate given stage of therapy and patient's presentation.
- All relevant standard items included (e.g. mood check, setting homework, session feedback).

1.2. Range and Scope - exemplar therapist behaviours

1. Inadequate

- Items far too numerous, large or overwhelming to be feasible / enable any within-session progress.
- Too few or insignificant items which could easily be covered in a very small time frame (e.g. reviewing a small piece of homework).

2. Poor

- Items too numerous, large or overwhelming to be feasible / enable much within-session progress.
- Too few or too insignificant items or items far too small to make effective use of time.

3. Satisfactory

- Range and scope of items slightly under or over optimistic, making it difficult to make effective use of time (e.g. unlikely that adequate time could be spent on all items).

4. Good

- Therapist prioritised appropriate number of items with suitable scope and significance, meaning it was likely that all items could be adequately covered within the session.

5. Excellent

- Therapist skilfully prioritised an optimal number of items with ideal scope and significance, making it highly likely that all items could be thoroughly covered in-session.

1.3. Collaboration - exemplar therapist behaviours

1. Inadequate

- Therapist made very little or no attempt to involve patient in setting agenda (e.g. simply told patient what they would be covering).
- Therapist provided no input to identifying agenda items where patient was not able to do so independently.
- If patient raised concerns these were completely ignored or disregarded by therapist.

2. Poor

- Therapist used an overly didactic style (e.g. appeared to come to session with a pre-prescribed agenda that was not open to sufficient discussion).
- Therapist did not attempt to share responsibility for setting agenda in a manner appropriate to stage of therapy or patient's presentation*.
- Therapist failed to identify patient's priorities or focused on areas that were not a priority for patient.
- If disagreements arose or patient raised concerns these were largely ignored or disregarded by therapist.

3. Satisfactory

- Therapist attempted to share responsibility for setting agenda with patient but either gave too much or too little responsibility given stage of therapy or patient's presentation (e.g. expecting patient to take the lead in very early sessions, not giving patient enough responsibility in later sessions).
- Therapist attempted to identify patient's session priorities but used complicated or overly vague questions which were not appropriate to stage of therapy or patient's presentation (e.g. "what shall we talk about today?").
- If disagreements arose or patient raised concerns these were given limited consideration by therapist (e.g. acknowledging concerns but failing to address them).

4. Good

- Therapist encouraged patient to share responsibility for identifying and prioritising items in a manner appropriate given stage of therapy and patient's presentation (e.g. evidence of responsibility being transferred to patient towards the end of treatment).
- Therapist used relevant questions to help patient identify and prioritise suitable items (e.g. "what do you think is most important to cover today? What would you like to spend most time on?").
- If disagreements arose or patient raised concerns these were acknowledged and given consideration.

5. Excellent

- Therapist engaged patient in a highly collaborative process and encourages them to share responsibility for identifying and prioritising items in a manner ideally suited to stage of therapy and patient's presentation.
- Therapist skilfully used relevant questions to help patient identify and prioritise suitable items.
- If disagreements arose or patient raised concerns these were openly acknowledged and fully explored by therapist and patient together.

* NB it may be appropriate for therapists to take more responsibility in initial sessions, but responsibility should be increasingly transferred to patient as therapy progresses, depending on patient's presentation.

Domain 2 - Effective Use of Time

Ability to structure, pace and manage time within the session in a manner which is appropriate to the therapeutic context and patient's capacity for learning and enables maintained focus on important issues.

2.1. **Pace -** exemplar therapist behaviours

1. Inadequate

- Pace was consistently either much too fast or much too slow for the therapeutic context and/or patient's optimal learning (e.g. therapist repeatedly belaboured a point long after patient had grasped the message, therapist's pace was so fast that important issues were rushed- making it difficult for patient to grasp or retain key concepts, therapist switched from topic to topic far too rapidly).

2. Poor

- Pace was frequently either too fast or too slow for the therapeutic context and/or patient's optimal learning (e.g. speed of therapist's pace meant important issues were somewhat rushed, making it questionable as to whether patient grasped or retained key concepts, therapist switched from topic to topic somewhat too rapidly without reaching conclusions).

3. Satisfactory

- Pace was largely suited to the therapeutic context and patient's learning capacity, though there were a number of times when the pace was somewhat too quick or slow for the therapeutic context and/or patient's optimal learning (e.g. therapist slightly overstressed a point, failed to slow the pace suitably where evidence of emotional or cognitive difficulties arose, moved patient along before they had fully understood concepts, switched topics before conclusions had been fully drawn).

4. Good

- For most of the session therapist set a pace which was well suited to therapeutic context and patient's learning capacity. However, there were one or two occasions when the pace was a little too quick or too slow to be optimally effective.

5. Excellent

- Therapist was responsive to and consistently matched the pace of the session to the therapeutic context and patient's learning capacity (e.g. gave more time and attention to issues where evidence of emotional or cognitive difficulties arose).

2.2. Time Management - exemplar therapist behaviours

1. Inadequate

- Inefficient, unbalanced allocation of time meant that a number of important issues were not covered at all (e.g. not enough time left to set homework at the end). These issues outstanding were not acknowledged or dealt with appropriately.
- Session started significantly late and/or ran very significantly over or under time without clinical justification (e.g. significant patient distress or expression of risk).

2. Poor

- Less than optimal use of time meant that some important issues were either not covered or were inadequately covered (e.g. insufficient time to review learning from a behavioural experiment or to plan how to take the work forward). These issues outstanding were not sufficiently acknowledged / rescheduled.
- Session started late and/or ran somewhat over or under time without clear clinical justification (e.g. significant patient distress or expression of risk).

3. Satisfactory

- Therapist's use of time was usually fairly efficient, meaning that most important issues were addressed. However, some unbalanced allocation of time meant that one or more important issues were still somewhat rushed and/or inadequately covered. Therapist made some attempt to acknowledge any issues outstanding.
- Session started and finished largely as scheduled or there was clear clinical justification for deviating from planned schedule (e.g. significant patient distress or expression of risk).

4. Good

- For most of session therapist's use of time was fairly well balanced and efficient, meaning all important issues were adequately covered. However, to make optimal use of time, it would have been necessary to spend a little less or a little more time on some issues. If issues were not covered, this was largely appropriate given the therapeutic context (e.g. it was decided to focus the entire session on one agenda item which was more distressing/significant than anticipated) and any issues outstanding were rescheduled for a different time or format (e.g. next session or homework).
- Session started and finished as scheduled or there was clear clinical justification for deviating from planned schedule (e.g. significant patient distress or expression of risk).

5. Excellent

- Consistently balanced and efficient allocation of time meant the most important issues were thoroughly covered. If issues were not covered, this was appropriate given the therapeutic context (e.g. it was decided to focus entire session on one issue which was more distressing/significant than anticipated), this decision was made collaboratively with the patient and issues outstanding were rescheduled for a different time or format (e.g. next session or homework).
- Session started and finished as scheduled or there was clear clinical justification for deviating from planned schedule (e.g. significant patient distress or expression of risk).

2.3. Maintained Focus - exemplar therapist behaviours

1. Inadequate

- Session lacked any clear focus and appeared aimless (e.g. patient dominated session completely and therapist simply followed patient's lead, making a very poor or no attempt to focus session back to appropriate topics).
- Overly rigid focus was unhelpful to patient (e.g. therapist completely disregarded patient's distress because it did not relate to an agenda item / fit with plan for session).
- Session lacked structure, with no clear introductory, middle or concluding phases (e.g. general unstructured 'chat' which did not appear to have any specified aims or reach any particular conclusions) or session was unduly regimented.

2. Poor

- Therapist repeatedly lost control over flow of dialogue, resulting in frequent repetitions, unproductive digressions or unjustified discussions of peripheral issues. Therapist either failed to address digressions or showed little skill in re-directing discussion to appropriate topics (e.g. abrupt, tactless).
- Therapist showed limited skill in maintaining appropriate focus, consequently parts of the session appeared aimless or unrelated to session priorities (e.g. therapist frequently switched topics without clear justification, session frequently and unnecessarily moved away from established agenda items).
- Therapist focused on established priorities in a rigid, inflexible or mechanistic manner (e.g. failed to acknowledge or respond to unanticipated issues).
- Session included some semblance of introductory, middle, and concluding phases, but these were not clear or discrete.

3. Satisfactory

- Therapist occasionally lost control over flow of dialogue, resulting in a number of repetitions, unproductive digressions or unnecessary discussions of peripheral issues. Although digressions were steered back to appropriate topics, therapist was slow to respond and/or did so with limited skill.
- At times the focus of the session moved unnecessarily away from established priorities (i.e. agenda items) or therapist focused on established priorities with little flexibility or responsiveness (e.g. to unanticipated issues or patient distress).
- Session included somewhat distinct introductory, middle, and concluding phases, but did not run smoothly between these.

4. Good

- Therapist maintained appropriate control over flow of dialogue throughout most of session. When repetitions, unproductive digressions or unnecessary discussions of peripheral issues occurred, these were steered back to an appropriate topic in a timely manner and with some degree of sensitivity and skill.
- Therapist largely maintained focus on established priorities (i.e. agenda items). However, if unanticipated issues arose, therapist showed appropriate flexibility (e.g. allowing the session to diverge from planned topics to make maximum therapeutic progress, or maintain client engagement/ therapeutic relationship).
- Session included clear introductory, middle, and concluding phases and ran smoothly between these.

5. Excellent

- Therapist showed highly skilled control over flow of dialogue throughout the session. Where repetitions, unproductive digressions or unnecessary discussions peripheral issues arose, these were quickly acknowledged, sensitively evaluated and steered back to an optimal focus (e.g. "We seem to be straying from our plan for the session- is this something that it is important that we focus on? Or would do you think it would be more useful to you for us to return the focus to [planned focus]?").
- Therapist consistently maintained focus on session priorities (i.e. agenda items). However, if unanticipated issues arose, therapist responded sensitively and flexibly.
- Session included clear introductory, middle, and concluding phases and flowed seamlessly between these.

Domain 3 - Homework

Ability to plan achievable homework which has clear and specific goals, is tailored to the individual patient's therapeutic context and is drawn directly from session material. To plan the homework in a collaborative manner which ensures the patient understands the purpose and content of the assignment and has the opportunity to discuss their reaction to it. To conduct a detailed review of previous homework in order to help the patient identify what they learned from the experience.

3.1. **Appropriate Content -** exemplar therapist behaviours

1. Inadequate

- Therapist failed to plan homework arising from the session despite clear opportunities to do so.
- Homework was highly inappropriate to patient's therapeutic context* (e.g. expecting patient to carry out a highly feared exposure after first session or planned a difficult or large assignment for a patient with low motivation).
- Content, practicalities and goals for homework were very vague and unclear (e.g. experiment with going out).

2. Poor

- Homework poorly suited to patient's therapeutic context (e.g. too threatening to be achievable, 'standard' homework used without necessary adaptation)
- Poorly specified content, practicalities and goals (e.g. go into a specific situation and see if I get anxious).
- Homework not clearly related to issues (e.g. problems, cognitions, behaviours) dealt with in-session, therefore restricting opportunities for relevant information gathering or generalising concepts and skills learnt into daily life.
- Limited skill in considering impact of likely outcomes from homework (e.g. significant risk patient would fail to benefit or would have predictions confirmed in a way that was counter therapeutic).

3. Satisfactory

- Homework was somewhat suited to patient's therapeutic context but required fine tuning (e.g. somewhat overly ambitious, easy, challenging or complicated for patient).
- Content, practicalities and goals were outlined, though some specific details were lacking (e.g. what, where, when, how, outcome evaluation).
- Homework somewhat related to issues (e.g. problems, cognitions, behaviours) dealt with in-session, therefore enabling some information gathering and/or generalising concepts and skills learnt into daily life. However, opportunities for more central homework tasks or further learning were missed.
- Therapist gave limited consideration to impact of likely outcomes of homework.

4. Good

- Homework was well suited to individual patient's therapeutic context (e.g. amount and difficulty of homework well suited to patient's motivation and cognitive functioning, homework represented a moderate and achievable challenge for patient).
- Content, practicalities and goals were clearly specified (e.g. what, how, where, when, frequency, duration, outcome measurement).
- Homework was drawn from important issues (e.g. problems, cognitions, behaviours) dealt with in-session, therefore enabling opportunity for relevant information gathering and/or generalising concepts and skills learnt into daily life.

5. Excellent

- Custom tailoring of homework meant it was optimally suited to individual patient's therapeutic context.
- Therapist skilfully worked with patient to agree a clear and detailed plan of exactly what homework involved (e.g. what, how, where, when, frequency, duration, outcome measurement).
- Homework was skilfully drawn from key issues (e.g. problems, cognitions, behaviours) dealt with in-session, therefore enabling clear opportunities for relevant information gathering and/or generalising central concepts and skills learnt into daily life.
- Therapist thoroughly considered the impact of likely outcomes of homework (e.g. set homework up as 'no lose' or worked with patient to make a plan for dealing with potential negative outcomes).

* Therapeutic context = patient's presenting problems, goals, stage of treatment, formulation & presentation.

3.2. Collaborative Planning - exemplar therapist behaviours

1. Inadequate

- Therapist made very little or no attempt to involve patient in devising homework (e.g. simply provided homework to complete without input from patient).
- Therapist provided no input to homework planning where patient was not able to devise appropriate homework independently.

2. Poor

- Therapist used a somewhat overly didactic style (e.g. appeared to come to session with prescribed homework that was not open to sufficient discussion).
- Therapist did not attempt to share responsibility for devising homework in a manner appropriate to stage of therapy or patient's presentation*.

3. Satisfactory

- Therapist attempted to share responsibility for setting homework with patient but either gave too much or too little responsibility given stage of therapy or patient's presentation (e.g. expecting patient to take the lead in very early sessions, not giving patient enough responsibility in later sessions).

4. Good

- Homework was jointly designed and planned.
- Therapist encouraged patient to share responsibility for devising homework in a manner which was suited to stage of therapy and patient's presentation.

5. Excellent

- Homework was jointly designed and planned.
- Therapist engages patient in a highly collaborative process and encourages them to share responsibility for devising homework in a manner which was ideally suited to stage of therapy and patient's presentation (e.g. asked "What ideas do you have about any ways in which you could take this work forward in the next week?" "What do you think would be important to do to continue to progress work on this issue?").

* NB it may be appropriate for therapists to take more responsibility for devising homework in initial sessions, but responsibility should be increasingly transferred to patients as therapy progresses, depending on patient presentation.

3.3. Patient Understanding - exemplar therapist behaviours

1. Inadequate

- Therapist made very little or no attempt to facilitate patient's understanding of the purpose and content of homework (e.g. failed to develop any rationale for importance/potential benefits).
- Therapist did not check patient's understanding of or reaction to homework.
- If patient raised reservations or potential obstacles, these were completely ignored or disregarded.

2. Poor

- Therapist attempted to facilitate patient's understanding of purpose and content of homework with very limited skill (e.g. developed a limited or unclear rationale for importance/potential benefits).
- Therapist checked patient's understanding of and reaction to homework with very limited skill (e.g. asked vague questions such as 'is that OK?'). Or, if patient raised reservations or potential obstacles, these were these were largely ignored or disregarded.

3. Satisfactory

- Therapist facilitated patient's understanding of purpose and content of homework with limited skill (e.g. developed a rationale for homework but did not relate this to patient's presenting problems and treatment goals).
- Therapist checked patient's understanding of and reaction to homework in a limited fashion (e.g. asked limited pertinent questions, such as "does this seem manageable?", "is the homework clear?").
- If patient raised reservations or potential obstacles, therapist gave these limited consideration (e.g. acknowledging an obstacle but failing to address it).

4. Good

- Therapist facilitated patient's understanding of purpose and content of homework with some degree of skill (e.g. developed a rationale for importance and potential benefits of homework which aligned with patient's presenting problems and treatment goals).
- Therapist used appropriate questions to check patient's understanding of and reaction to homework (e.g. asked "how does the idea of doing this strike you?", "how do you feel about attempting this task?") and to identify potential obstacles (e.g. asked "What might get in the way of you completing this task?").
- If patient raised reservations or potential obstacles these were acknowledged and given adequate consideration (e.g. asked "How might we overcome that?").

5. Excellent

- Therapist skilfully facilitated patient's understanding of the purpose and content of homework (e.g. developed a clear rationale for the importance and potential benefits of homework which clearly aligned with patient's presenting problems and treatment goals) and checked patient's understanding of this by asking skilful question such as "can you explain to me why it is important that you do this?", "can you tell me why it is useful for you to do this?").
- If relevant, therapist used guided imagery techniques or worked through examples with patient in-session to provide a clear picture of what is expected and highlight any potential difficulties or questions (e.g. completing a thought record or behavioural experiment in-session before expecting patient to carry out for homework).
- Therapist skilfully and thoroughly checked patient's understanding of and reaction to homework.
- Therapist worked with the patient to identify any doubts, reservations or potential obstacles and (where possible) resolved these issues (e.g. breaking homework into smaller tasks if patient felt overwhelmed, discussing advantages and disadvantages of homework).

3.4. Learning Review - exemplar therapist behaviours

1. Inadequate

- Therapist failed to review or conducted extremely brief or cursory homework review which did not help patient identify learning.
- Therapist responded to completed or uncompleted homework in an unhelpful or negative manner (e.g. scolding, judgemental, dismissed homework as unimportant).

2. Poor

- Therapist conducted an inadequate, rushed or superficial homework review which was of limited use in helping patient identify learning (e.g. asked "how did you get on with the homework?" with no further follow up).
- If patient did not complete, or only partially completed homework, therapist made little or no attempt to reflect on why this occurred and did not identify ways to overcome future 'blocks' to completion or use the opportunity to facilitate broader learning.
- Completed homework was given limited recognition and/or therapist responded to non-completion with limited sensitivity and understanding.

3. Satisfactory

- Therapist adequately reviewed most previous homework assignments and attempted to identify learning, but with limited skill (e.g. did not facilitate patient reaching their own conclusions).
- If patient did not complete, or partially completed homework, therapist made some attempt to identify reasons for this and to identify ways to overcome future 'blocks' to completion but with limited skill (e.g. failed to identify some emotional, cognitive or practical difficulties, overly didactic manner).
- Therapist made limited attempt to highlight broader learning implications from completing (or not completing) homework (e.g. to draw conclusions within the context of patient's conceptualisation and treatment goals).
- Therapist responded completed homework with somewhat appropriate encouragement and praise and to non-completion in a generally supportive and sensitive manner.

4. Good

- Therapist reviewed all previous homework in order to help patient identify learning which was useful and relevant to issues being addressed (e.g. examined outcomes in relation to prior expectations and predictions).
- If patient did not complete, or partially completed homework, therapist skilfully identified relevant emotional, cognitive or practical reasons for non-completion and discussed ways to overcome future 'blocks' to completion (e.g. using a calendar for a patient who has difficulty remembering to complete homework, breaking homework down into less threatening components).
- Therapist highlighted broader learning implications from completing (or not completing) homework and linked this learning to patient's conceptualisation and treatment goals (e.g. identified what it tells patient about how best to understand or make progress with their difficulties).
- Therapist responded to completed homework with appropriate encouragement and praise and to non-completion in a supportive and sensitive manner.

5. Excellent

- Therapist skilfully engaged patient in a highly collaborative, thorough and comprehensive review of all previous homework in order to help patient identify learning which was useful and relevant to issues being addressed.
- If patient did not complete, or partially completed homework, therapist worked skilfully with the patient to identify all relevant emotional, cognitive or practical reasons for non-completion and discussed relevant ways to overcome future 'blocks' to completion.
- Therapist highlighted broader learning implications from completing (or not completing) homework and helped the patient to link this learning to their conceptualisation and treatment goals (e.g. identified what it tells patient about how best to understand or make progress with their difficulties).
- Therapist responded to completed homework with optimal levels of encouragement, enthusiasm and praise and to non-completion in a highly supportive, sensitive and open manner.

Domain 4 - Delivering CBT Interventions

Ability to select suitable evidence-based cognitive-behavioural interventions to target clearly specified key cognition(s) or maintaining factors, skilfully apply and review the impact of interventions in a manner which ensures that the aim of the intervention is clear to, and understood by, the patient and that helps the patient identify what they learned from the experience.

NB A number of factors other than therapist competence can influence whether an intervention is successful in bringing about change. Please, therefore, focus on the quality of the cognitive-behavioural interventions and not whether change actually occurred.

4.1. **Suitable Interventions** - exemplar therapist behaviours

1. Inadequate

- Therapist did not select any CBT interventions to target cognition(s)* linked to distressing emotions or associated maintaining factors, despite clear opportunity to do so.
- Interventions highly inappropriate to patient's therapeutic context** (e.g. therapist challenged core beliefs too early in therapy, intervention far too complex for current capacity of the patient).
- Therapist's selection of interventions was not guided by evidence-based CBT practice***.

2. Poor

- Therapist missed a number of valuable opportunities where CBT interventions could have usefully been administered.
- Therapist selected interventions to target irrelevant or peripheral cognition(s) and/or associated maintaining factors which were of limited therapeutic benefit (e.g. targets were not linked to distressing emotions).
- Interventions poorly suited to patient's therapeutic context (e.g. too threatening to be achievable, 'standard' intervention used without necessary adaptation).
- Therapist's selection of interventions was guided by somewhat inappropriate evidence-based CBT practice (e.g. a treatment protocol which was only vaguely related to the patient's presentation) or therapist seemed to significantly misinterpret or misapply appropriate evidence-based CBT practice.

3. Satisfactory

- Therapist selected interventions to target somewhat relevant cognition(s) and/or associated maintaining factors but targets were not clearly specified (e.g. were somewhat vague/ambiguous).
- Therapist missed one or two opportunities where interventions could have usefully been administered.
- Interventions somewhat suited to patient's therapeutic context but required fine tuning (e.g. somewhat overly ambitious, easy, challenging or complicated for patient).

4. Good

- Therapist selected interventions to target clearly specified, relevant cognition(s) and/or associated maintaining factors.
- Interventions well suited to individual patient's therapeutic context (e.g. complexity/difficulty well suited to patient's motivation and cognitive functioning, represented a moderate and achievable challenge to patient at this time).
- Therapist's selection of interventions was accurately guided by appropriate evidence-based practice.

5. Excellent

- Therapist selected interventions to target key cognition(s) which were clearly linked to distressing emotions and/or associated maintaining factors. These targets were skilfully defined, clarified and specified.
- Appropriate timing, grading & tailoring of interventions meant they were optimally suited to individual patient's therapeutic context.
- Therapist's selection of interventions was consistently and accurately guided by the most appropriate evidence-based practice.

* Target cognitions = thoughts, images, assumptions, beliefs.

** Patient's therapeutic context = patient's presenting problem, current capacity of the patient, stage of treatment etc.

*** Evidence-based CBT practice = theoretical frameworks, treatment packages/protocols.

4.2. Rationale - exemplar therapist behaviours

1. Inadequate

- Therapist employed interventions without an adequate and explicit rationale.
- Therapist did not check patient's understanding of or reaction to interventions.
- If patient did not understand or was sceptical about the rationale for interventions or highlighted questions or potential obstacles, these were completely ignored or disregarded or therapist responded in a negative manner (e.g. defensive, hostile).

2. Poor

- Therapist developed an incomplete or unclear rationale for completing interventions.
- Therapist checked patient's understanding of and reaction to interventions with very limited skill (e.g. asked vague questions such as 'is that OK?').
- If patient did not understand or was sceptical about the rationale for interventions or highlighted questions or potential obstacles, these were these were largely ignored or disregarded.

3. Satisfactory

- Therapist developed a somewhat suitable rationale for completing interventions, though it was somewhat vague or confusing.
- Therapist checked patient's understanding of and reaction to interventions in a limited fashion.
- If patient did not understand or was sceptical about the rationale for interventions or highlighted questions or potential obstacles, therapist gave these limited consideration (e.g. acknowledging a question but failing to answer it).

4. Good

- Therapist provided a clear explanation of the rationale underpinning interventions.
- Therapist used appropriate questions to check patient's understanding of and reaction to interventions (e.g. asked "how does the idea of doing this strike you?", "how do you feel about attempting this task?") and to identify potential obstacles (e.g. asked "What might get in the way of completing this task?").
- If patient did not understand or was sceptical about the rationale for interventions or highlighted questions or potential obstacles, these were acknowledged and given adequate consideration.

5. Excellent

- Therapist clearly and succinctly conveyed an excellent explanation of the rationale underpinning interventions.
- Therapist skilfully and thoroughly checked patient's understanding of and reaction to interventions.
- If patient did not understand or was sceptical about the rationale for interventions or highlighted questions or potential obstacles, therapist responded openly and (where possible) resolved these issues (e.g. clarified any ambiguities).

4.3. **Learning Review** - exemplar therapist behaviours

1. Inadequate

- Therapist failed to review or conducted extremely brief or cursory review of the results of interventions which did not help patient identify learning.

2. Poor

- Therapist conducted an inadequate, rushed or superficial review of the results of interventions which was of limited use in helping patient identify learning.

3. Satisfactory

- Therapist adequately reviewed the results of interventions and attempted to identify learning, but with limited skill (e.g. did not facilitate patient reaching their own conclusions, struggled to relate the results back to the cognitions/maintaining factors they were designed to target, brushed over negative results).
- Therapist made a limited attempt to highlight broader learning implications from interventions (e.g. to draw conclusions within the context of patient's conceptualisation and treatment goals).

4. Good

- Therapist helped patient to review the results of interventions (whether positive or negative) and to relate the results back to the cognitions/maintaining factors they were designed to target in order to identify the impact it had on their thinking or behaviour.
- Therapist highlighted broader learning implications from interventions and helped patient link this learning to patient's conceptualisation and treatment goals (e.g. identified what it tells patient about how best to understand or make progress with their difficulties).

5. Excellent

- Therapist skilfully engaged patient in a highly collaborative, thorough and comprehensive review of the results of interventions (whether positive or negative) and seamlessly helped patient relate the results back to the cognitions/maintaining factors they were designed to target in order to help the patient re-evaluate previous conclusions or construct new ideas.
- Therapist skilfully worked with patient to help them identify broader learning implications from interventions and linked this learning to patient's conceptualisation and treatment goals.

Domain 5 - Measuring Change

Ability to select appropriate, clinically relevant methods for measuring change in symptoms and associated features (beliefs, behaviours, feelings) and movement towards goals. Ability to administer these methods at suitable time points across and within session and to skilfully review, interpret and respond to the information gleaned.

NB

A range of methods can usefully be employed to measure change, these include standardised questionnaires, diaries, frequency counts, ratings of duration of event or experience, self-ratings of emotions and cognitions etc. These methods should be employed across sessions to measure change in symptoms and movement towards goals over the course of treatment as a whole. However, these methods should also be employed within treatment sessions in order to measure the impact of cognitive-behavioural interventions completed within session and for homework (e.g. belief ratings pre/post intervention).

In addition to viewing session recording, it is advisable to review any relevant written in-session materials in order to make an informed rating on this item (e.g. completed questionnaires, diaries, belief ratings, treatment goals, etc.).

5.1. **Appropriate Measure** - exemplar therapist behaviours

1. Inadequate

- Therapist neglected to measure change in symptoms, associated features or movement towards goals.
- Therapist selected standardised measures with inadequate reliability, validity or usability.
- Therapist created inadequate idiosyncratic measures with poor usability (e.g. too complicated for patient to complete), completely unsuitable response scales (e.g. yes/no to measure symptom change), or undefined targets (e.g. rate your 'feelings').
- Measures were irrelevant and of no clinical use (e.g. depression questionnaire for anxiety patient, belief ratings of irrelevant cognitions, frequency of irrelevant behaviours).

2. Poor

- Therapist selected standardised measures with poor validity, reliability or usability.
- Therapist created idiosyncratic measures which were difficult to use, had inappropriate response scales or had poorly defined targets.
- Measures were of limited relevance and clinical use (e.g. general anxiety questionnaire for panic disorder patient, belief ratings of peripheral cognitions).

3. Satisfactory

- Therapist selected standardised measures with adequate reliability, validity and usability, though better measures were available.
- Therapist created usable idiosyncratic measures with appropriate response scales and clear targets (e.g. 0-100 emotional intensity ratings, 0-10 problem severity ratings).
- Measures were somewhat relevant and clinically useful (e.g. belief ratings for related cognitions, rating severity of relevant problem).

4. Good

- Therapist selected reliable, valid and usable standardised measures.
- Therapist created idiosyncratic measures with clearly defined targets and appropriate response scales (e.g. 0-100 rating of operationalised behaviour).
- Measures were relevant and clinically useful (e.g. panic cognitions questionnaire for panic disorder patient, belief ratings for important cognitions).

5. Excellent

- Therapist selected the most reliable, valid and usable standardised measures available.
- Therapist created well-constructed idiosyncratic measures with clearly specified and defined targets and the most appropriate response scales.
- Measures were highly relevant and served an important clinical function (e.g. belief ratings for key cognitions, diary provided useful insight into relationship between different aspects of patient's problems).

5.2. Implementation - exemplar therapist behaviours

1. Inadequate

- Schedule for the use of measures was random, haphazard, highly inappropriate or not clinically useful (e.g. no baseline function to measure change against [e.g. thought record without pre-intervention belief rating], patient overly burdened by completing measures too frequently, progress not regularly checked).
- Therapist failed to review or conducted extremely brief or cursory review of completed measures which did not provide useful clinical information (e.g. didn't compare post-intervention scores with pre-intervention scores, lack of understanding of completed measures).
- Therapist made little or no attempt to identify or explore barriers (e.g. difficulty understanding or completing measures, doubts about the purpose of monitoring) or responded to barriers in an unhelpful or negative manner (e.g. scolded patient for incorrectly completing measure).

2. Poor

- Therapist implemented a monitoring schedule which was unsuitable or of little clinical use (e.g. measures not used regularly enough or used more frequently than necessary, measures implemented at inappropriate time points).
- Therapist conducted an inadequate, rushed or superficial review of completed measures which provided little useful clinical information (e.g. therapist showed a limited ability to interpret completed measures).
- Therapist made some attempt to identify barriers to monitoring and responded to these in a somewhat understanding manner. However, therapist made little or no attempt to explore or overcome barriers.

3. Satisfactory

- Therapist implemented a somewhat appropriate monitoring schedule which served a limited clinical purpose (e.g. provided some insight into on-going change or impact of interventions).
- Therapist adequately reviewed completed measures attempted to use this information for clinical purposes, but with limited skill (e.g. overly didactic style, struggled to relate the results to the individual patient).
- Therapist identified barriers to monitoring, responded in a supportive manner and made some attempt to explore or overcome them.

4. Good

- Therapist implemented an appropriate monitoring schedule which served a useful clinical purpose (e.g. provided useful insight into changes in patient's symptoms).
- Therapist reviewed completed measures and used information gleaned for relevant clinical purposes (e.g. exploring improvements or deteriorations to better understand maintaining factors, giving patient feedback about progress to encourage motivation, making decisions about where best to focus therapeutic efforts).
- Therapist used measures to encourage patient to engage in self-monitoring.
- Therapist identified, explored and discussed ways to overcome barriers to monitoring with some degree of sensitivity and skill.

5. Excellent

- Therapist systematically implemented a highly appropriate monitoring schedule which served an important clinical function (e.g. provided clear and relevant feedback as to effectiveness of interventions).
- Therapist thoroughly and comprehensively reviewed completed measures, skilfully viewed the results in the context of the individual patient and used them for relevant clinical purposes.
- Therapist used measures to help patient become an active, collaborative participant in their own therapy by engaging in self-monitoring.
- Therapist skilfully, sensitively and thoroughly identified and explored barriers to monitoring. Where possible therapist worked with patient to resolve issues (e.g. working through measure with patient, modifying or simplifying measures).

Domain 6 - Conceptualisation

Ability to develop a mutually agreed conceptualisation (or 'formulation') which provides a bridge between CBT theory and the individual patient and offers a coherent, meaningful and relevant account of the key features of the patient's problems. Ability to draw upon the conceptualisation in order to facilitate understanding for both the patient and therapist and to develop a coherent treatment strategy.

NB

An initial conceptualisation is typically developed in the early phase of treatment as part of the assessment process. However, a conceptualisation should not simply precede treatment but should evolve throughout therapy as further clinical information arises and recurrent patterns and themes emerge. This information will need to be added to the conceptualisation or may require the conceptualisation to be modified or revised. Additionally, the therapist should draw upon and refer to the conceptualisation throughout treatment to help the patient understand their problems within a CBT framework and to highlight potential treatment strategies. Thus a therapist's ability to conceptualise a patient's problems should be evident not only in the initial assessment phase but throughout the course of treatment.

In addition to viewing the session recording, it is advisable to review any written conceptualisations or diagrams which were used by the therapist in order to check that the therapist's treatment choices are in line with these conceptualisations.

6.1. **Coherent Strategy -** exemplar therapist behaviours

1. Inadequate

- Therapist did not draw upon the formulation in order to select interventions. Consequently the session lacked an overall strategy for bringing about therapeutic change (e.g. session consisted of a set of unrelated interventions, therapist's use of interventions was disjointed, haphazard or erratic, therapist lacked a clear rationale for selecting interventions, therapist appeared to select interventions in a hit-and-miss fashion).
- Therapist made little or no attempt to plan any session work or further homework to build upon completed interventions, despite clear opportunity to do so.

2. Poor

- Therapist made some attempt to draw upon the formulation in order to select interventions but with limited skill. Consequently the overarching strategy for bringing about therapeutic change lacked coherence and/or showed very little promise in bringing about therapeutic change (e.g. interventions were tenuously linked to treatment targets identified in conceptualisation, treatment strategy was vague, unclear or disjointed).
- Therapist inappropriately planned further session work or homework to build upon completed interventions (e.g. further assignments were very tenuously linked) or did so with limited skill.

3. Satisfactory

- Selected interventions formed part of a somewhat coherent therapeutic strategy that showed some promise in bringing about therapeutic change and followed logically from the agreed conceptualisation of the patient's problems.
- Therapist planned appropriate further therapy tasks (in session or for homework) which were somewhat linked to completed interventions.

4. Good

- Conceptualisation was used as a platform for the selection of a generally coherent therapeutic strategy which showed reasonable promise in bringing about therapeutic change.
- Where appropriate, further therapy tasks (in session or for homework) were planned in a way which built upon completed interventions (e.g. developing a behavioural experiment relating to a belief identified in a thought record).

5. Excellent

- Conceptualisation was clearly and consistently used as a platform for the selection of a logical, coherent and unified treatment strategy which showed a great deal of promise in bringing about therapeutic change.
- Where appropriate, therapist worked with the patient to plan further therapy tasks (in session or for homework) which clearly built upon learning from completed interventions and enabled continued momentum.

6.2. Patient Theory Bridge - exemplar therapist behaviours

1. Inadequate

- Therapist did not use a CBT model* to inform conceptualisation, used a highly inappropriate model (e.g. appeared to force patient into a model) or inaccurately reproduced a relevant CBT model.
- Conceptualisation included information which had no direct relevance to patient's problems.
- Conceptualisation failed to incorporate key features of individual patient's presentation** and/or to explain how patient's difficulties developed or what is currently maintaining them, despite opportunity to do so.
- Conceptualisation was unnecessarily complex or overly simplistic given stage of treatment or patient's ability to understand (e.g. therapist introduced complex model of OCD very early in treatment).

2. Poor

- Conceptualisation was informed by a CBT model which was somewhat appropriate to patient. However, therapist failed to link model to individual's presentation, resulting in a vague or abstract conceptualisation.
- Conceptualisation included information which was of limited relevance to patient's problems.
- Conceptualisation incorporated some key features of patient's presentation but failed to show how these features related to each other / the problem in a meaningful way and/or provided a poor explanation of how these difficulties developed and what is currently maintaining them.
- Conceptualisation was somewhat too complex or simplistic given stage of treatment or patient's ability to understand.

3. Satisfactory

- Therapist utilised an appropriate CBT model and attempted to link this model to patient with limited skill (e.g. lack of personalised, concrete and specific examples).
- Conceptualisation incorporated most key features of patient's presentation but provided an unclear picture of how these features related to each other / the problem and/or provided a limited explanation of how the difficulties developed and what is currently maintaining them.
- Complexity of conceptualisation was somewhat appropriate, though a more straightforward or detailed conceptualisation would have been better suited to stage of treatment and patient's ability to understand.

4. Good

- Therapist flexibly utilised an appropriate CBT model to provide a personalised account of patient's problems.
- Conceptualisation drew together key features of patient's presentation into a well-defined overall picture which provided a sound explanation of how different components related to each other / the problem and how patient's difficulties developed / what is currently maintaining them in a way the patient could understand.
- Complexity of conceptualisation was well suited to stage of treatment and patient's ability to understand.

5. Excellent

- Therapist flexibly utilised the most appropriate CBT model to provide a nuanced and highly personalised account of patient's unique experiences.
- Conceptualisation seamlessly drew together key features of patient's presentation into a concise, comprehensive and coherent overall picture which eloquently demonstrated how different components related to each other / the problem in a meaningful way. Conceptualisation provided a full and coherent explanation of how patient's difficulties developed and what is currently maintaining them in a way that was understandable and meaningful to patient.
- Conceptualisation integrated relevant strength and resilience factors which could be used to promote change.
- The complexity of conceptualisation was optimal to stage of treatment and patient's ability to understand.

** CBT model = relevant disorder-specific model, vicious cycles etc.

* Key features of patient's presentation = behaviours, specific cognitions, emotions, physiological experiences, environmental factors & interpersonal issues.

6.3. Shared Understanding - exemplar therapist behaviours

1. Inadequate

- Therapist made very little or no attempt to involve patient in conceptualising their problems (e.g. did not elicit or value patient's ideas) or inappropriately followed patient's lead in developing conceptualisation.
- Therapist did not adequately check patient's understanding of and agreement with conceptualisation (e.g. simply showed patient conceptualisation). Or, if patient raised doubts or held different opinions, therapist largely ignored or disregarded these.
- Therapist made very little or no attempt to use conceptualisation to facilitate patient's understanding of their problems within a CBT framework (e.g. did not share conceptualisation with patient) or drew upon conceptualisation in a way that was counter therapeutic (e.g. told patient how their problem worked).

2. Poor

- Therapist took an overly didactic, leading role in conceptualisation (e.g. appeared to come to session with a pre-planned idea which was not open to sufficient discussion).
- Therapist checked patient's understanding of and agreement with conceptualisation with very limited skill (e.g. asked vague questions such as 'is that OK?') and/or if patient raised doubts or held different opinions, therapist gave these limited consideration (e.g. acknowledged but failed to explore them).
- Therapist made vague, unclear or limited references to conceptualisation in an attempt to facilitate patient's understanding of their problems within a CBT framework.

3. Satisfactory

- Therapist attempted to elicit patient's ideas and involve them in conceptualisation but there was some degree of imbalance (e.g. therapist paid minimal attention to patient's ideas).
- Therapist checked patient's understanding of and agreement with conceptualisation in a limited fashion. If patient raised doubts or held different opinions, therapist made some attempt to explore these.
- Therapist drew on conceptualisation in order to help patient gain some understanding of their problems within a CBT framework, though some opportunities to facilitate understanding were missed.

4. Good

- Therapist worked effectively with patient to jointly conceptualise their problems (e.g. elicited, attended to and incorporated patient's ideas).
- Therapist thoroughly checked patient's understanding of and agreement with conceptualisation and, if patient raised doubts or held different opinions, these were explored and incorporated.
- Therapist drew upon conceptualisation in a way which facilitated patient's understanding of the history, triggers and maintaining features of their problems within a CBT framework and promoted understanding of the CBT rationale for treatment.

5. Excellent

- Clear and consistent evidence of highly interactive, skilful collaboration between therapist and patient in conceptualising patient's problems.
- Therapist skilfully and thoroughly checked patient's understanding of and agreement with conceptualisation (e.g. asked patient to explain conceptualisation). Therapist welcomed any doubts or differences of opinion and worked with patient to resolve these in order to reach a mutually agreed conceptualisation.
- Therapist skilfully drew upon relevant specific examples in conceptualisation to effectively facilitate patient's understanding of the history, triggers and maintaining features of their problems within a CBT framework and promote understanding of the CBT rationale for treatment.

Domain 7 - Managing Therapeutic Relationship

Ability to foster a positive therapeutic alliance by (i) accurately grasping the content and emotional tone of the patient's understanding of themselves and the world around them and sensitively and appropriately conveying this understanding and (ii) embodying an excellent interpersonal style which is congruent with the therapeutic context and patient's needs.

7.1. **Interpersonal style -** exemplar therapist behaviours

1. Inadequate

- Therapist consistently demonstrated a negative interpersonal style (e.g. impatient, aloof, insincere, patronising, distant, distracted, cold, uncaring, intimidating, reprimanding, hostile, demeaning etc.). This style seemed to be destructive to the therapeutic alliance (e.g. caused patient to disengage, become distrustful or hostile, made patent unduly anxious and inhibited).

2. Poor

- Therapist demonstrated a somewhat undesirable interpersonal style (e.g. slightly aloof or distant, limited warmth and sincerity etc.) which had a detrimental impact on the therapeutic relationship (e.g. seemed to cause patient to lose trust in or respect for therapist).
- Therapist's interpersonal style was not appropriate to the therapeutic context* (e.g. therapist praised patient excessively, or demonstrated high level of warmth to a highly avoidant or socially anxious patient thus making them feel uncomfortable).

3. Satisfactory

- Therapist generally demonstrated a reasonable interpersonal style (e.g. was friendly, warm, positive, constructive etc.). However, at times therapist's style seemed to hinder the therapeutic alliance (e.g. inappropriate use of humour invalidated patient's feelings, inappropriate self-disclosure, inappropriate disclosure of ideas about patient's problems [e.g. told patient with eating disorder idea for final weight early in treatment]).
- Therapist showed limited flexibility in adapting their interpersonal style in response to therapeutic context.

4. Good

- Therapist demonstrated a good interpersonal style (e.g. reasonably open, warm, genuine etc.) which seemed to foster a good therapeutic alliance.
- Therapist's interpersonal style was appropriate to the therapeutic context.

5. Excellent

- Therapist consistently demonstrated an excellent interpersonal style (e.g. embodied and effectively communicated sincerity, open-minded curiosity, genuineness, honesty, sensitive concern, respect, optimism, professionalism, encouragement, unconditional positive regard etc.). This style seemed to promote a trusting working relationship between therapist and patient (e.g. appeared to be mutual liking, respect, confidence and commitment between therapist and patient, patient seemed at ease with therapist, patient seemed engaged).
- Therapist seamlessly adapted their interpersonal style so that it was congruent with the therapeutic context (e.g. optimal levels of warmth, concern etc. given the individual patient's current needs).

* Therapeutic context = patient's interpersonal style, presenting problems, history, stage of treatment, formulation & presentation.

7.2. Empathic Understanding - exemplar therapist behaviours

1. Inadequate

- Therapist frequently projected their own attitudes, conventional attitudes, or attitudes derived from particular theoretical systems onto patient.
- Therapist repeatedly ignored or completely misinterpreted obvious elements of patient's viewpoint.
- Therapist negated, undermined or was critical of patient's viewpoint (e.g. appeared disapproving, disparaging or judgemental) or over-identified with patient's perspective (i.e. was not objective).
- Therapist seemed unwilling or unable to revise their understanding of patient's viewpoint in response to new clinical information or in-session changes.

2. Poor

- Therapist accurately reflected more obvious, verbal cues (i.e. what patient explicitly said), but repeatedly ignored or completely misinterpreted more subtle emotional or nonverbal cues* and/or responded inappropriately (e.g. excessive empathy).
- Therapist found it difficult or seemed reluctant to revise their understanding of patient's viewpoint in response to new clinical information or in-session changes.
- If relevant, therapist showed insufficient awareness of how patient's cultural background** impacted on their viewpoint and/or of patient's perception of the therapist.

3. Satisfactory

- Therapist picked up on most verbal and more subtle emotional or nonverbal cues to gain an understanding of patient's viewpoint. However, therapist conveyed this understanding with limited skill (e.g. restating patient's words unnecessarily, overly long summaries, unsuitable tone of voice, intellectualising issues) and/or at times therapist slightly misinterpreted patient or focused on peripheral aspects of what patient said.
- If relevant, therapist showed limited awareness of patient's perception of how patient's cultural background impacted on their viewpoint and/or of patient's perception of the therapist.

4. Good

- Therapist picked up on all verbal, emotional and nonverbal cues to gain an accurate understanding of patient's 'inner world'. Therapist used appropriate verbal (e.g. paraphrasing, summarising, reflecting) and non-verbal responses (e.g. tone of voice, facial expression) to convey this understanding in order to make patient feel listened to and understood.
- If relevant, therapist demonstrated an awareness of patient's perception of how patient's cultural background impacted on their viewpoint and of patient's perception of the therapist.

5. Excellent

- Therapist picked up on all verbal, emotional and nonverbal cues to gain an accurate and intimate understanding of patient's viewpoint. Therapist skilfully employed appropriate verbal (e.g. paraphrasing, summarising, reflecting) and non-verbal responses (e.g. tone of voice, facial expression) to sensitively convey their understanding of the explicit and implicit content, meaning and emotional tone of patient's perspective in order ensure they felt listened to and understood on an emotional and cognitive level.
- If relevant, therapist demonstrated an in-depth awareness of how patient's unique cultural background impacted on their viewpoint and of patient's perception of the therapist and responded accordingly (e.g. reducing eye gaze in response to excessively self-conscious patient.

* Subtle emotional or nonverbal cues = tone of voice, body language etc.

** Patient's cultural background = religion/spirituality, culture, socioeconomic status, age, gender etc.

Domain 8 - Effective Two-way Communication

Ability to elicit, explore and respond to in-depth feedback about the patient's reaction to all aspects of session and to work with the patient to reflect upon and summarise salient session content in order to facilitate joint understanding of crucial therapeutic material and consolidate key learning.

8.1. Patient Feedback - exemplar therapist behaviours

1. Inadequate

- Therapist made little or no attempt to ask for feedback or discouraged patient from giving feedback.
- Therapist repeatedly ignored or disregarded patient feedback or responded in a negative manner (e.g. became defensive or argumentative).

2. Poor

- Therapist infrequently or superficially elicited patient feedback which was of limited use in checking patient's reaction to session.
- Therapist failed to ask for patient feedback about any negative reactions to session, despite clear need to do so (e.g. patient's comments indicated potential concerns).
- Therapist acknowledged patient feedback but made little or no attempt to discuss or respond to it.

3. Satisfactory

- At somewhat regular intervals, therapist elicited patient feedback which provided some insight into patient's understanding of and reaction to session.
- Therapist made some attempt to elicit potentially negative reactions to session (e.g. doubts, scepticism or concerns) as well as positive reactions.
- Therapist listened to, acknowledged and discussed patient feedback in a non-judgemental, supportive manner but failed to respond clearly or appropriately.

4. Good

- Throughout the session, therapist used appropriate questions to determine patient's reaction to session (e.g. asked "how do you think the session is going so far", etc.).
- Therapist encouraged patient to express positive and negative reactions to session.
- Therapist carefully listened to, acknowledged and openly discussed patient feedback in a non-judgemental, supportive and sensitive manner. Therapist was somewhat responsive to patient's feedback.
- If relevant, therapist picked up on subtle cues (e.g. tone of voice, hesitation etc.) to identify difficulties giving authentic feedback (e.g. patient responded in accordance with what they thought therapist wished to hear).

5. Excellent

- Throughout the session, therapist used skilful questioning to elicit in-depth patient feedback which was useful in helping therapist appraise patient's reaction to all aspects of session (e.g. relationship with therapist, interventions, problem formulation etc.).
- Therapist actively encouraged patient to express positive and negative reactions to all aspects of session in a warm and friendly manner which invited exploration.
- Therapist carefully listened to, acknowledged and thoroughly and openly discussed patient feedback in a non-judgemental, supportive and sensitive manner. Therapist responded appropriately to patient's feedback (e.g. adapted interventions, adjusted their own behaviour, clearly and fully answered questions).
- If relevant, therapist picked up on subtle cues to identify difficulties giving authentic feedback and openly discussed these with patient.

8.2. Reflective Summaries - exemplar therapist behaviours

1. Inadequate

- Therapist made little or no attempt to encourage patient to review and reflect on session content.
- Therapist made little or no attempt to provide capsule summaries (i.e. chunking important information) or provided very minimal, highly inappropriate or inaccurate capsule summaries.

2. Poor

- Therapist prompted patient to review and reflect on session content and/or provided capsule summaries either far too rarely (e.g. only once or twice) or far too frequently (e.g. continually providing summaries or asking patient to reflect on session content).
- Reflective summaries were superficial and/or focused on irrelevant information or peripheral learning points and thus were of limited therapeutic benefit.

3. Satisfactory

- At the beginning and end of session, and at regular intervals within session, therapist prompted patient to review and reflect on session content and provided appropriate and accurate capsule summaries.
- Reflective summaries focused on relevant information and learning points and thus were of some use in helping to reinforce understanding of important therapeutic material and consolidate key learning.

4. Good

- At the beginning and end of session, and at appropriately regular intervals within session, therapist encouraged patient to review and reflect on session content (e.g. asked relevant questions such as “to see if we are both on the same track, can you summarise xx?") and provided clear capsule summaries.
- Reflective summaries focused on important information and learning points and thus were useful in helping to facilitate joint understanding of important therapeutic material and consolidate key learning.

5. Excellent

- At the beginning and end of session, and at appropriately regular intervals within session, therapist used skilful questioning to elicit in-depth patient reflections on and summaries of session content and provided clear, succinct and meaningful capsule summaries.
- Reflective summaries skilfully drew together salient information and key learning points in a way which helped to reinforce and facilitate joint understanding of crucial therapeutic material, consolidate key learning and enabled the patient to gain new insights and make therapeutic shifts.

Patient Difficulty Rating

How difficult did you feel this patient was to work with?

1. Very straightforward	**3. Somewhat Difficult**
2. Somewhat Straightforward	**4. Verv Difficult**

Global Performance Rating

How would you rate the therapist's overall performance within this session?

1. Inadequate	• Therapist fails to include feature outlined. • Therapist demonstrates a significant absence of skill or an extremely inappropriate performance which is likely to have negative therapeutic consequences.
2. Poor	• Therapist's performance is somewhat appropriate. • However, major substantive problems or numerous minor problems are evident.
3. Satisfactory	• Therapist demonstrates a satisfactory degree of skill with no major problems. • However, minor problems or inconsistencies are evident in the therapist's performance.
4. Good	• Therapist consistently demonstrates good skill with evidence of only a few minor problems.
5. Excellent	• Therapist consistently demonstrates a very high level of skill with minimal or no problems.

References

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. doi: 10.1017/S1352465801004040

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*(2), 218-244. doi: 10.1037/0033-2909.90.2.218

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012a). Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, 1-11. doi: 10.1080/10503307.2012.716528

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

Keen, A., & Freeston, M. (2008). Assessing competence in cognitive-behavioural therapy. *British Journal of Psychiatry, 193*(1), 60-64. doi: 10.1192/bjp.bp.107.038588

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

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognititve and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use (3rd ed.)*. Oxford: Oxford University Press.

Vallis, T., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54*(3), 381-385. doi: 10.1037/0022-006X.54.3.381

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), 620-630. doi: 10.1037/0022-006X.61.4.620

Weck, F., Bohn, C., Ginzburg, D. M., & Ulrich, S. (2011). Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire sessions. *Psychotherapy Research, 21*(6), 658-669. doi: 10.1080/10503307.2011.602751

Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *Journal of Nervous and Mental Disease, 199*(4), 276-279. doi: 10.1097/NMD.0b013e3182124617

Assessment of Core CBT Skills (ACCS)

FEEDBACK FORM

Version 1- July 2013

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus and was conducted in affiliation with the University of Oxford's Department of Psychiatry and the Oxford Cognitive Therapy Centre (OCTC).

Domain 1 - Agenda Setting

Ability to collaboratively set a realistic, mutually agreed agenda which prioritises suitable, specific target problems and includes all appropriate routine items.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
1.1. Suitable Items	☐	☐	☐	☐	☐
1.2. Range and Scope	☐	☐	☐	☐	☐
1.3. Collaboration	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 7 – 9

Domain 2 - Effective Use of Time

Ability to structure, pace and manage time within the session in a manner which is appropriate to the therapeutic context and patient's capacity for learning and enables maintained focus on important issues.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
2.1. Pace	☐	☐	☐	☐	☐
2.2. Time Management	☐	☐	☐	☐	☐
2.3. Maintained Focus	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 10 – 12

Domain 3 - Homework

Ability to plan achievable homework which has clear and specific goals, is tailored to the individual patient's therapeutic context and is drawn directly from session material. To plan the homework in a collaborative manner which ensures the patient understands the purpose and content of the assignment and has the opportunity to discuss their reaction to it. To conduct a detailed review of previous homework in order to help the patient identify what they learned from the experience.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
3.1. Appropriate Content	☐	☐	☐	☐	☐
3.2. Collaborative Planning	☐	☐	☐	☐	☐
3.3. Patient Understanding	☐	☐	☐	☐	☐
3.4. Learning Review	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 13 – 16

Domain 4 - Delivering CBT Interventions

Ability to select suitable evidence-based cognitive-behavioural interventions to target clearly specified key cognition(s) or maintaining factors, skilfully apply and review the impact of interventions in a manner which ensures that the aim of the intervention is clear to, and understood by, the patient and that helps the patient identify what they learned from the experience.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
4.1. Suitable Interventions	☐	☐	☐	☐	☐
4.2. Rationale	☐	☐	☐	☐	☐
4.3. Learning Review	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 17 – 19

Domain 5 - Measuring Change

Ability to select appropriate, clinically relevant methods for measuring change in symptoms and associated features (beliefs, behaviours, feelings) and movement towards goals. Ability to administer these methods at suitable time points across and within session and to skilfully review, interpret and respond to the information gleaned.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
5.1. Appropriate Measure	☐	☐	☐	☐	☐
5.2. Implementation	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 20 – 21

Domain 6 - Conceptualisation

Ability to develop a mutually agreed conceptualisation (or 'formulation') which provides a bridge between CBT theory and the individual patient and offers a coherent, meaningful and relevant account of the key features of the patient's problems. Ability to draw upon the conceptualisation in order to facilitate understanding for both the patient and therapist and to develop a coherent treatment strategy.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
6.1. Coherent Strategy	☐	☐	☐	☐	☐
6.2. Patient- Theory Bridge	☐	☐	☐	☐	☐
6.3. Shared Understanding	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 22 – 24

Domain 7 - Managing Therapeutic Relationship

Ability to foster a positive therapeutic alliance by (i) accurately grasping the content and emotional tone of the patient's understanding of themselves and the world around them and sensitively and appropriately conveying this understanding and (ii) embodying an excellent interpersonal style which is congruent with the therapeutic context and patient's needs.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
7.1. Interpersonal Style	☐	☐	☐	☐	☐
7.2. Empathic Understanding	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 25 – 26

Domain 8 - Effective Two-way Communication

Ability to elicit, explore and respond to in-depth feedback about the patient's reaction to all aspects of session and to work with the patient to reflect upon and summarise salient session content in order to facilitate joint understanding of crucial therapeutic material and consolidate key learning.

	Inadequate	Poor	Satisfactory	Good	Excellent
	1	2	3	4	5
8.1. Patient Feedback	☐	☐	☐	☐	☐
8.2. Reflective Summaries	☐	☐	☐	☐	☐

Total Domain Score (*add items*)

See user manual page 27 – 28

Overall Session Score

Total Score (*add domains*)	(range 22 – 110)

Patient Difficulty

Therapist's performance was rated on the basis of a session with a patient who was:

1. Very straightforward	☐	**3. Somewhat Difficult**	☐
2. Somewhat Straightforward	☐	**4. Very Difficult**	☐

Global Performance Rating

Therapist's overall performance within the session was:

Inadequate	Poor	Satisfactory	Good	Excellent
1	2	3	4	5
☐	☐	☐	☐	☐

Please explain *why* you chose this global performance rating:

Skills Profile

Therapist's Key Strengths

Draw on the exemplar therapist behaviours provided under each domain and the session material being assessed to provide examples of what made some aspects of the therapist's performance particularly successful.

Therapist's Learning Needs

Draw on the exemplar therapist behaviours provided under each domain and the session material being assessed to provide examples of what made some aspects of the therapist's performance less successful. Suggest alternative, more skilful way of working within any weaker domains and highlight ways in which the therapist could further develop their skills in this area.

CHAPTER FOUR:

Obtaining feedback from expert and novice clinicians: Examining content validity, face validity and usability of the ACCS.

Chapter Overview

The Assessment of Core CBT Skills (ACCS) is a novel rating scale for assessing the competence with which therapists deliver Cognitive Behavioural Therapy (CBT). The ACCS can be used by assessors to provide formative and summative feedback regarding therapists' performance within observed CBT treatment sessions and can also be used by therapists themselves to rate and reflect on their own performance. Development of the ACCS scale involved three key stages: (1) scale development, (2) expert review, and (3) scale evaluation. This chapter outlines stage two in this process: an 'expert' review of the ACCS scale. Specifically, this chapter outlines a questionnaire study which sought feedback about the ACCS from both experts (*N* = 41) and novices CBT therapists (*N* = 25). The feedback obtained suggest that the ACCS has good face validity, content validity and usability. Thus, the ACCS appears to provide an appropriate and useful measure of CBT competence. Participants' feedback also identified a number of areas where the ACCS required further refinement. Modifications were therefore made to the ACCS in order to further improve the scoring system, the usability of the scale and the quality of the items included in the scale.

Acknowledgements:

With thanks to all of the participants who gave up their valuable time to take part in this study and without whom this research would not have been possible. These include (in alphabetical order) Peter Armstrong, Elisavet Athanasia, Natalie Barnes, Mary Bennett, Simon Blackwell, Shelley Bradley, Gillian Butler, Lorna Cameron, Gavin Clark, Sue Clohessy, Amanda Cole, Maggie Cole, Myra Cooper, Anna Coughtrey, Felicity Cowdrey, Torrey Creed, Gerald Cunningham, Sally Davenport, Annabel David, Christine D'netto, Keith Dobson, Fjola Dogg Helgadottir, Daniel Freeman, Kareena Gates-Bennett, James Gregory, Nick Grey, Brynjar Halldorsson, Sean Harper, Steve Hollon, Natasha Honeybell, Zoe Houston, Helen Kennerley, Carly Keys, Jo Kucharska, Mark Latham, Laura Lockhart, Rachel Manser, Lesley Maunder, Joy McGuire, Rebecca McLoughlin, Elizabeth Napier, Patrick Needham, Leeanne Nicklas, Carmel Oshea, Christine Padesky, Lisa Palmer, Jill Roberts, Isabelle Rudolf Von Rohr, Louise Rutter, David Sandford, Melanie Shadbolt, Jayne Shears, Suzanne Straebler, Maggie Stanton, Donna Sudak, Michelle Suskind, Christina Surawy, John Swan, Katie Tattershaw, Karl Taylor, Liane Thakur, Richard Thwaites, Hakan Turkcapar, Christian Webb, Adrian Whittington, and Andrew Wickson.

Contents

Introduction 229

Method 230

Ethics 230

Participants 230

Procedure 232

Feedback Questionnaire 232

Validity questions. 232

Usability questions. 233

Qualitative feedback. 233

Results 234

Validity 234

Usability 237

Qualitative Feedback 238

Discussion 273

Summary of Results 273

Scale Revisions 273

Limitations 274

References 275

Appendices 277

Appendix 1. Ethical Consent Approval 277

Appendix 2. Participant Information Sheet 278

Appendix 3. Participant Consent Form 281

Appendix 4. Template Email for Initial Invitation to Participate 282

Appendix 5. Feedback Questionnaire 283

Appendix 6. Supporting Information Cover Sheet 291

Appendix 7. Revised ACCS Feedback Form and Scale Manual (version 2) 293

Introduction

As outlined in Chapter one, there are significant limitations with the scales that are currently used to assess CBT therapists' performance within treatment sessions (Barber, Sharpless, Klostermann, & McCarthy, 2007; Kazantzis, 2003; McGlinchey & Dobson, 2003; Muse & McManus, 2013; Rakovshik & McManus, 2010; Sharpless & Barber, 2009; J. Waltz, Addis, Koerner, & Jacobson, 1993). This thesis aims to address this issue by developing of a novel CBT competence assessment scale (Assessment of Core CBT Skills: ACCS). The ACCS provides a framework for assessors to provide formative and summative feedback regarding therapists' performance within observed CBT treatment sessions and for therapists to rate and reflect on their own performance. The central aim of the ACCS is to build upon existing rating scales (particularly the Cognitive Therapy Scale [CTS, or Cognitive Therapy Rating Scale: CTRS, www.beckinstitute.org] and the Cognitive Therapy Scale-Revised: CTS-R [Blackburn et al., 2001]) in order to provide a tool with improved validity, reliability and usability. Development of the ACCS followed a classical measurement theory approach involving three key stages: (1) scale development, (2) expert review, and (3) scale evaluation (DeVellis, 2012). This chapter outlines stage two of the process: an 'expert' review of the ACCS scale.

Review from subject matter experts is an essential ingredient of efforts to improve the quality of rating scales during the developmental phase (Brewer & Hunter, 2005). It is also useful to gain feedback from the target population to better understand how they comprehend and respond to items (Campanelli, Martin, & Rothgeb, 1991). The current study therefore collected formal feedback about the ACCS from both field experts (i.e. experts within the field of CBT who have experience of assessing CBT competence) and the target population (i.e. relative novices with limited CBT experience [e.g. trainees, recently qualified CBT therapists] who are likely to receive feedback on the ACCS and use the tool to rate their own competence). The study had two key aims. First, to examine face validity (i.e. the inclusion of appropriate items that are a credible and plausible measures of CBT competence), content validity (i.e. adequate representation of CBT competence), and perceived usability. Second, to identify areas where the ACCS required further improvement or refinement. This was achieved by asking participants to read the ACCS manual and feedback form and then to complete a questionnaire consisting of a series of open and closed questions about the domains included in the ACCS, as well as the rating scale overall.

Method

Ethics

Ethical permission for the study was obtained by the University of Oxford Research Ethics Committee (see Appendix one). All participants were provided with an information sheet (see Appendix two) and provided written informed consent (see Appendix three).

Participants

The study sought to recruit and collect feedback from two groups: (1) experts within the field of CBT who have experience of assessing CBT competence and (2) relative notices with limited CBT experience (e.g. trainees, recently qualified CBT practitioners) who are likely to use the tool to rate their own competence. There are no established guidelines regarding the number of participants that should be used for a scale review, with variable recommendations depending on the length and style of the scale and practical considerations such as the availability of experts (Haynes, Richard, & Kubany, 1995). However, it is generally agreed that using more than five participants makes it easier to detect and exclude rater outliers and increases the robustness of the feedback ratings (Haynes et al., 1995). Thus, the current study sought to recruit a minimum of five expert and five novice participants, but did not impose an upper limit.

Expert and novice participants were invited to participate via email (see Appendix four). 'Expertise' in this context was broadly defined as individuals with significant experience in the provision of CBT interventions and involvement in evaluating the competence of CBT therapists. Individuals were identified as experts through their professional involvement in the training, selection, or evaluation of CBT therapists' and/or publication of research in the domain of CBT. Novice participants were broadly defined as individuals who were new to and inexperienced in delivering CBT. Individuals were identified as novices through their current or recent involvement in training courses that included a significant CBT training component (e.g. clinical psychology doctorate courses, post graduate diplomas in CBT). Snowball sampling, whereby participants were asked to forward the information about the study to relevant colleagues, was then used to reduce the potential for researcher bias and to increase the number of participants (Atkinson & Flint, 2001).

Due to the recruitment strategy implemented, it cannot be known how many therapists were given information about the study. Seventy expert participants responded to the invitation, 41 (58.57%) of whom returned the feedback questionnaire. Thirty four novice participants

responded to the invitation, 25 (73.53%) of whom returned the feedback questionnaire. Demographics for novice and expert participants are outlined in Table one.

Table 1.

Demographic characteristics for novice and expert participants.

	Novices *n* = 25	**Experts** *n* = 41
Gender - % female	85.00%	58.54%
Age – Mean (*SD*)	36.97 (10.00)	47.10 (11.07)
[range]	[25 – 55]	[30 – 71]
Years practicing CBT	2.03 (2.23)	19.09 (9.67)
Mean (*SD*) [range]	[0.00 – 6.00]	[5.00 – 38.00]
Number of CBT cases:		
% treated 0 cases	44.00 %	0.00 %
% treated 1 - 50 cases	44.00 %	0.00 %
% treated 50 – 200 cases	12.00 %	24.39 %
% treated > 200 cases	0.00 %	75.61 %
Profession: %		
Clinical psychologist	0.00 %	64.10 %
Psychologist	16.00 %	4.80 %
Psychotherapist	12.00 %	0.00 %
Mental health nurse	16.00 %	2.40 %
Nurse	4.00 %	11.90 %
Psychiatrist	4.00 %	7.20 %
CBT therapist	16.00 %	9.60 %
Assistant psychologist	8.00 %	0.00 %
General practitioner	12.00 %	0.00 %
Psychological wellbeing practitioner	4.00 %	0.00 %
Physiotherapist	4.00 %	0.00 %
Speech and language therapist	4.00 %	0.00 %

In line with exceptions, novices were significantly younger than experts (U = 215.50, p = <.001)[1] and had been practicing CBT for significantly less time than experts (U = 3.00, p = <.001)[1]. The majority of novices had treated less than 50 CBT cases, whereas the majority of experts had treated more than 200 CBT cases. As can be seen in Table one, there was some overlap between the novice and expert group in terms of their CBT experience. This was due to three 'outliers' in the groups: one therapist in the expert group had been practicing CBT for five years (during which time they had treated over 200 CBT cases), one therapist in the novice group had been practicing CBT for six years (during which time they had treated under 50 CBT cases), and one therapist in the novice group had treated over 200 cases during the three years they had been practicing CBT. The expert group were primarily clinical psychologists, whereas the profession of the novice group was spread more broadly amongst psychologists, psychotherapists, mental health nurses, CBT therapists and general practitioners. There were also significantly more females in the novice group than in the expert group (X^2[1] = 5.34, p = <.05).

Procedure

Participants were asked to read and familiarise themselves with the ACCS rating scale and manual. They were then asked to complete a questionnaire to provide feedback about the face and content validity for each item and the usability of the scale overall. Copies of the questionnaire and scale were emailed or posted to participants, allowing them to complete the measures in their own time and in a location of their choosing. These anonymised materials were then emailed or posted to the researcher in a pre-paid envelope.

Feedback Questionnaire

In line with recent guidelines (Anastasi & Urbina, 1997; DeVellis, 2012; Haynes et al., 1995) the feedback questionnaire assessed (1) the face and content validity of the ACCS and (2) the usability of the ACCS (see Appendix five).

Validity questions.

Participants rated the degree to which the items in each domain were relevant (1- *not relevant*, 2- *somewhat relevant*, 3- *quite relevant*, 4- *very relevant*) and clear (1- *not clear*, 2- *somewhat clear*, 3- *quite clear*, 4- *very clear*). An index of content validity (CVI: Lynn, 1986; C. F. Waltz & Bausell, 1983; Yagmale, 2003) was then calculated for each domain by calculating the

[1] 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.

percentage of experts who rated the item as being both relevant and clear (i.e. a rating of three or four on the four-point scale). Participants were asked whether there were any important aspects of CBT competence they felt were missing from the scale (i.e. any key competences which the scale neglected) and, if so, what these aspects of competence were. Participants were given the opportunity to identify any domains that overlapped (i.e. appeared to be measuring the same construct). Finally, a yes/no response was used for participants to indicate whether any items in each domain inappropriately assessed multiple aspects of CBT competence (rather than a single, specific and discrete aspect of CBT competence). If participants circled yes, they were asked to indicate which aspects were assessed by the item.

Usability questions.

Participants used a four-point scale to rate how easy they thought the scale would be to use (1- *not easy*, 2- *somewhat easy*, 3- *quite easy*, 4- *very easy*), the overall style, appearance and layout of the scale (1- *poor*, 2- *fair*, 3- *good*, 4- *very good*), and how appropriate they found the scoring system (1- *not appropriate*, 2- *somewhat appropriate*, 3- *quite appropriate*, 4- *very appropriate*). Participants also indicated whether they felt the scale provided ample opportunity to provide both in-depth summative and formative feedback using a yes/no response. If participants circled no, they were asked to indicate what they felt was missing.

Qualitative feedback.

Where participants provided a rating of three or below on the four-point scales for the relevance or clarity of the domain, they were asked to provide qualitative feedback to indicate what revisions could improve the items in the domain. Similarly, where participants provided a rating of three or below on the four-point scale for the appropriateness of the scoring system, style appearance and layout, or ease of use, they were asked to provide qualitative feedback to indicate what revisions could improve the scale. Participants were also provided with space to indicate whether they had any other comments, feedback or suggestions for improvements for each domain as well as the rating scale overall.

Patterns (or themes) which appeared within the written feedback obtained throughout the questionnaire were identified using thematic analysis. This analytic process followed the guidelines outlined by Braun and Clarke (2006) and involved three key phases. First, the feedback provided by each participant was read in order 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 themes were outlined in tabular format and were organised by issues specific to each individual domain and generic issues relating to the entire scale. The number of participants who endorsed the theme was also noted. The analytic process did not provide in-depth description and interpretation of the data (i.e. no attempt was made to identify the broader meanings and implications of the themes or to relate these themes to previous literature), as the intention of the analysis was simply to identify 'surface level' meaning of the participants' comments; i.e. problems with and suggestions for improvements to the ACCS scale (Braun & Clarke, 2006).

Following this analytic process, revisions to the scale were made in response to each theme, or an explanation and justification was provided as to why no revisions were made in response to the issue. The primary author (Kate Muse) took the lead in revising the scale in light of the feedback obtained. However, these revisions were carried out under the supervision of, and in close consultation with, Dr Freda McManus and Dr Sarah Rakovshik.

Results

Validity

Content validity results for each domain in the ACCS scale are presented in Table two. Mean relevance scores for individual domains ranged from 3.83 to 3.98 for experts and from 3.76 to 4.00 for novices, indicating that both novices and experts found all of the domains in the scale at least 'quite' relevant. Mean clarity scores for individual domains ranged from 3.41 to 3.76 for experts and from 3.68 to 3.88 for novices, indicating that both novices and experts found all of the domains in the scale at least 'quite' clear. Mann-Whitney tests revealed no significant differences between the scores for relevance or clarity assigned by novices and experts for any of the scale domains, indicating that novices and experts did not differ significantly in their views regarding the relevance and clarity of the items in the scale. The content validity index (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) ranged from 82.90 % to 100 % for experts and from 84.00 % to 96.00 % for novices, with the CVI being above the suggested threshold of 70 % required to establish content validity (Lynn, 1986) for all domains in the scale. Thus, all domains in the scale were viewed as having acceptable content validity by both novices and experts. In summary, none of the validity analysis indicated that any of the items in the eight domains should be removed from the scale.

Table 2 – Content validity results for each domain in the Assessment of Core CBT Skills (ACCS)

	CVI [1]		Relevance (1 - 4)			Clarity (1 - 4)		
Domain	Novices *n* = 25 %	Experts *n* = 41 %	Novices *n* = 25 Mean (*SD*)	Experts *n* = 41 Mean (*SD*)	Mann-Whitney *U* [2]	Novices *n* = 25 Mean (*SD*)	Experts *n* = 41 Mean (*SD*)	Mann-Whitney *U* [2]
1- Agenda Setting	96.0	85.4	3.96 (.20)	3.83 (.50)	469.50 *p = .25*	3.68 (.56)	3.46 (.67)	426.00 *p* = .18
2- Effective Use of Time	92.0	95.1	3.88 (.44)	3.93 (.35)	496.50 *p = .61*	3.72 (.54)	3.59 (.55)	440.50 *p* = .25
3- Homework	96.0	100	3.96 (.20)	3.93 (.26)	495.50 *p = .59*	3.68 (.56)	3.76 (.44)	489.00 *p* = .68
4- Delivering CBT Interventions	96.0	87.8	3.96 (.20)	3.83 (.44)	457.50 *p = .17*	3.68 (.56)	3.41 (.71)	410.50 *p* = .19
5- Measuring Change	84.0	82.9	3.76 (.52)	3.68 (.57)	478.00 *p = .54*	3.68 (.80)	3.59 (.71)	455.00 *p* = .31
6- Conceptualisation	92.0	92.7	4.00 (.00)	3.98 (.16)	500.00 *p = .44*	3.72 (.74)	3.61 (.63)	440.50 *p* = .21
7- Managing Therapeutic Relationship	92.0	95.1	3.76 (.60)	3.90 (.44)	457.00 *p = .15*	3.88 (.33)	3.71 (.46)	424.00 *p* = .10
8- Effective Two-way Communication	96.0	97.6	4.00 (.00)	3.95 (.22)	487.00 *p = .27*	3.84 (.47)	3.71 (.51)	440.50 *p* = .18

[1] 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.
[2] 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.

Eight participants (12.12 % of the total sample: seven experts and one novice) indicated that domains in the scale inappropriately overlapped (i.e. appeared to be measuring the same construct). The number of participants who indicated that items within a domain inappropriately assessed multiple aspects of CBT competence ranged from 4.80 % to 26.20 % for experts and from 4.00 % to 12.00 % for novices (see Table three for a breakdown per domain). Only items identified as assessing multiple concepts or as overlapping with other items by the majority (>50%) of participants need to be removed from a scale (DeVellis, 2012). Thus no items or domains were removed on this basis. However, further refinements to overcome problems highlighted by participants were carried out on the basis of the qualitative feedback provided in response to these questions (see qualitative results section below).

Table 3 – Participants who indicated that items in the domain inappropriately assessed multiple aspects of CBT competence

Domain	Novices *n* = 25	Experts *n* = 41
1- Agenda Setting	4.00 %	26.20 %
2- Effective Use of Time	8.00 %	11.90 %
3- Homework	8.00 %	16.70 %
4- Delivering CBT Interventions	4.00 %	9.50 %
5- Measuring Change	8.00 %	16.70 %
6- Conceptualisation	8.00 %	14.30 %
7- Managing Therapeutic Relationship	12.00 %	7.10 %
8- Effective Two-way Communication	12.00 %	4.80 %

Nineteen participants (28.79 % of the total sample, all experts) indicated that there were important aspects of CBT 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 feedback section below.

Usability

Usability results for the ACCS scale are presented in Table four. These results indicate that both experts and novices felt the scale would be at least 'quite' easy to use, had at least 'good' style, appearance and layout, and had at least a 'quite' appropriate scoring system. Mann-Whitney 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 views. However, novices assigned a significantly higher rating for ease of use compared to experts. All novice participants and 87.80 % of experts felt the scale provided ample opportunity for both formative and summative feedback. On the basis of the qualitative feedback provided by participants, revisions to the scale were implemented to overcome limitations in relation to the ability of the ACCS to provide formative feedback (see the qualitative feedback section below).

Table 4 - Usability results for the Assessment of Core CBT Skills (ACCS)

	Novices *n* = 25	Experts *n* = 41	Mann-Whitney *U* [1]
Ease of Use (1 - 4) Mean (*SD*)	3.56 (.65)	3.20 (.71)	366.50 *p* = .03
Style, appearance and layout (1 - 4) Mean (*SD*)	3.88 (.33)	3.68 (.52)	422.50 *p* = .10
Appropriate scoring system (1 - 4) Mean (*SD*)	3.80 (.41)	3.51 (.71)	407.50 *p* = .08
Opportunity for feedback (Y / N) (% yes [2])	100 %	87.80 %	-

[1] 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.

[2] Yes indicated that participants felt there was "ample opportunity" to provide formative and summative feedback.

Qualitative Feedback

Patterns (i.e. themes) that appeared within the written qualitative feedback obtained throughout the questionnaire are reported in Tables five – 14. These tables outline the themes specific to each individual domain in turn (Tables five – 12), followed by generic themes relating to the entire scale (Table 13), and finally general comments about the ACCS (Table 14). Themes within the tables are organised in relation to nine 'overarching issues': (1) elements to incorporate (i.e. issues or components which need to be added to the scale), (2) anchor definitions (i.e. the item-specific definitions provided for the scale anchor points require clarifying or revising), (3) structuring issue (i.e. revisions need to be made to the order, layout, or format of the scale), (4) usability issue (revisions need to be made to the scale or manual to improve usability), (5) re-phrasing required (the wording in the item or manual needs improving), (6) typographical errors (i.e. spelling or grammatical errors), (7) multiple aspects assessed (i.e. the item or domain inappropriately assesses multiple different aspects of CBT competence), and (8) overlap within/between domains (i.e. the aspects of competence assessed within the item/domain overlap with the aspects of competence assessed in another item/domain), and (9) positive comments. The revisions to the ACCS made in response to the each theme, or an explanation as to why the issue in the theme was not resolved, are also outlined in the tables.

Table 5 – Themes emerging from qualitative feedback: Issues specific to domain one (agenda setting)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Patient difficulty (addition throughout the items in the domain).		1	Patient difficulty/complexity has not been incorporated within the scoring system of individual items as this may unduly confound the ratings. Furthermore, reserving the highest ratings for performance demonstrated in the face of difficulties or added complexity may unfairly prevent those working with more straightforward cases from being awarded a high performance rating. However, assessors are asked to provide a separate patient complexity rating at the end of the scale. This is so that the context within which the therapist's performance has been evaluated can be taken into account when viewing the ACCS ratings (i.e. it identifies whether the therapist has competently delivered CBT with a patient who is straightforward to work with vs. a patient who has more complex difficulties). This has now been clarified within the instructions to the manual.
	The complete agenda should be jointly summarised and agreed upon (addition within the collaboration item).		1	A bullet point relating to whether the agenda was jointly designed and agreed upon by the patient and therapist was added within the upper bandings in the anchor description. However, in response to other feedback, collaboration was subsequently removed from individual domains and included as a separate, independent item within the scale.
	Incorporate the issue of conveying empathy/the importance of focusing on the issues most relevant to the patient within a rating of 5.		1	A bullet point about therapist's ability to convey the importance of working on what was most relevant and important to patient was added within the anchor description of the highest banding.
	Writing the agenda down is important and should be included.		1	It was decided not to incorporate these issues within the domain. This decision was made because competently setting an agenda involves multiple different

elements to incorporate continued.	Suitable Items could include the requirement for the therapist to provide a very brief rationale for the choice of agenda items.		1	components. So, as it stands, the domain includes a number of different items, each of which refer to different skills which are required to set an agenda. Feedback has highlighted the complexity of this domain and the need to simplify and streamline the item further. Thus, it was felt that adding additional components (and ones which were not highlighted by a large number of participants) would cause further confusion.
	Include the issue of whether therapists' allocated time to items.		2	
	Assess whether therapists' included learning points from the previous session on the agenda.	1		Key learning from the previous session was added as a standard item which should usually be included on the agenda.
	It is also important to prepare the patient for agenda setting / provide a rationale for agenda setting.	1		This is indeed an important aspect of using the agenda setting process effectively. However, a rationale for agenda setting should be provided in an early treatment session and so would not be evident in a mid-treatment session. As this scale is designed to assess competence on the basis of a mid-treatment session this would not be an appropriate aspect of competence to include.
anchor definitions	The examples of well-defined, specific agenda items are unrealistic as they relate to interventions and not topics of conversation.		2	The examples given were changed to reflect broader agenda topics rather than specific interventions (e.g. improving difficulty getting to sleep, addressing financial concerns).
	Vague agenda items are sometimes appropriate (e.g. to try and understand what the person is struggling with).		1	To try and clarify this issue, it was made clear that the agenda also needed to be suited to the stage of therapy and patient's presentation (e.g. therapists would be marked down if they included a very concrete, specific item that was not appropriate given the individuals current presentation). Also, in line with the comment above, it was made clear that well defined, specific items did not need to refer to very specific interventions or issues, but could instead reflect broader topics.
	Clarify what is meant by "skilfully" in relation to defining agenda items (see excellent rating).	1	1	The item was re-phrased to remove the term skilfully. Instead- the item now reads "Therapist consistently helped the patient identify and prioritise items which met all of the following criteria: Well-defined, specific, and measurable. etc.".

anchor definitions continued ...

Rename the item “suitable items” to read "specificity”.		1	It was decided not to re-name this item as there are many components that relate to the agenda items being suitable other than the specificity of the items. Furthermore, only one participant highlighted a problem with the title, indicating that other participants found the title appropriate.
Is increased collaboration necessary?		1	Further guidance has been added to indicate that *it may be appropriate* for therapists to take more responsibility in initial sessions, but responsibility should be increasingly transferred to the patient as therapy progresses, *depending on patient's presentation*. As to whether increased collaboration is necessary to obtain positive therapeutic benefits- this seems to be an empirical question which could be answered by examining the relationship between collaboration on this scale and patient outcome. In addition, collaboration was subsequently removed from individual domains and included as a separate, independent item within the scale. However, this information was retained within the new collaboration item.
Expand on the issue of prioritising.		1	Reference to prioritising of agenda items has been added within each anchor description for suitable items, rather than only appearing in the overarching domain definition.
Agenda items may not always be amenable to change.	1	2	Clarification was added to indicate that attention may sometimes need to be given to items the patient deems important, even if these do not relevant to their presenting problem / treatment goals or it is unlikely that progress can be made in solving the problem (e.g. allowing the patient time to ‘vent’ when they have nobody else to talk to about the issue).
Agenda items may not be amenable to *within-session* change (e.g. item may refer to work being conducted for homework outside of the session).		2	The term ‘within session’ was removed from the anchor description and instead reads “likely to make progress in solving the problem”.
Define amenable to change.	2		To clarify the phrase, the term ‘within session’ was removed from the anchor description and instead reads “likely to make progress in solving the problem”.

anchor definitions continued …	Define ‘ideal’ scope / significance.	1		The term ideal is somewhat ambiguous. The definition was, therefore, re-phrased to read instead “therapist skilfully prioritised an optimal number of items with ideal scope and significance, making it highly likely that all items could be thoroughly covered in-session”.
	Define 'standard items'.	1		A definition of the term was added: “Standard items = items which should usually be included on the agenda (e.g. mood check, week review, key learning from previous session, homework review, setting homework, session feedback etc.)”.
	It may not always be appropriate to include all standard items in the agenda.		1	Clarification about occasions when the inclusion of standard items would not be appropriate was added. “NB: There may be occasions when it is appropriate *not* to include all standard items on the agenda of a session (e.g. if a routine for carrying out a mood check at the beginning of every session has been firmly established, it may not be useful to restate this agenda item)”. In addition, wording of scale anchors was changed to indicate that lower marks should only be given when a therapist *inappropriately* omits standard items.
	Refer to therapists being passive or didactic rather than only didactic within the collaboration item.		1	The term passive was added to the anchor description. Although the collaboration was subsequently removed from individual domains and included as a separate, independent item within the scale, this information was retained within the new collaboration item.
usability issue	How do you score items 1.2 and 1.3 if there is no agenda?		1	Clarification has been added that a rating of 1-limited should be provided where the therapist did not set an agenda (“therapist made little or no attempt to outline any specific topics to be covered in the session”).
re-phrasing required	Phrase 'under or over optimistic' is unclear.		1	This wording was re-phrased to “mismatch between available time and number of items or scope and significance of items”.
	Some phrasing in items 1.1 and 1.2 sounds uncollaborative (e.g. ‘therapist prioritised’).		3	The wording for both items was re-phrased wherever the phrasing gave a sense of primacy to the therapist in agenda setting. (e.g. “therapist identified” change to “therapist help patient identify”).

re-phrasing required continued ...	The term 'patient concerns' in 1.3 collaboration is ambiguous.		1	This wording was removed from the collaboration item.
	More clearly define "range/scope" in title of item 1.2	1	4	Item renamed "feasible agenda" to improve clarity.
	Replace 'didactic' with 'instructive'.		1	The term "didactic" was replaced with the term "instructive" throughout the scale.
typo	-		1	The typo was corrected.
multiple aspects assessed	Range and scope are two issues in one item (1.2).		1	The item has now been re-named 'feasibility'. As both of these issues (inappropriate range of items and scope of items) could create an agenda which is unfeasible given the time available, both were retained within the item.
	Consider breaking down the domain into further micro- items.		1	Collaboration was removed from the domain. It was decided not to further break down 'feasible agenda' or 'suitable items' because feedback indicated that the scale already included a large number of items. Further splitting items would, therefore, serve to further complicate and add to the length of the scale. However, further clarification about how to use the multiple different components involved in developing a suitable agenda were added (e.g. "a number of items had major substantive problems in one or more of the following areas", "therapist consistently helped the patient identify and prioritise items which met all of the following criteria" etc.).
	Overlap between suitable items and range and scope (particularly in terms of items being amenable to within session change).		1	To reduce the degree of overlap between the items the phrase "amenable to within session change" has now been removed from item 1.1 and the phrase "likely to make progress in solving the problem" has been used instead.
positive comments	Clarity of anchor definitions		3	-
	Useful addition/extension		3	-
	Generic		3	-

Table 6 – Themes emerging from qualitative feedback: Issues specific to domain two (effective use of time)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	It would be useful to consider whether the therapist was explicit about any decision to show "appropriate flexibility" in changing the focus of the session in item 2.3.		1	The therapist's ability to make the decision to change focus openly with the patient was added as a skill evident within those rated the highest banding "any decisions to change focus were negotiated openly and explicitly with the patient".
	Therapists often do not leave enough time to get feedback on homework- could this be included here?		1	The issue of therapists not leaving some time at end of the session to complete tasks, such as getting feedback on homework, was included. However, the issue of getting feedback was not used as an example as this was felt to be too specific. Instead, the example referred to homework planning more broadly "e.g. not quite enough time left at end of session to plan homework thoroughly".
	Add something in 2.1 to capture tendency for some trainees to move on from emotion too fast.		1	An example was given in the scale anchors of 2.1 of therapists' moving on from emotionally difficult topics too quickly (in lower bandings ") vs. allowing patient to experience emotions rather than rushing onto an intervention/another topic (in upper bandings).
	Use of paraphrasing and summarising to keep patient focused.	1		Paraphrasing and summarising is one of the methods that can be used to keep the patient focused and, therefore, to stick to the agenda. However, this is not the only way of maintaining focus. Furthermore, these skills are already highlighted in the reflective summaries item. Thus, it was decided not to include this issue within he domain.
	Make client aware of time.	1		This was not included as this is a somewhat minor issue, only one participant highlighted the issue, and it overlaps considerably with the collaboration item.

	Therapist's rate of speech could be too fast/slow.		1	This was added as an example of inappropriate pacing ("spoke to rapidly or too slowly").
anchor definitions	Define therapeutic context.	1		A definition of the term was included: "therapeutic context = patient's presenting problems, goals, stage of treatment, formulation, presentation etc.".
	Rigidly following an agenda is problematic and could alienate the patient – sometimes flexibility is appropriate.	2	2	Rigidly following an agenda was added as an example of inappropriate performance or a major substantive problem in the lower bandings of the item.
	Clarity is needed as to the reason for the late start to the session as this may not be evident from watching the session. Suggestion by two participants to delete this. One participant also noted that allocation of time and starting late were distinct issues.		7	Reference to the session starting late was removed from the anchor description.
	Add more examples of 'important items not covered' in time management item.	1		More examples of important items which could be omitted from the session due to poor time management were provided.
	Define "ran very significantly over time".	1	1	It was felt that it would be inappropriate to specify an exact time frame as this will vary depending on context. For this reason, the following examples were given "e.g. 50 minute session lasted 70 minutes or ended after 30 minutes".
	Define "unduly regimented".	1		Aspects of the item relating to structure of session were removed in response to feedback that structure was a separate issue from maintaining focus.
	Add more examples of pace being too slow/fast.	1		Further examples of inappropriate pace were added to the pace item.

	Clarification needed that sessions which run over time are acceptable if this was in the patient's best interests.	2		It was made clear that a session running over or under scheduled time was only problematic if there was not "sufficient clinical justification" for doing so. Examples of clinical justification were also provided "significant patient distress, expression of risk etc.".
Structuring issue	Remove maintained focus and amalgamate core elements within time management.	1		Although there is some degree of overlap between maintained focus and time management, it was decided that these items should remain as two separate items as only one participant felt that this separation was not appropriate.
usability issue	Rating guidance needed as to whether the assessor should continue to listen to the recording if it significantly overruns (e.g. switch off after 60 minutes).		1	It was decided not to provide guidance regarding this issue. Instead, individual courses or services can make decisions about the most appropriate course of action to take if a session overruns.
re-phrasing required	The word 'control' is unhelpful.		1	This word was removed.
	The domain definition needs to be re-worded.		1	This wording was re-phrased. Additionally, definitions were provided for each item individually, rather than for the domain as a whole.
	The phrase 'issues outstanding' sounds clumsy.		1	This wording was re-phrased to "omitted issues".
	The therapist's comment in the excellent range of maintained focus is too wordy.		1	This wording was re-phrased and the example was reduced in length.
typo	-	1		The typo was corrected.
multiple aspects assessed	Item 2.3 assesses the issue of structure and maintaining focus which are distinct issues. This also means that the heading "maintained focus"		4	Aspects of the item relating to structure of session were removed.

	doesn't adequately represent the issues covered in the item.		
positive comments	Useful addition/extension.	4	-
	Clarity of anchor definitions.	2	-
	Generic.	1	-
	Flexibility of the item (accepting that therapists can deviate but in a productive way).	1	-

Table 7 – Themes emerging from qualitative feedback: Issues specific to domain three (homework)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Could item 3.1. assess how realistic homework was given the context and patient's abilities / motivation?		1	The issue of how realistic the homework was had already been mentioned briefly within the anchor description outlining the degree to which homework was suited to the individual patient's therapeutic context. However, this definition was extended/re-phrased to make this more explicit.
	Include reference to how well the therapist prepared the patient to do the homework within the planning item.		1	Some reference to how well the therapist prepared the patient to do the homework was already included in the 'patient understanding' item. However, this was expanded upon and re-framed in order to make it clear that these points referred to how well the therapist prepared the patient to complete homework. In addition, the issue of planning homework was created as an independent item.
	Should a homework review also involve identifying any problems arising and taking the opportunity to plan future work?		1	This may be a useful part of the homework review process. However, the item already includes multiple components and many participants commented that the item needs to be further streamlined. Thus, as only one participant suggested the addition, this has not been incorporated within the item.
	Comment on whether the therapist asked the patient to summarise the homework at the end of the session.	1		Therapist's ability to use summaries to facilitate patient understanding was added to the 'good' and 'excellent' rating of the 'patient understanding' item (changed later to 'rationale for homework'.
	I do wonder in the excellent exemplar whether it should also include something about working with the client to translate the leaning back to the formulation. It might already be implied, but could be more explicit.		1	This issue was made more explicit: e.g. "to link this learning to their formulation / treatment goals (e.g. identified what it tells them about how best to understand or make progress with their difficulties)".

	Is increased independence necessary?		1	Further guidance has been added to indicate that *it may be appropriate* for therapists to take more responsibility in initial sessions, but responsibility should be increasingly transferred to the patient as therapy progresses, *depending on patient's presentation*. As to whether increased collaboration is necessary to obtain positive therapeutic benefits- this seems to be an empirical question that could be answered by examining the relationship between collaboration on this scale and patient outcome. In addition, collaboration was subsequently removed from individual domains and included as a separate, independent item within the scale. However, this information was retained within the new collaboration item.
	Learning review not specific to homework item		1	Conducting a learning review is indeed relevant for other items and has, therefore, been included elsewhere in the scale (e.g. CBT interventions domain).
	Add working through homework example in higher ratings.		1	The idea of the therapist making some attempt to work through a homework example with the patient in-session was added to the upper bandings.
anchor definitions	Define "likely outcomes of homework".	1		This phrase was re-wording and further definition of the term was provided in the anchor definition.
	Define when learning review should happen.	1		Further guidance as to when a learning review should occur has not been provided. This is because, although it is often likely to occur at the beginning of a session, this is context dependent.
	Define "overly didactic".		1	Further clarification about what is meant by the term overly didactic was provided in the anchor description. Also, the term "didactic" was replaced with the term "instructive" throughout the scale.
	Define collaborative in learning review item.	1		
	Give examples of "jointly designed".	1		N/A - Collaboration has been removed from individual domains and included as a separate, independent item within the scale.
	Collaborative planning is not always appropriate.		1	

anchor definitions continued ..	Homework is not always suitable / necessary.	1		Whether setting and reviewing homework is necessary to obtain positive therapeutic benefits is an important empirical question that could be answered by examining the relationship between homework as measured by this scale and patient outcome. However, both theory and research indicate that between-session work is an important aspect for promoting therapeutic change so the items were retained.
	Is reviewing homework necessary?	1	1	
structuring issue	Homework review appears as the last item in domain, yet is likely to happen before more homework is planned.		3	The items were re-ordered so that the review item appears first.
re-phrasing required	Clumsy phrasing identified.		1	Wording was re-phrased.
	Long sentence		1	Sentence was re-phrased and shortened.
	Therapist statement awkwardly worded.		1	Wording was re-phrased.
	Simplify domain definition.		1	Wording was re-phrased to clarify and simplify the domain definition. Additionally, definitions were provided for each item individually, rather than for the domain as a whole.
typo	-	1		The typo was corrected.
overlap within domain	Overlap between patient understanding and collaborative planning (because of the collaboration elements).		1	Collaboration has been removed from individual domains and included as a separate, independent item within the scale.
multiple aspects assessed	Patient understanding and learning review items assess multiple aspects of competence (too complex).	1	1	Patient understanding was simplified and renamed within the item ‘rationale for homework’. The learning review was also simplified and re-named ‘reviewing homework’.

positive comments	Useful addition / extension.		5	-
	Clarity.	1	2	-
	Satisfactory feels more achievable than on the CTS-R.		1	-

Table 8 – Themes emerging from qualitative feedback: Issues specific to domain four (delivering CBT interventions)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Skilful delivery of interventions does not seem to be measured by this domain.	1	5	An item assessing the implementation of interventions was added.
	A rationale should be drawn from the formulation.		2	This was added to the rationale item.
	Range/frequency of interventions		1	An item assessing the implementation of interventions was added which incorporated this issue.
	Include reference as to whether the interventions were appropriate to the stage of therapy.	1		This is already included in suitable interventions item which assesses whether the interventions were suited to the patient's therapeutic context. Therapeutic context was defined to clarify the use of the term "patient's therapeutic context = patient's presenting problem, current capacity of the patient, stage of treatment etc.".
anchor definitions	Is the specifier "opportunity to do so" necessary? I can imagine a (poor) session that is mostly chatty, and where there aren't necessarily clear opportunities to select interventions due to the therapist's poor skills.		1	This point also applies to domain three where the same specifier is used in reference to homework. This has been removed from both items.
	Define interventions.		2	Further definition of the term along with examples of interventions is provided at the beginning of the domain.

anchor definitions continued ...

How can you distinguish peripheral cognitions / maintaining factors from central cognitions / maintain factors?	1		The wording of this item has now been re-phrased to indicate that irrelevant or peripheral cognitions / behaviours are those which are of limited therapeutic benefit to target (e.g. those which are not linked to distressing emotions or are unlikely to be maintaining problems).
Use of the phrase 'evidence based practice' questionable- there may not always be an evidence base to drawn from and we don't necessarily know which the most appropriate evidence is.		2	The anchor definition has been re-phrased to make it clear that interventions should be "accurately guided by appropriate theory-based practice or practice based on evidence when possible".
Define / give examples of 'maintaining factors'.	1	1	The wording has been re-phrased to provide clarification as to the meaning of the term.
Suitable interventions relates to a number of different factors- outline these.		1	Further clarity has been provided to indicate what is being assessed within the item choosing suitable interventions.
It might be quite difficult to determine what constitutes "missing" an opportunity to administer interventions or deciding to let it pass at that moment due to prioritising/time management.		1	This was removed from the item.
Include phrase "whether positive or negative" in the initial definition of the item reviewing interventions.		1	This phrasing was included in the definition of the item.
I think it is important to not only link interventions to the evidence base but to the formulation.		2	This was added to the item.

	Define the term “excellent”	1		The term excellent was removed and less ambiguous, more specific descriptions were included.
re-phrasing required	Use of the word ‘task’ is indicative of homework	1		This wording was re-phrased.
	The meaning of the term ‘interventions’ is not clear. Would ‘technique’ or ‘method’ be a better term?	1		The word interventions was retained. However, the alternative term CBT ‘techniques’ was also given and further definition of the term was provided.
	The heading “learning review” is too vague.	1		This heading was changed to ‘reviewing interventions’.
	Re-phrase term "challenged" as this implies confrontation / conflict.	1		The phrase was re-worded to read “therapist tried to re-evaluate or develop new core beliefs too early in therapy”.
	Unclear sentence identified.	1		This sentence was re-phrased to improve clarity.
usability issue	Footnote system confusing.		1	The number of footnotes was reduced and the asterisks system was changed to superscript numbers.
typo	-		1	The typo was corrected.
overlap within domain	Providing a rationale and problem solving obstacles are distinct items.		1	The issue of problem solving obstacles was removed from the rationale item.

positive comments	Useful addition/extension	1	-
	Clarity	1	-
	Useful rating guidelines	1	-

Table 9 – Themes emerging from qualitative feedback: Issues specific to domain five (measuring change)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Include reference to 0-8 scales as well as 0-100 and 0-10 scales in appropriate measure item.		1	Reference to 0-8 scales was added in the anchor description.
	Include reference to the clinical context (i.e. do measures make sense to the client and track the most appropriate beliefs, behaviours or moods).		1	This issue was incorporated throughout the choosing suitable measures item.
clarification or revision of anchor definitions	List valid/reliable questionnaires.	1	2	It was not felt that this would be appropriate as a wide range of different questionnaire measures can be used within clinical practice and updated version of questionnaires and new questionnaires are constantly being added.
	Define 'well operationalised'.	1		This has been defined "clearly specified behaviour".
	This implies measures must be used every session.		2	It has now been made clear that this is not the case.
	Expand on examples of types of measures.		1	Further information about the types of methods which can be used to measure change are outlined "these include standardised questionnaires, diaries, frequency counts, ratings of duration of event or experience, self-ratings of emotions and cognitions etc.".

Usability issue	What if no measures was used- how would you score the second item?		1	A qualifier was added to indicate that the lowest banding should be provided if the therapist neglected to measure change.
	Impact of service context – sometimes services may not have access to best methods or sometimes services may choose measures for therapists.		3	A footnote was added to indicate that "therapists should select the most reliable, valid and usable standard measures which were *available* to them. Thus therapists should *not* be penalised if they were unable to access 'gold standard' questionnaire measures (e.g. due to service constraints, cost)".
	The criteria of "most appropriate" seems too subjective (i.e. there is not one "most reliable, valid" etc. measure but many widely used measures).		3	This phrase has been removed from the upper banding removed.
	Is this relevant to early treatment sessions?	2		The ACCS is designed to assess a mid-treatment session, rather than a very early treatment session.
	Is it necessary to measure change?		1	Many experts in the field believe that this is an important element of CBT – indeed many participants in this study commented that it was a useful addition/extension. The domain was, therefore, retained.
re-phrasing required	The headings 'appropriate measure' and 'implementation' are somewhat confusing.		1	The headings were changed to "choosing suitable measures" and "implementing measures".
overlap within domain	Two issues measured- treatment progress and in session change – needs clarification of this at least.	3	1	Further clarification of this issue was provided in the introduction to the domain.
	Confounds standardised / personalised measures.	1		It was decided to retain both of these issues because either may be legitimately used to measure change, depending on the therapeutic context.
positive comments	Useful addition/extension		7	-
	Generic		1	-
	Clarity		2	-

Table 10 – Themes emerging from qualitative feedback: Issues specific to domain six (conceptualisation)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Reformulation		1	
	The clarity and parsimony of formulation.		1	These issues have been incorporated within the 'coherent formulation' item.
	The relevance and usefulness of the conceptualization used.		1	
	Discourage use of double headed arrows in diagrammatic formulations.		1	This is a minor issue which was not highlighted by any other participants. Thus, in an attempt to reduce the length of the item descriptors, this has not been included.
	Perhaps the use of diagrammatic case formulations should be emphasised?		1	This is a minor issue which was not highlighted by any other participants. Thus, in an attempt to reduce the length of the item descriptors, this has not been included.
anchor definitions	Patient theory bridge exemplars are too abstract.		1	More concrete examples were provided.
	Definition problem- strengths / resilience occurs only in band five.		1	This was not changed. This is because it was initially only included in the upper ratings so that therapists who did not incorporate strengths and resilience factors would not be penalised. This decision was made because this is a relatively new concept which has not been incorporated within all theoretical models / treatment packages and is not viewed by all CBT experts as being necessary.
	The wording in this domain is too complex / there is too much jargon.		1	The wording used in this domain was carefully checked and simplified where appropriate.

anchor definitions continued …	Underline the need for appropriate complexity of formulation.	1		This issue was not specifically underlined as it was not thought to be of greater importance than other issues relevant to the formulation.
	Vague terms used in poor / satisfactory (e.g. “some”).		1	Where vague terms were used, these were removed and replaced with clearer, more specific wording.
Structuring issue	Patient theory bridge, followed by shared understanding and finally coherent strategy.	2		The order of items was changed in line with the suggestion. However, the domain was later changed to include only one item.
re-phrasing required	6.1 (2) poor- second bullet point not clearly expressed.	1		This bullet point has now been removed.
	Main heading – use the term formulation rather than conceptualisation.	1	1	The domain heading has been changed to ‘formulation’.
	Item 6.1 – the heading ‘coherent strategy’ is unclear.	1	2	This item has now been removed.
	Item 6.2 – the heading ‘patient-theory bridge’ is confusing.		2	This item has now been re-named ‘coherent formulation’.
	The use of the term 'model' is confusing.		1	The word model has been changed to theory and extra explanation as to the meaning of this term “relevant evidence-based theory should be used to inform the formulation. If a relevant evidence-based model is available, this should be used to develop the formulation (e.g. disorder-specific models such as Clark’s model of panic). If not, the formulation should be based on an appropriate generic model (e.g. Beck’s cognitive model of emotional disorder, personalised vicious cycles)”.
	Re-phrasing: consistent use of formulation or conceptualisation (prefer former).	1	3	The term ‘formulation’ is now used throughout.

	Domain definition is unclear.	1	The domain definition has been re-phrased to improve clarity.
usability issue	Some of the issues outlined in the domain would not be evident in every session (i.e. therapist does not explicitly refer to the formulation in every session as this is not necessary).	2	Issues which were highlighted as being unlikely to be evident in every session were removed from the domain. This was especially apparent in the shared understanding item. As collaboration was removed from this item also, there was very little substance remaining for this item. Thus, the shared understanding item was removed.
typo	-	1	The typo was corrected.
positive comments	Generic	2	-

Table 11 – Themes emerging from qualitative feedback: Issues specific to domain seven (managing therapeutic relationship)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Agreement on goal and task.		2	It was decided not to add this aspect to the domain as this may not be evidenced frequently within treatment sessions.
	Congruency with the therapeutic context.		2	This issue was added to the interpersonal style item.
	Congruence with patient's interpersonal style.	1		This was already included. However, this has now been more clearly highlighted.
	Therapist flexibility / adaptability.		1	This was already included. However, this has now been more clearly highlighted.
	Management of ruptures.		3	Although this is an important issue, it has not been included as this will not always occur during a session.
	Collaboration.		1	Collaboration is included in the scale as a separate item.
	Respect.		1	Being respectful was included as an example of a positive interpersonal style.
	Ethnicity part of cultural background?		1	N/A - reference to cultural competence was removed from the domain.
	Pitch of voice, amount of therapist talking, active listening skills, empathic statements, use of technical terms/jargon, & dealing with distress.	1		These aspects of competence were included in the communication and therapeutic relationship items.

	Give more specific examples of how standard CBT aspects can be used empathically within the empathy item.		1	More examples were provided.
anchor definitions	Define ‘viewpoint’ / Interchangeable use of "inner world" and "participant’s viewpoint".		2	The term “inner world” was changed to viewpoint consistently and this term was defined (i.e. their understanding of themselves and the world around them).
	Define "sensitively".		1	This term is now explained within the item anchors.
structuring issue	Include the relationship within other domains on the scale.		1	This was not done as it was felt that this would be too repetitive.
	Separate understanding + conveying/communicating skilfully into 2 items.		1	This was not carried out as it was felt that it would create too many items.
re-phrasing required	Some sentences were too long.	1		The domain was carefully checked and any long sentences were reduced in length.
	Term ‘managing’ seems impersonal.		1	The title was re-named to “fostering therapeutic relationship”.
	Remove terms good and excellent from rating.		1	These have been removed throughout the scale.
	Domain definition		1	
usability issue	Culture/diversity may not always be relevant- include as a separate item?		2	The idea of ‘cultural competence’ was removed from the domain.
	Eye gaze can’t be seen in audio recordings.		1	This example has been removed and a non-visual example included instead.
typo	-		1	The typo was corrected.

overlap within domains	Empathy and cultural competence are separate domains – possibly include cultural competence as a separate item.	1	The idea of 'cultural competence' was removed from the domain. It was decided not to include this a separate item as it may not always be relevant within treatment sessions.

Table 12 – Themes emerging from qualitative feedback: Issues specific to domain eight (effective two-way communication)

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Feedback sought on content and process issues for excellent feedback.		1	It was made clear that an upper rating requires the therapist to consistently check the patient's understanding of and reaction to *all aspects of session*.
	Response to negative feedback.		1	An appropriate response to feedback (whether positive or negative) was included in the upper ratings.
anchor definitions	One of the examples sounds contrived.		1	This example was replaced.
	Define term "useful".		1	This term was removed and the sentence was re-phrased to improve clarity.
re-phrasing required	Domain definition needs re-phrasing.		1	The domain definition was re-phrased.
	Use 'accurate' instead of 'in-depth'.		1	This wording has been changed.
	Re-phrase the example question.		1	The example question was re-phrased.
	How "session was going" sounds clunky.	1		This wording has been changed.
	Some of the wording in reflective summaries is a little complex		1	The item was carefully checked and, wherever possible, the wording was simplified.
positive comment	Clarity		1	-

Table 13 – Themes emerging from qualitative feedback: Generic scale issues.

Overarching Issue	Theme *Summaries reflecting core content of feedback (i.e. not direct quotations).*	No. endorsed Novices ***n* = 25**	 Experts ***n* = 41**	Solutions or Responses
elements to incorporate	Guided discovery / Socratic method.		6	This was not added as an additional item in the scale. An item was initially created to assess this aspect of CBT competence in response to the feedback. However, it was found that there was relatively little information to include in the anchor descriptions for the item (because of the ambiguity and specificity of the issue) and there was a great deal of overlap between the content of the item and other items in the scale (particularly the newly created collaboration item and the communication items). In addition, the scale already included a large number of items. Thus, it was decided to further emphasise guided discovery / the Socratic method within these other items instead of creating an additional item.
anchor definitions	One participant felt that the satisfactory ratings were too generous throughout, one felt the satisfactory rating was too generous for 3.1, and two felt the satisfactory rating was too generous for 7.1.	1	3	In response to comments regarding the lack of discrimination between the anchor descriptions, the five-point scale was reduced to a four-point scale. The changes made to the anchor descriptions to reduce the response scale to four-points varied depending for each item. These item-specific changes were based on participants' comments about the problems with the anchor descriptions for that specific item.
	Little discrimination between upper rating bands (4 / 5) in general and also a recurring theme appearing across individual items.	5	12	
	Little discrimination between bands 3 / 4 for domain 8.		1	

anchor definitions continued . . .

Little discrimination between lower rating bands. Highlighted particularly for domain 1, domain 2, and domain 6.		8	
Excellent rating unobtainable.		1	Reducing the scale to four anchor points meant that the scale was no longer 'negatively' skewed (i.e. it included two anchors on the side of 'incompetence' and two anchors on the side of 'competence'). This also removed the mid-point response option, forcing the respondent to make a commitment in the direction of competence or incompetence. Given that some respondents will choose a neutral response in order to avoid making a choice (Van Vaerenbergh & Thomas, 2012) and that the purpose of this scale is to determine whether a therapist can demonstrate competence or not, an even number of response may be more appropriate for the ACCS.
Code 1-4 (so that you have a 'true 0').		1	
Descriptors too negative (3 negative bands, 2 positive).		1	
Label for lower bandings too negative / plus repeatedly poor performance would also be inadequate so use of this term doesn't seem appropriate.	2	3	Lower bandings were re-named with labels which had less negative associations.
Terms "somewhat" and "ideally" too vague- remove?		2	These terms were removed and replaced with less ambiguous phrasing. The scale was also carefully checked for ambiguous phrasing and further revisions were made where appropriate.
Include more example questions therapists should be asking.	2	1	Where possible example questions or exemplar behaviours were included in the scale. However, as many questions are dependent on the therapeutic context, it was not possible to provide a large number of example questions.
Use examples for each item/anchor consistently to avoid bias.		1	Each anchor was checked to identify the number of examples provided – where the balance between the number of examples provided in each anchor description was unevenly weighted, examples were added to redress the balance.

anchor definitions continued. . .	Describe /define patient difficulty more fully.		4	The following guidance has been provided "How complex do you feel the patient was to work with? Issues that may influence complexity include level of patient motivation, level of engagement, the presence of social or environmental problems, the number of problems, the severity of problems, and suitability for CBT".
	Term "patient difficulty" has negative connotations.	1	2	The term patient difficulty has been changed to 'patient complexity'.
	Too much bias toward cognitive aspect of CBT, missing out the behavioural aspect of CBT. This was particularly highlighted in relation to the conceptualisation item and the interventions item. One participant also mentioned this in relation to agenda setting.		13	For the intervention and formulation (formally conceptualisation) domains, the anchor descriptions for all of the items were re-phrased throughout to reflect the importance of behavioural aspects of CBT (e.g. targeting cognitions *or* behaviours inked to distressing emotions). Within the agenda setting, the example agenda items provided were changed to reflect broader agenda topics, rather than specific cognitive interventions. The entire scale was also carefully scrutinised to identify any potential bias towards cognitive aspects of CBT.
structuring issues	Re-order domains so that they are more in line with the order in which they are likely to occur within a session.		6	Domains were re-ordered in line with participants' suggestions.
	Include collaboration as a separate item.		8	Collaboration was removed from individual items and was instead included as a distinct item.
usability issues	When there are multiple exemplars in each level of performance, can you be any more explicit about how many of these need to be present to score at that level?		1	Further guidance has been provided as to how to deal with the multiple aspects included in the items. In particular, the overarching definition for each anchor has been provided within each item (e.g. "absence of skill or an inappropriate performance in one or more of the following areas:").
	More guidance for how to use exemplars in each banding and the multiple dimensions they assess.		1	

usability issues continued . . .	Multiple dimensions assessed within each item.		1	
	What if a therapist shows 1 poor, 1 satisfactory and 1 good behaviour in a particular sub domain?	2	1	Guidance has been provided for assessors as to what rating to provide if a therapist demonstrates behaviours from more than one performance banding.
	Anchor descriptions too long.	3	6	Anchor descriptions have been reduced in length wherever possible.
	Experienced assessors required.		1	Previous research has demonstrated that some degree of expertise is required to assess CBT competence, although adherence can be adequately assessed by trained novices. It was, therefore, felt to be appropriate to expect experienced CBT therapists / assessors to use the scale.
	Relevance to all disorders/protocols?		2	The scale was developed primarily for Axis 1 disorders (in particular anxiety and depression). As the scale has not yet been piloted, the degree to which the scale is relevant across disorders and protocols is unknown.
	Uneven item weighting in the final total score due to the different number of items in the domains.		2	A mean score was provided for each domain, rather than a total score being provided for each domain. This means that the overall ACCS total score provides even weighting to all domains.
	Provide more space for formative feedback. A free text section after each domain is needed.	1	9	A section to provide formative feedback was provided for each domain as well as having an overall formative feedback section at the end of the scale.
	More guidance for narrative feedback.		1	Further guidance as to the type of narrative feedback to provide was added.
	A cover sheet for therapist to include supporting information (context etc.) would be useful.		1	A supporting information cover sheet was created (see Appendix six).
item overlap	Time management / setting agenda.		1	It is true that agenda setting and time management are linked (e.g. setting an unrealistic agenda would make it more difficult to effectively manage time within the session). However, it was felt that these were both important aspects of CBT

				competence and were distinct enough to warrant inclusion of two distinct items. Furthermore, only one participant highlighted this overlap as being problematic.
item overlap continued . . .	Shared understanding / patient feedback & empathic understanding.		1	Item 6.3- shared understanding (within the conceptualisation domain) has now been removed.
	Conceptualisation and selection of strategies.		4	On reflection, it seemed that the conceptualisation domain (particularly the coherent strategy item) and the interventions domain (particularly the suitable interventions item) did seem to overlap too much. It was, therefore, decided to remove the coherent strategy item from the conceptualisation (now formulation) domain and instead to expand upon the links to the formulation within the coherent strategy item (which became 'choosing suitable interventions').
	Some highlighted general overlap between the domains/items.		6	These participants noted that this was inevitable given the overlap between different aspects of CBT competence. They also commented that the ACCS did a good job of separating these aspects of competence (as much as is possible) and did not feel this overlap would impact on the usability of the scale or the ability of the scale to provide useful feedback to therapists. Thus, no amendments were made in response to this feedback.
positive comments	Comprehensive and detailed.	2	10	-
	Intuitive.		8	-
	Good appearance/style.	4	20	-
	Clarity.	2	10	-
	Useful self-reflection / formative feedback tool.	11	8	-

Table 14 – General feedback about the Assessment of Core CBT Skills (ACCS).

Participant Group	General Comments: Illustrative Quotations [1]
expert group (n = 24) [2]	I think the scale is excellent and manages to specify some core CBT skills that are only inferred in the CTS and CTS-R, making it a more useful learning tool. It appears more user-friendly and I imagine would be a tool that assessors would more easily be able to be trained in and achieve inter-rater reliability. I have included a number of comments above about issues to consider, but would say that even in its draft current form it is superior to the CTS and CTS-R. In my role I have had to mark numerous students on the aforementioned scales and believe that this would have be a more helpful scale to score and for students to receive feedback on.
	I think it will be a useful addition to our box of tools in supervision.
	As a rating scale it is clear and extremely comprehensive. As an aid to training it would be valuable and highlight important areas to cover- particularly with more straightforward cases. My concern would be more it could limit- or not acknowledge the value of flexibility and creativity in therapy. It would be a pity if people believed they just had to "tick the boxes" rather than do good therapy.
	Fabulous. Just made sense. Seems so relevant.
	I think there are some areas I have outlined but overall I think it is very helpful.
	This could be improved greatly by relating it more directly to Roth and Pilling's CBT generic, specific and meta competencies. The scale is excessively bulky and wordy. I do, however, like the layout and the fact that items are split into different domains. More examples in the manual would be helpful.
	I think this is an exciting measure with lots of potential.
	This is an important step forward in scale development for rating CBT competence. Very nice work.
	Nice job overall – maybe a few more 'trees" than "forest" but competence is always hard to rate.

expert group continued (n = 24) [2]

Overall, my comments aside, I think you're making an important step forward with this scale. Very nicely done.
I really like this scale and I think it adds something that is missing from the CTS-R in terms of its coherence and separates some of the skills into more discrete sections. It's great that you have worked hard to develop this, thank you!
The General Guidelines at the beginning of the scale are very clear and helpful.
I think you have done a fantastic job in conceptualising what is important in delivering CBT therapy.
Really like the scale -very comprehensive.
The whole thing is really great and very impressive!!!
A great piece of work; hopefully will be accepted in due course nationally.
No additional comments. I wish you good success with this Q.
The general feeling I had was that this has been developed with the standard treatment of Axis 1 patients in mind – especially those for whom there are existing treatment models and protocols. But many therapists (inexperienced as well as experienced, in training and not in training) work with patients for whom Axis II problems co-exist with Axis 1 problems, or for whom Axis II problems dominate.
Overall think the scale is fantastic- very clear and useful. Would definitely move to using this when completed.
Interesting work and definitely an improvement from the CTS-R. Well done.
I was thinking of issues within recorded CBT sessions that had been tricky to rate on the CTS-R as I read the manual and I think this scale has great potential to clarify these when used in applied settings. I look forward to hearing more about the scale development in due course.

expert group continued (n = 24) [2]	None. Good job!
	I think this is a significant improvement over the other existing scales.
	The greater degree of specificity across the scale strikes me as a clear step forward in assessing and providing feedback to therapists.
novice group (n = 9)	I believe that this is an excellent rating scale – better than the usual one used currently (the CTS-R). It is very easy to use and very thorough with clear examples. I believe that it should be a formal requirement to be used by supervisors regularly with trainee CBT therapists in all accredited CBT courses (level one and two).
	Much easier to make sense of quickly in comparison to CTS-R.
	Looks very comprehensive and useful.
	The layout was clear.
	As a trainee I found the examples given in the CTSR useful but I like the simplicity and organisation of the different domains in the ACCS. I also really like the fact it allows for qualitative feedback which would be essential for learning. Overall I am aware I've given very positive responses but it is difficult to me to identify issues when I have never used it and I have limited experience of CBT at present.
	Full of opportunity to on the one hand provide constructive feedback, while on the other can provide a standard to work towards and better oneself by.
	I really like the clear, comprehensive design of this rating scale, particularly the colour-coded icons. This scale will certainly help me in developing my skills.
	From a trainees perspective the item descriptions provide good guidance as to what issues to consider.
	For formative feedback, I think the scale offers the therapist a huge amount of detailed information to reflect upon which is useful.

[1] Illustrative quotations are full and direct quotes taken from the 'general comments' section at the end of the feedback form.

[2] Twenty-four participants (58.54 %) in the expert group nine participant in the novice group (36.00 %) provided 'general comments'.

Discussion

Summary of Results

The present study sought to obtain feedback from expert and novice CBT therapists in order to examine the usability, face validity and content validity of the ACCS scale. The majority of novice and expert participants found the domains in the scale both relevant and clear and only a very small percentage of participants indicated that items in the scale inappropriately assessed multiple concepts or overlapped with other items. Qualitative feedback about the ACCS was generally very positive. In particular, participants commented that the ACCS: is comprehensive, has clear anchor definitions and rating guidelines, is intuitive and user-friendly, is appealing in appearance and layout, includes useful information that is not captured in previous scales (e.g. CTS/R), and provides a useful tool for self-reflection and formative feedback. In terms of usability, both experts and novices felt the scale would be easy to use, was visually appealing (i.e. had good style, appearance and layout), and had an appropriate scoring system. Taken together, these results suggest that the ACCS has good face validity, content validity and usability. Thus, the ACCS appears to provide an appropriate and useful measure of CBT competence.

Scale Revisions

The study also aimed to identify areas where the ACCS required further improvement or refinement and a number of modifications were undertaken in light of the feedback received from both expert and novice participants. These modifications can be broadly categorised within the following three areas. First, changes were made to the scoring system. This included adding space for assessors to provide formative feedback for each domain, reducing the five-point scale to a four-point scale, using more positive banding titles, and using an average item scoring system rather than a total sum scoring system. Second, changes were made to improve the usability of the scale. This included re-phrasing and clarification of anchor descriptions when required, reducing the length of anchor descriptions, reducing ambiguity and increasing behavioural specificity of anchor descriptions, including additional rating guidance, adding a submission coversheet to be submitted with session recordings, and updating the order in which domains appeared in the scale. Finally, changes were made to the specific content of the scale items. These changes included redressing the 'cognitive bias' apparent in the scale by focusing more explicitly on behavioural elements of CBT therapies, including collaboration as a separate item (rather than collaboration being embedded within other items in the scale), providing further clarification and guidance for the measuring change domain, re-naming, re-structuring and reducing the formulation domain, expanding the CBT interventions domain, and re-structuring

and extending the homework domain. The revised version of the ACCS feedback form and scale manual can be seen in Appendix seven.

Limitations

The results of this study must be viewed in the context of a number of possible limitations. The feedback questionnaire included questions about the validity of each of the eight domains in the scale (i.e. the relevance and clarity of the items in the domain and whether any items in the domain inappropriately assessed multiple aspects of CBT competence). As each domain included multiple items, the questionnaire would ideally have asked these questions in reference to each individual item. However, the questionnaire already spanned over 39 pages and took approximately 30 minutes to complete. It was, therefore, not feasible to ask busy therapists to complete any more questions than this. Indeed, it is likely that this would have reduced the number of participants willing to take part. It is also worth noting that participants were asked about how easy they thought the scale was to use. However, participants had only read the scale rather than having used it in practice. In order to be confident that the scale was indeed easy to use, this question would need to be answered by individuals who had experience of using the scale to assess CBT competence.

It is possible that the feedback presented in this study is positively skewed due to the participant sample. As a snowball sampling method was employed, a large number of participants would have seen the invitation to participate. Thus, it is possible that those who responded to the invitation and subsequently completed the questionnaire were those who were interested in the development of a new CBT competence scale and viewed the ACCS in a favourable light. Demand characteristic could also have influenced participants' responses (e.g. participants may not have wanted to give negative feedback about the scale which was then sent back to the scale authors). Any such influences would be more likely to have influenced novice participants, who were less experienced and less senior than the expert participants. However, there were no differences in the questionnaire ratings assigned by novices and experts. The only exception was ease of use, where novices assigned a significantly higher rating (although this difference could also be explained by the novices' relative lack of experience in using rating scales- i.e. they may have underestimated the difficulty and complexity of the task). Steps taken to mitigate against participants providing positively biased feedback included providing therapists with anonymised questionnaires to return, confirming on the information sheet 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.

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Macmillan.

Atkinson, R., & Flint, J. (2001). Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies. *Social research*.

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology-Research and Practice, 38*(5), 493-500. doi: 10.1037/0735-7028.38.5.493

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. doi: 10.1017/S1352465801004040

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77 - 101. doi: 10.1191/1478088706qp063oa

Brewer, J., & Hunter, A. (2005). *Foundations of multimethod research: Synthesizing styles*. Thousand Oaks, CA: Sage Publications.

Campanelli, P. C., Martin, E. A., & Rothgeb, J. M. (1991). The Use of Respondent and Interviewer Debriefing Studies as a Way to Study Response Error in Survey Data. *Journal of the Royal Statistical Society. Series D (The Statistician), 40*(3), 253-264. doi: 10.2307/2348278

DeVellis, R. F. (2012). *Scale development: Theory and Applications*. Thousand Oask, CA: Sage.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247. doi: 10.1037/1040-3590.7.3.238

Kazantzis, N. (2003). Therapist competence in cognitive-behavioural therapies: Review of the contemporary empirical evidence. *Behaviour Change, 20*(1), 1-12. doi: 10.1375/bech.20.1.1.24845

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-385. doi: 10.1097/00006199-198611000-00017

McGlinchey, J. B., & Dobson, K. S. (2003). Treatment integrity concerns in cognitive therapy for depression. *Journal of Cognitive Psychotherapy Winter, 17*(4), 299-318. doi: 10.1891/jcop.17.4.299.52543

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

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. doi: 10.1016/j.cpr.2010.03.004

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. doi: 10.1016/j.cpr.2008.09.008

Van Vaerenbergh, Y., & Thomas, T. D. (2012). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal of Public Opinion Research*. doi: 10.1093/ijpor/eds021

Waltz, C. F., & Bausell, R. B. (1983). *Nursing research: Design, statistics and computer analysis* (2nd ed. ed.). Philadelphia: FA Davis Company.

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), 620-630. doi: 10.1037/0022-006X.61.4.620

Yagmale, F. (2003). Content validity and its estimation. *Journal of Medical Education, 3*(1), 25 - 27.

Appendices

Appendix 1. Ethical Consent Approval

MEDICAL SCIENCES INTER DIVISIONAL RESEARCH ETHICS COMMITTEE
Medical Sciences Divisional Office

University of Oxford, Level 3, John Radcliffe Hospital, Headington, Oxford, OX3 9DU
Tel: +44(0)1865 228974 Fax: +44(0)1865 750750
ethics@medsci.ox.ac.uk http://www.medsci.ox.ac.uk

CONFIDENTIAL

Ref: MSD-IDREC-C1-2013-134

Miss Kate Muse
Powic Building
Department of Psychiatry
Warneford Hospital

28th August 2013

Dear Miss Muse

CUREC checklist

I am writing to acknowledge receipt of your CUREC-1 form for your project: **A review of a novel Cognitive Behavioural Therapy (CBT) competence rating scale**

On the basis of the information you have provided this has now been approved by the Medical Sciences IDREC **subject to:**

a) **it is your responsibility to comply with the requirements for administering any tests or questionnaires and if in doubt to contact the publisher of those tests or questionnaires.**

The reference number for this project is **MSD-IDREC-C1-2013-134** and may I remind you that your project may be reviewed at some stage during an annual audit of projects.

Amendments

Should you at some stage alter some of the techniques or procedures then you should first undertake a checklist (CUREC-1) to see whether these changes alter the ethics of the research. If these remain the same then the committee will require notification of the changes to lodge with the project. If they do not remain the same then you may need to complete a CUREC-2 form and undergo further scrutiny by the committee.

Please do not hesitate to contact me if you have any queries about this.

Yours sincerely

Rosie Mortimer
Senior Assistant Registrar & Secretary to IDREC

Appendix 2. Participant Information Sheet

UNIVERSITY OF OXFORD

DEPARTMENT OF PSYCHIATRY

WARNEFORD HOSPITAL
OXFORD
OX3 7JX

TEL: 07834817388
EMAIL: kate.muse@psych.ox.ac.uk

PARTICIPANT INFORMATION SHEET

A review of a novel CBT competence rating scale

Principal Researcher: Kate Muse, Supervisor: Dr Freda McManus

You are being invited to take part in a review of a CBT 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. Please contact the principal researcher (Kate Muse) using the contact details outlined on page 3 with any questions. If you do decide that you would like to take part then please reply to the email invitation to participate and we will contact you with further details.

Why have I been contacted?

You have been invited to take part in the study because you have been identified as *having a great deal of experience in making judgments about the competence of CBT therapists* **/ or /** *being an individual who practices CBT and who may therefore use the scale to rate their own competence.*. We would therefore like to ask you for your feedback on a competence rating scale that we have developed. It is hoped that at least twenty individuals will choose to be involved in the study in order to highlighting areas where the scale requires refinement.

What is the purpose of the study?

There is a need for improved methods of assessing CBT competence, particularly those which assess whether therapists can demonstrate the skills necessary to effectively deliver CBT. We have developed a CBT 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 improves upon currently available scales (particularly the Revised Cognitive Therapy Scale: CTS-R) in terms of validity, reliability and usability. 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 Kate Muse as a part of a DPhil at the Department of Psychiatry, University of Oxford which is funded by the British Association of Behavioural and Cognitive Psychotherapists (BABCP). 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 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.

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 (ii) to read through and familiarise yourself with the rating scale and manual and then to complete a feedback questionnaire about the scale. Taking part in this study is expected to last approximately one hour. Once completed, we will ask you to post the completed documentation to Kate Muse in the pre-paid envelope provided. Please note that all questionnaires will be identified by a unique ID number, rather that 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.

That being said, we appreciate you taking time from your busy schedule in order to take part in the study and recognise your valuable contribution to this research. Thus, although your name will not be linked to any results, we would like to thank you for your contribution to the research by including your name within the acknowledgements of the rating scale and any published papers which may occur as a result of this study. We have therefore included a tick box option within the consent form for you to indicate whether you would like to have your name included in the acknowledgements or not. Please note that this decision will not affect your ability to participate in the study.

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 the University of Oxford 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?

There will not be any direct benefit to you for taking part in the study. However, the information you provide will be beneficial in helping to highlighting areas where the scale requires refinement and thus will help to improve current methods of assessing CBT 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 Research Ethics Committee at the University of Oxford (ethics@medsci.ox.ac.uk; Medical Sciences Inter-Divisional Research Ethics Committee, Medical Sciences Divisional Office, Level 3 John Radcliffe Hospital, Oxford, OX3 9DU, UK).

Who has reviewed this study?

The study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number MSD-IDREC-C1-2013-134).

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:

Kate Muse
Department of Psychiatry
Warneford Hospital
Oxford
OX3 7JX
Tel: 078348 17388
Email: kate.muse@psych.ox.ac.uk

Supervisor:

Dr. Freda McManus
Oxford Cognitive Therapy Centre
Warneford Hospital
Oxford
OX3 7JX
Tel: 01865 738816
Email: freda.mcmanus@psych.ox.ac.uk

Appendix 3. Participant Consent Form

UNIVERSITY OF OXFORD

DEPARTMENT OF PSYCHIATRY

WARNEFORD HOSPITAL
OXFORD
OX3 7JX

TEL: 07834817388
EMAIL: kate.muse@psych.ox.ac.uk

CONSENT FORM

A review of a novel CBT competence rating scale

Principal Researcher: Kate Muse, Supervisor: Dr Freda McManus

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 25.07.13 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 the University of Oxford and other regulatory authorities, for the purpose of 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.	☐
In understand that the study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number MSD-IDREC-C1-2013-134.	☐
I understand how to raise a concern or make a complaint.	☐
I agree to take part in the above study.	☐

We would like to thank you for your contribution to the research by including your name within the acknowledgements of the rating scale and any published papers. Please indicate whether you would like to have your name included in the acknowledgements or not. This will not influence your ability to take part in the study.

☐ Yes, please acknowledge me ☐ No, I would rather remain anonymous

Name of Participant	Date	Signature
Researcher	Date	Signature

Appendix 4. Template Email for Initial Invitation to Participate

Dear *Insert Name*,

Re: An invitation to participate in an expert review of a novel Cognitive Behavioural Therapy (CBT) competence rating scale

We have developed a Cognitive Behavioural Therapy (CBT) competence rating scale which can be used to provide formative and summative feedback regarding therapists' performance within observed treatment sessions. 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 scale. We have identified you as *having a great deal of experience in making judgments about the competence of CBT therapists* ***/ or /*** *being an individual who practices CBT and who may therefore use the scale to rate their own competence*. We are therefore writing to ask whether you would be able to participate in the study.

Taking part would essentially involve spending about an hour reading through and familiarising yourself with the scale and completing a questionnaire to give feedback on the scale. I have attached an information sheet which provides further details about why the study is being conducted and what your involvement would be.

Please let us know if you are interested in taking part, or if you have any queries.

Thank you very much for taking the time to consider participating.

Best wishes,

Kate Muse & Freda McManus

Appendix 5. Feedback Questionnaire

Please answer the following questions for **each scale domain individually:**

Domain One- Agenda Setting

Please read the description for domain one in the attached manual before answering the next questions.

ITEM 1, Q1. Please rate how relevant the items in this domain are to the assessment of competence in the provision of standard, individual CBT interventions to adults experiencing mental health problems.

Please Circle

1 not relevant

2 somewhat relevant

3 quite relevant

4 very relevant

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

Space given for written feedback

ITEM 1, Q2. Please rate the clarity of the items in this domain.

Examples of problems with clarity include lengthy, awkward or complex wording, the use of ambiguous or vague terms, undefined technical terms and confusing definitions.

Please Circle

1 not clear

2 somewhat clear

3 quite clear

4 very clear

If you circled a three or below, what revisions do you feel could improve the clarity of this domain?

Space given for written feedback

ITEM 1, Q3. Does this domain inappropriately assess multiple aspects of CBT competence, rather than a single, specific and discrete aspect of CBT competence?

Items which assess multiple aspects typically include double-barrelled questions which touch upon more than one issue, whilst only allowing for one answer. The respondent can only respond to one of the issues (i.e. can only rate one aspect of CBT competence) and cannot indicate which one is being answered.

Please Circle **Yes / No**

If you circled yes, what aspects of CBT competence are assessed by the item?

Space given for written feedback

ITEM 1, Q4. Do you have any other comments or feedback about this domain (e.g. other problems with the item descriptions or response scale) or any other suggestions of ways the domain could be improved?

Space given for written feedback

QUESTION 1 – 4 REPEATED FOR ALL EIGHT SCALE DOMAINS

General Questions

Please answer the following questions about the rating scale overall:

GENERAL, Q1: Are there any important aspects of CBT 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 CBT competence do you feel are missing?

Space given for written feedback

GENERAL, Q2: Using the table below, please indicate any domains that are overlapping (i.e. appear to be measuring the same aspect of CBT competence) by placing an **X** in the cell corresponding to the two overlapping domains:

	Agenda Setting	Planning Homework	Reviewing Homework	Effective Use of Time	Measuring Change	Managing Therapeutic	Conceptualisation	Effective Two-way Communication	Intervention Strategy	Quality of Interventions?
Agenda Setting	X									
Planning Homework		X								
Reviewing Homework			X							
Effective Use of Time				X						
Measuring Change					X					
Managing Therapeutic Relationship						X				
Conceptualisation							X			
Effective Two-way Communication								X		
Intervention Strategy									X	
Quality of Interventions?										X

If you placed an X in any cell, which aspect(s) of CBT competence do you think overlap? *Space given for written feedback*

GENERAL, Q3: How appropriate do you find the scoring system?

Examples of problems with the scoring system include an inappropriate number or range of response categories or an unsuitable response format.

Please Circle

1 not appropriate

2 somewhat appropriate

3 quite appropriate

4 very appropriate

If you circled a three or below, what revisions do you feel could improve the scoring system?

Space given for written feedback

GENERAL, Q4: Do you feel the scale provides ample opportunity to provide both in-depth summative and formative feedback?

Summative feedback provides an overall judgment of competence (e.g. a numerical rating) whilst formative feedback provides more detailed, corrective feedback in order to promote self-reflection and guide future learning.

Please Circle **Yes / No**

If you circled no, what did you feel was missing?

Space given for written feedback

GENERAL, Q5: Please rate the overall style, appearance and layout of the scale.

Examples of problems with the overall appearance of the scale include the use of text which is difficult to read or a confusing layout.

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?

Space given for written feedback

GENERAL / QUESTION 6: How easy do you think the rating scale is to use?

Examples of problems with usability include unclear or confusing instructions, difficulty understanding what is being asked or how to respond or unnecessary complexity.

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?

Space given for written feedback

GENERAL / QUESTION 6: Do you have any other comments or feedback about the rating scale or any other suggestions of ways the rating scale could be improved?

Space given for written feedback

Appendix 6: Supporting Information Cover Sheet

ACCS Submission Cover Sheet

Therapist's Name: _____

Date of Submission: _____

Assessor's Name: _____

Supporting Material. The recording should be accompanied by any materials completed in the session or for homework (e.g. diaries, thought records etc.) and any completed questionnaires.

Formulation. Please attach a copy of the formulation (no more than one A4 sheet). The formulation should demonstrate the therapist's understanding - at that point in time - of how the patient's problems developed (background experiences, beliefs and assumptions, critical incidents) and what is maintaining them (vicious circles).

Session Summary:

- Stage of therapy / session number

- Patient's presenting problem(s)

- Patient's treatment goals

- Agenda

- Homework review (What were the assignments? Outcome?)

- Main topic(s) – What were the main things you worked on? What did you do (e.g. worked through problem on thought record, carried out behavioural experiment, looked for alternative explanations)? What is the rationale for this work?

- New homework (What are the assignments? What is the rationale for them?)

- Patient's feedback on the session (reactions, take home messages)

Appendix 7: Revised ACCS Feedback Form and Scale Manual (version 2)

Assessment of Core CBT Skills (ACCS)

USER MANUAL

Version 2- October 2013

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus and was conducted in affiliation with the University of Oxford's Department of Psychiatry and the Oxford Cognitive Therapy Centre (OCTC).

Contents Page

Domain 1 - Agenda Setting 294
1.1. Suitable Items 294
1.2. Feasible Agenda 294
Domain 2 - Formulation 294
2.1. Coherent Formulation 294
Domain 3 - CBT Interventions 294
3.1. Appropriate Intervention Targets 294
Ability to 294
3.2. Choosing Suitable Interventions 294
3.3. Rationale for Interventions 294
3.4. Implementing Interventions 294
3.5. Reviewing Interventions 294
Domain 4 - Homework 294
4.1. Reviewing Homework 294
4.2. Choosing Suitable Homework 294
4.3. Rationale for Homework 294
4.4. Planning Homework 294
Domain 5 - Measuring Change 294
5.1. Choosing Suitable Measures 294
5.2. Implementing Measures 294
Domain 6 - Effective Use of Time 294
6.1. Pace 294
6.2. Time Management 294
6.3. Maintained Focus 294
Domain 7 - Fostering Therapeutic Relationship 294
7.1. Interpersonal style 294
7.2. Empathic Understanding 294
7.3. Collaboration 294
Domain 8 - Effective Two-way Communication 294
8.1. Patient Feedback 294
8.2. Reflective Summaries 294
Patient Complexity Rating 294
References 294

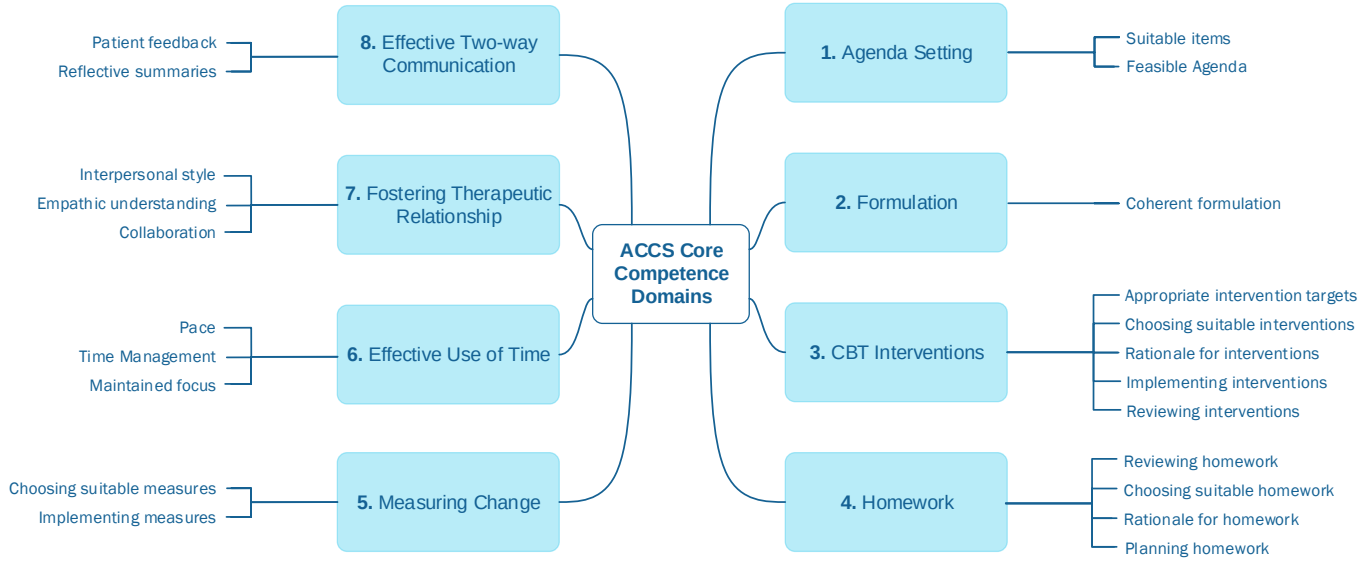
General Guidelines

Scale Scope

The Assessment of Core CBT Skills (ACCS) provides a framework for assessors to deliver formative and summative feedback about a therapist's performance within an observed treatment session. The scale assesses core general therapeutic and CBT-specific skills required to appropriately deliver individual cognitive-behavioural therapy to adults experiencing mental health problems. The scale focuses on competences which are transdiagnostic, rather than skills which are specific to a particular disorder or treatment protocol. The scale is also intended to assess a mid-treatment session and so it does not include skills which, although important, do not occur in most active treatment sessions (e.g. goal setting, relapse prevention etc.).

Core competences were identified for inclusion in the scale on the basis of a review of relevant literature. In particular, the authors drew upon the Cognitive Therapy Scale (CTS, or Cognitive Therapy Rating Scale: CTRS, www.beckinstitute.org), the Cognitive Therapy Scale-Revised (CTS-R: Blackburn et al., 2001) and Roth and Pilling's (2007) competence framework.

As demonstrated in the figure below, the scale features eight competence domains, each of which includes one to five items:



The scale provides a numerical total score which is designed to identify the therapist's competence across the session overall. Numerical scores are also provided for each competence domain and their respective sub-components in order to provide more detailed feedback as to the therapist's specific areas of strength and weakness. In addition, the feedback form includes space for assessors to include further qualitative feedback and to indicate what steps could be taken to further develop competence. Thus the scale can be used both to establish whether therapists have reached an adequate standard of competence (useful for accreditation, evaluating the impact of training, selecting and monitoring research trial therapists) and to provide detailed feedback to aid a therapist's development and progression (useful within supervision and training settings, for continuing personal development, and examining skill acquisition).

How to use the ACCS

This manual is intended to provide guidance for assessors as to how to make judgements about the quality or skilfulness of therapists' performance using the ACCS rating scale. It is recommended that assessors spend some time reading and familiarising themselves with the manual before conducting an assessment. Initially, the manual should be consulted on an item by item basis alongside completing the feedback form. Once the assessor becomes familiar with the scale, the manual can then be used as a frame of reference which can be consulted whenever questions or uncertainties arise. It is recommended that assessors periodically re-review the manual to prevent assessor-drift.

Step 1: Provide a numerical rating for each item

Each of the eight domains is broken down into one to five items, each of which is scored on a four-point scale. The overarching definitions of the four-point scale are as follows:

Performance Band	Generic Definition of Performance Band
1. Limited	• Therapist fails to include feature outlined. • ***Or*** therapist demonstrates a significant absence of skill or an inappropriate performance which is likely to have negative therapeutic consequences.
2. Basic	• Therapist's performance is somewhat appropriate with some degree of skill evident. • However, major substantive problems are evident.
3. Good	• Therapist demonstrates a good degree of skill with no major problems. • However, minor problems or inconsistencies are evident in the therapist's performance.
4. Advanced	• Therapist consistently demonstrates a high level of skill with only very few and very minor problems.

Concrete examples of behaviours consistent with each scale point are provided for each item individually. These 'exemplar therapist behaviours' are intended to provide guidance as to the type of performance which is consistent with each point on the rating scale for that particular item. However, there may be occasions where a therapist demonstrates behaviours from more than one performance banding. Under these circumstances, the assessor is asked to use the overarching definitions outlined above to inform their decision as to which banding to assign. For example, within the 'choosing suitable homework' item, the therapist may have selected homework which related to key maintenance cycles discussed in-session (band 4) but the homework task was too challenging for the patient at this time (band 2). As the homework being too challenging for the patient constitutes a 'major substantive problem', the assessor would assign a rating of 2- Basic. When providing qualitative feedback the assessor has the opportunity to explain why the rating of 2-Basic was given as well as highlighting aspects of the therapist's performance which were consistent with the 4-advanced banding.

The feedback form should be used to assign a rating for each item on the scale. A total domain score can then be provided by calculating the average score of items within the domain. Each total domain score can also be added to provide an overall session score.

Issues to consider-

- The exemplar behaviours are not able to provide an exhaustive list of possible performance indicators. Furthermore, ratings of competence will inevitably include a degree of subjectivity. However, in order to promote consensus between different assessors, it is important that assessors base their ratings on the item-specific exemplar therapist behaviours wherever possible.
- Assessors may find it helpful to make notes of about the therapist's performance that correspond with each item as they view the session. However, assessors should only score items after they have listened to the entire session.
- The 'halo effect' indicates that, when an assessor rates individual aspects of a therapist's competence, they are often unduly influenced by their overall impression of the therapist's performance (Cooper, 1981). As a result the therapist is provided with a uniformly high or low rating, rather than specific feedback as to their strengths and weaknesses. It is important for assessors to bear in mind that therapist's performance may well vary across domains- indeed a therapist can perform poorly in one competence domain and then do very well in another. It is therefore important to score each item independently of other items, and to avoid relying on an overall global impression.

Step 2: Provide qualitative feedback for each domain:

After completing the numerical ratings for each domain, assessors are given the space to highlight areas of strength as well as learning needs. Assessor can draw on the exemplar therapist behaviours provided and the session material being assessed to provide examples of what made some aspects of the therapist's performance more successful than others. Where areas of weakness have been identified, assessors can suggest alternative, more skilful ways of working within this domain and highlight ways in which the therapist could further develop their skills in this area. Such in-depth feedback is invaluable in aiding further development and progression.

Step 3: Provide a global performance rating

The scale also includes a global competence item which is rated using the generic four-point scale outlined above. This global rating should reflect the assessor's overall impression of the therapist's performance within the session as a whole. This global rating will normally correspond to domain ratings. However, this will not always be the case. The assessor may, for example, wish to provide a global rating which reflects the presence or absence of specific aspects of competence which, although important, have not been included within the core competence domains of this scale because they are not relevant across patients and/or treatment sessions.

Step 4: Rate patient complexity:

Patient complexity has not been incorporated within the scoring system of individual items as this may unduly confound the ratings. Furthermore, reserving the highest ratings for performance demonstrated in the face of difficulties or added complexity may unfairly prevent those working with more straightforward cases from being awarded a high performance rating. However, assessors are asked to provide a separate patient complexity rating at the end of the scale. This is so that the context within which the therapist's performance has been evaluated can be taken into account when viewing the ACCS ratings (i.e. it identifies whether the therapist has competently delivered CBT with a patient who is straightforward to work with vs. a patient who has more complex difficulties).

Minimum Competence Score

A minimum score above which a therapist can be considered competent has not been provided for the ACCS. This is because further research is required to establish an empirically proven 'tipping point' for competence using the ACCS. Furthermore, the level of competence and thus the score on the ACCS which is required may vary according to the context and purpose of the assessment. For example, when determining whether candidates are suitable for inclusion on a CBT training course, a lower level of performance may be acceptable, whereas those seeking accreditation may be required to demonstrate a higher level of performance.

Assessment Materials

The ACCS is designed to assess therapists' performance within an 'active' CBT treatment session. Assessments should therefore be made on the basis of a mid-treatment session which involves CBT intervention strategies, rather than an initial assessment session or end of treatment/review session.

Ratings should be based upon source material which retains the most information possible, as this is more likely to provide the most accurate representation of the treatment session and thus the most accurate assessment of the therapists' competence (Waltz, Addis, Koerner, & Jacobson, 1993). As it is typically not possible to physically sit in on treatment sessions, rating are usually made by viewing recordings of a treatment session. Visual recordings provide the richest source of information and can therefore be viewed as the optimal source material, followed by audio recordings. Ratings should not be made on the basis of session transcripts or therapists' verbal or written reports of the session.

It is recommended that assessors base their ratings on a full treatment session, rather than segments of a session. This is because the reliability of individual item ratings is poorer when assessors view session segments rather than full sessions, with a number of aspects of competence not being assessed at all using this method (Weck, Bohn, Ginzburg, & Ulrich, 2011).

In addition to submitting a session recording, it is recommended that therapists provide a short written case summary. This should briefly outline key contextual information such as the stage of therapy, the patient's presenting problem, the session agenda, an outline of homework assignments, the main focus of the session, a brief summary of interventions carried out so far, a rational for chosen treatment interventions, a treatment plan based on the formulation and treatment targets/patient's goals, and any relevant

patient outcome data. Any materials completed in the session or for homework (e.g. questionnaires, diaries, thought records etc.) should also be submitted alongside the recording. Viewing this additional contextual information will help assessors make an informed rating of the therapist's performance.

Assessor Selection

It is unlikely that assessors who are themselves novice therapists would be able to provide detailed formative feedback about a therapist's specific areas of strengths and weaknesses. Indeed research suggests that therapist competence cannot be reliably rated by trained novices (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011). However, at present it is unclear what expertise is necessary to assess competence. As a minimum, it is recommended that assessments are carried out by therapists who have received formal training or accreditation as a CBT therapist.

There is some evidence that assessors rate competence differently depending on their level of independence (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012a), although it is not known whose ratings are more accurate. There are a number of pros and cons involved with using either supervisors who have a relationship with the therapist or independent assessors to rate a therapist's performance using the ACCS. Independent assessors are less likely to be influenced by social demands, relationship dynamics or information other than that obtained through viewing treatment sessions (e.g. prior competence, ability in other domains). On the other hand, supervisors may have access to a greater wealth of contextual information (e.g. treatment context, therapist's broader work) and be better placed to provide appropriate formative feedback. Those selecting assessors will therefore need to weigh up these pros and cons and make a decision based on pragmatics and the purpose of the assessment. If it is decided to use supervisors, one possibility for overcoming potential biases is to implement reliability checks by independent assessors.

Therapist Self-ratings

The scale can also be used by therapists to rate their own competence. As the ability to accurately rate one's own competence has been called into question, it is not recommended that self-ratings be used as a formal measure of therapist competence (Muse & McManus, 2013). However, self-ratings can be a useful learning tool. This is particularly the case when they are used in combination with assessor-ratings as this process of comparison can help to foster self-reflection and highlight on-going learning needs.

Cautions

The ACCS is designed to assess whether a therapist has demonstrated the core generic and CBT specific skills required to deliver effective CBT within a given treatment session. Thus a number of aspects of competence are not assessed by the ACCS. First, the ACCS does not assess therapist's knowledge or understanding of CBT, aspects of competence which can instead be assessed using multiple choice questionnaires, essays, case reports or clinical vignettes (Muse & McManus, 2013). Second, as the ACCS focuses on 'intervention competence', broader professional skills (e.g. ethical practice, effective use of supervision) are not covered. Third, the scale focuses on competences which are transdiagnostic, rather than skills which are specific to a particular disorder or treatment protocol. Fourth, as the ACCS assesses core CBT skills evident during active, mid-treatment therapy sessions, it does not assess therapist's assessment or relapse-prevention skills. Hence it is recommended that, where the scale is used for summative assessment purposes, the ACCS should not be used as a stand-alone measure of competence. Instead the ACCS should form part of a multi-method competence assessment programme.

Although competence is often inferred from rating one or two isolated treatment sessions per therapist, recent studies suggest that a much higher sample of clinical work needs to be assessed in order to reliably assess therapist competence (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012b; Keen & Freeston, 2008). Furthermore, collecting multiple assessors' ratings of the same treatment session can increase reliability (Vallis, Shaw, & Dobson, 1986) and reduce the impact of the halo effect (Streiner & Norman, 2003). Thus in order to be confident of a reliable judgement as to a therapist's competence, multiple assessors would be used to assess a number of therapy sessions drawn across different patients (between 15 and 24 sessions rated by a single assessor, or 19 sessions rated by two assessors: Keen & Freeston, 2008). Although financial and time constraints often make this unfeasible, it is important to recognise the limitations involved in drawing conclusions about a therapist's competence and thus fitness to practice on the basis of a limited number of session ratings.

Domain 1 - Agenda Setting

NB This item refers to **planning and setting** an appropriate agenda **not adhering to it** which should be rated under Item 6.3 'Maintained Focus'.

1.1. Suitable Items

Ability to help the patient identify and prioritise specific, relevant and appropriate agenda items.

1. Limited

Absence of skill or an ***inappropriate performance:***

- Therapist made little or no attempt to outline any specific topics to be covered in the session (e.g. asked "How would you like to spend our time today?" with no further follow-up).

2. Basic

Therapist helped the patient identify and prioritise topics to be covered but a ***number of items*** had ***major substantive problems*** in one or more of the following areas:

- Vague, unclear or overly general (e.g. my anxiety).
- Not suited to stage of therapy or patient's presentation.
- Not relevant to patient's presenting problem or treatment goals [1].
- Unlikely to make progress in solving the problem [1].
- Inappropriate given potential risk factors [2].

- ***And/or*** a number of relevant standard items inappropriately omitted [3].

3. Good

Therapist helped the patient identify and prioritise items which met ***most*** of the following criteria, with only ***minor problems or inconsistencies***:

- Well-defined, specific, and measurable (e.g. improving difficulty getting to sleep, addressing financial concerns).
- Suited to stage of therapy and patient's presentation.
- Relevant to patient's presenting problem / goals.
- Likely to make progress in solving the problem.
- Appropriate given potential risk factors.

- ***And*** included most relevant standard items.

4. Advanced

Therapist ***consistently*** helped the patient identify and prioritise items which met ***all*** of the following criteria:

- Well-defined, specific, and measurable.
- Suited to stage of therapy and patient's presentation.
- Relevant to patient's presenting problem / goals.
- Likely to make progress in solving the problem.
- Appropriate given potential risk factors.

- ***And*** included all relevant standard items.

[1] NB: Attention may sometimes need to be given to items the patient deems important, even if these do not appear relevant to their presenting problem / treatment goals or it is unlikely that progress can be made in solving the problem (e.g. allowing the patient time to 'vent' when they have nobody else to talk to about the issue).

[2] Potential risk factors = risk of harm to self or others (e.g. expression of suicidal wishes), degree of distress associated with problem.

[3] Standard items = items which should usually be included on the agenda (e.g. mood check, week review, key learning from previous session, homework review, setting homework, session feedback etc.). NB: There may be occasions when it is appropriate *not* to include all standard items on the agenda of a session (e.g. if a routine for carrying out a mood check at the beginning of every session has been firmly established, it may not be useful to restate this agenda item).

1.2. Feasible Agenda

Ability to set an agenda which is realistic and feasible given the time available.

1. Limited

- Therapist made little or no attempt to outline any specific topics to be covered in the session

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas***:***

- Items far too numerous, large or overwhelming to be feasible / enable any progress in solving the problem.
- ***Or*** too few or insignificant items which could easily be covered in a very small time frame (e.g. reviewing a small piece of homework).

2. Basic

Major substantive problems:

- Mismatch between available time and number of items ***or*** scope and significance of items- making it unlikely that adequate time could be spent on all items ***or*** items likely to be covered well before the end of the session.

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- Appropriate number of items with suitable scope and significance given the time available, making it likely all items could be adequately covered in- session.

4. Advanced

Consistently high level of skill:

- Optimal number of items with ideal scope and significance given the time available, making it highly likely that maximum use could be made of the session and that all items could be thoroughly covered.

Domain 2 - Formulation

An initial formulation is typically developed as part of the assessment process. However, a formulation should not simply precede treatment but should evolve throughout therapy as further clinical information arises and recurrent patterns and themes emerge. Thus a therapist's ability to formulate a patient's problems should be evident throughout the course of treatment.

NB

In addition to viewing the session recording, it is necessary to review any written or diagrammatic formulation materials in order to make an informed rating on this item.

Whether the therapist drew upon the formulation to highlight potential treatment targets and intervention strategies should be rated under domain 3 'CBT Interventions'.

2.1. Coherent Formulation

Ability to develop a clear formulation which draws upon appropriate evidence-based theory to offer a concise, comprehensive and personalised explanation of relevant history, triggers and maintaining features of the patient's problems.

1. Limited

- Therapist made very little or no attempt to formulate the patient's problems.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist's formulation of the patient's problems had major flaws (e.g. did not use evidence-based CBT theory [1], attempted to fit the patient into an inappropriate model, inaccurately reproduced a relevant CBT model, included largely irrelevant information, failed to incorporate key features of the patient's presentation [2], failed to identify links between different features of the patient's problems, formulation overly complex or far too simplistic given stage of treatment or patient's ability to understand, formulation incoherent or disjointed).
- If new clinical material emerged, therapist seemed unwilling or unable to review or revise the formulation.

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist utilised appropriate evidence-based theory to provide an account of the patient's problems. However, this account was limited (e.g. vague or abstract, unclear or confusing, too complex or simplistic given stage of treatment or patient's ability to understand, included information of limited relevance, failed to show how features of the patient's presentation related to each other or the patient's problems in a meaningful way).
- If new clinical material emerged, therapist found it difficult to or seemed reluctant to revise the formulation.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist utilised appropriate evidence-based theory to provide an explanation of the history, triggers and maintaining features of the patient's problems. This explanation was clear and logical but had some problems (e.g. lack of personalisation, lack of concrete and specific examples, missed opportunities to include relevant strength and resilience factors, a more straightforward or detailed formulation would have better suited the stage of treatment or patient's ability to understand).
- If new clinical material emerged, therapist made some attempt to integrate this into the formulation.

4. Advanced

Consistently high level of skill in the following areas:

- Therapist flexibly utilised appropriate evidence-based theory to provide a clear, concise, comprehensive and personalised account of relevant history, triggers and maintaining features of the patient's problems, as well as any relevant strength and resilience factors. This formulation was well suited to the stage of treatment and patient's ability to understand.
- If new clinical material emerged, therapist responded flexibly and fully incorporated this within the formulation.

[1] Relevant evidence-based theory should be used to inform the formulation. If a relevant evidence-based model is available, this should be used to develop the formulation (e.g. disorder-specific models such as Clark's model of panic). If not, the formulation should be based on an appropriate generic model (e.g. Beck's cognitive model of emotional disorder, personalised vicious cycles).

[2] Presentation features = behaviours, cognitions, emotions, physiological experiences, environmental factors, interpersonal issues.

Domain 3 - CBT Interventions

This item refers to the selection and application of cognitive and/or behavioural interventions (also referred to as CBT 'techniques') designed to promote therapeutic change by targeting cognitions, behaviours or emotions likely to be maintaining the patient's problems. A range of CBT interventions can be used to promote change. These include, but are not limited to, cognitive restructuring, exposure/response prevention, addressing safety behaviours, imagery re-scripting, examining pros and cons, thought records, activity monitoring/scheduling, behavioural experiments, role play, graded task assignment, problem solving, assertiveness or other skills training, behaviour modification.

NB A number of factors other than therapist competence can influence whether an intervention is successful in bringing about change. Please, therefore, focus on the quality of the cognitive-behavioural interventions and not whether change actually occurred.

3.1. Appropriate Intervention Targets

Ability to skilfully define, clarify and specify intervention targets which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were highly likely to be maintaining problems.

1. Limited

- Therapist did not identify any intervention targets (e.g. cognitions, behaviours, emotions).

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist lacked a clear rationale for selecting intervention targets (e.g. focused on cognitions, behaviours or emotions which neither evidence-based theory or the patient's formulation suggested were likely to be maintaining the patient's problems).
- Intervention targets were not appropriate to the patient's therapeutic context [1] (e.g. therapist tried to re-evaluate or develop new core beliefs too early in therapy).

2. Basic

Major substantive problems in one or more of the following areas:

- Interventions targeted cognitions, behaviours or emotions which were of limited therapeutic benefit (e.g. cognitions or behaviours were not linked to distressing emotions, relevant evidence-based theory suggested the targets could be maintaining the patient's problems but the idiosyncratic formulation suggested that the targets were irrelevant or peripheral to the patient's problems).
- Intervention targets were largely appropriate to the patient's therapeutic context but required fine tuning (e.g. therapist targeted safety behaviours which the patient was not yet ready to consider changing).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Interventions targeted relevant cognitions, behaviours or emotions which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were maintaining problems. However, these targets were not clearly specified (e.g. were generic, vague or ambiguous).
- Interventions targets were suited to the patient's therapeutic context (e.g. therapist targeted thoughts that were more amenable to change before progressing to more strongly held beliefs).

4. Advanced

Consistently high level of skill in the following areas:

- Interventions targeted relevant cognitions, behaviours or emotions which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were highly likely to be maintaining problems. These targets were skilfully defined, clarified and specified.
- Interventions targets were suited to the patient's therapeutic context.

[1] Patient's therapeutic context = patient's presenting problem, current capacity of the patient, stage of treatment etc.

3.2. Choosing Suitable Interventions

Ability to select cognitive-behavioural interventions which form part of a logical, coherent and unified treatment strategy which is likely to bring about therapeutic change in the treatment target(s) and is suited to the patient's therapeutic context. This selection was accurately guided by appropriate theory-based practice or practice based on evidence when possible.

1. Limited

- Therapist did not select any CBT interventions

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Interventions formed part of an incoherent treatment strategy (e.g. therapist appeared to select interventions in a random, hit-and-miss fashion). ***And/or*** this strategy was highly unlikely to bring about therapeutic change in the treatment targets (e.g. interventions were not at all matched to treatment targets, such as selecting imagery re-scripting to increase activity levels).
- Interventions were poorly suited to the patient's therapeutic context [2] (e.g. intervention far too complex for current capacity of the patient).
- Selection of interventions was not guided by theory or evidence-based CBT practice.

2. Basic

Major substantive problems in one or more of the following areas:

- Interventions formed part of a vague, unclear or disjointed treatment strategy. ***And/or*** this strategy showed little promise in bringing about therapeutic change in the treatment targets (e.g. interventions were poorly matched to treatment targets, such as selecting a core belief worksheet to work with automatic thoughts).
- Interventions were largely appropriate given the patient's therapeutic context but required fine tuning (e.g. too easy or overly ambitious, challenging or complicated for the current capacity of the patient).
- Selection of interventions was guided by inappropriate theory or evidence-based CBT practice (e.g. a treatment protocol which was only vaguely related to the patient's presentation) or therapist significantly misinterpreted or misapplied appropriate theory or evidence-based CBT practice.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Interventions formed part of a generally coherent therapeutic strategy which showed some promise in bringing about therapeutic change in the treatment targets (e.g. interventions were appropriate to the treatment targets, such as selecting activity scheduling to improve mood by increasing activity levels)
- Interventions were appropriate given the patient's therapeutic context (e.g. complexity and difficulty matched with patient's motivation and cognitive functioning).
- Selection of interventions was accurately guided by appropriate theory-based practice or practice based on evidence when possible.

4. Advanced

Consistently high level of skill in the following areas:

- Interventions formed part of a logical, coherent and unified treatment strategy which showed a great deal of promise in bringing about therapeutic change in the treatment targets.
- Appropriate timing and grading meant that interventions were optimally suited to the individual patient's therapeutic context.
- Selection of interventions was accurately guided by appropriate theory-based practice or practice based on evidence when possible.

[2] Patient's therapeutic context = patient's presenting problem, current capacity of the patient, stage of treatment etc.

3.3. Rationale for Interventions

Ability to facilitate the patient's understanding of the importance and potential benefits of interventions.

1. Limited

- Therapist did not implement any CBT interventions.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to facilitate the patient's understanding of the rationale underpinning interventions (e.g. did not explain the importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, these were ignored or disregarded or therapist responded in a negative manner (e.g. defensive, hostile).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist facilitated the patient's understanding of the rationale underpinning interventions with limited skill (e.g. developed a limited or confusing rationale for importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, therapist gave these limited consideration (e.g. acknowledging a question but failing to answer it).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist facilitated the patient's understanding of the rationale underpinning interventions with some degree of skill (e.g. developed a rationale for importance / potential benefits) but some minor problems were evident (e.g. did not relate rationale to the patient's presenting problems / formulation / treatment goals, conveyed the rationale with limited clarity).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, these were acknowledged and therapist made some attempt to discuss and resolve the issues.

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully facilitated the patient's understanding of the rationale underpinning interventions (e.g. clearly and succinctly conveyed a rationale for the importance and potential benefits which aligned with the patient's presenting problems / formulation / treatment goals).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, therapist responded openly, fully discussed the issues and (where possible) resolved the issues (e.g. clarified any ambiguities).

3.4. Implementing Interventions

Ability to systematically implement intervention(s) in a fluent and articulate manner. To be sensitive and responsive to the therapeutic context and provide optimal levels of support, encouragement and praise.

1. Limited

- Therapist did not implement any CBT interventions.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Interventions were misused or were implemented incorrectly in a way that was not justified by therapeutic context (e.g. omitting 'evidence for the belief' when doing a thought record, terminating exposure before any anxiety habituation occurred).
- Therapist lacked skill in implementing interventions (e.g. therapist was inarticulate or incoherent, appeared unfamiliar with techniques).
- Therapist showed very little or no awareness of the therapeutic context ***or*** responded in an unhelpful or negative manner (e.g. therapist was too forceful, appeared discouraging, ignored or dismissed very high levels of arousal or patient distress).

2. Basic

Major substantive problems in one or more of the following areas:

- Interventions were not implemented in the manner in which they were intended and there was no clinical justification for these deviations (e.g. interventions were incomplete, were abandoned prematurely or were missing key components [e.g. therapist neglected to get ratings of pleasure/achievement in activity scheduling]).
- Therapist implemented interventions with limited skill (e.g. at times the therapist was unclear, difficult to understand or appeared confused or muddled).
- Therapist showed some awareness of the therapeutic context but responded with limited sensitivity (e.g. provided limited encouragement or was overly pushy) ***and/or*** appropriate flexibility (e.g. failed to adapt intervention despite indication that it was too distressing or complex for the patient to grasp).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Interventions were mainly carefully and methodically implemented in the manner in which they were intended, unless there was clinical justification for deviating from this. However, some minor skill deficits or inconsistencies were evident (e.g. thought record could have been more detailed).
- Therapist implemented interventions with adequate skill (e.g. therapist was articulate and comprehensible).
- Therapist was largely sensitive to the therapeutic context (e.g. provided appropriate support and encouragement) and made some attempt to adapt interventions to suit the patient's needs where necessary.

4. Advanced

Consistently high level of skill in the following areas:

- Clear and consistent evidence that interventions were carefully, thoroughly and systematically implemented from start to finish in the manner in which they were intended, unless there was clinical justification for deviating from this.
- Therapist consistently implemented interventions with a high level of skill (e.g. therapist was fluent, articulate, clear and comprehensible).
- Therapist was consistently sensitive to the therapeutic context (e.g. provided optimal levels of support, encouragement and praise) and was able to custom tailor interventions to suit the patient's needs where necessary (e.g. broke down a thought record into sections to make the task more manageable/less overwhelming, recognised and addressing subtle safety behaviours, modelled the intervention for a patient who felt it was too threatening).

3.5. Reviewing Interventions

Ability to conduct a comprehensive review of the results of interventions (whether positive or negative) in order to help the patient identify what they learned from the experience.

1. Limited

- Therapist did not implement any CBT interventions.

Or absence of skill or an ***inappropriate performance***:

- Therapist failed to review ***or*** conducted a brief or cursory review of the results of interventions which did not help the patient identify learning (e.g. asked "how do you think that went?" with no further follow up).

2. Basic

Major substantive problems:

- Therapist reviewed the results of interventions and made some attempt to highlight learning implications, but with limited skill (e.g. struggled to relate the results back to the relevant cognition, behaviours or emotions that they were designed to target, did not address negative results).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- Therapist adequately reviewed the results of interventions (whether positive or negative). This review helped the patient identify some useful learning implications (e.g. examined outcomes in relation to prior predictions).

4. Advanced

Consistently high level of skill:

- Therapist skilfully engaged the patient in a comprehensive review of the results of interventions (whether positive or negative) which helped the patient identify useful and relevant learning implications. Therapist helped the patient link this learning to the relevant cognition, behaviours or emotions they were designed to target ***and*** to their formulation / treatment goals (e.g. helped the patient re-evaluate previous conclusions or construct new ideas).

Domain 4 - Homework

4.1. Reviewing Homework

Ability to conduct a comprehensive review of previous homework (whether completed or not) in order to help the patient identify what they learned from the experience.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist failed to review ***or*** conducted a brief or cursory homework review which did not help the patient identify learning (e.g. asked "how did you get on with the homework?" with no further follow up).
- Therapist responded to completed or uncompleted homework in an unhelpful or negative manner (e.g. scolding, judgemental, dismissed homework as unimportant).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist reviewed homework and made some attempt to highlight learning implications from completing (or not completing) homework, but with limited skill (e.g. learning had limited relevance to the patient's problems).
- If relevant, therapist made some attempt to reflect on why non-completion occurred but failed to identify relevant emotional, cognitive or practical difficulties ***and/or*** did not identify ways to overcome future 'blocks' to completion.
- Completed homework was given limited recognition ***and/or*** therapist responded to non-completion with limited sensitivity and understanding.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist adequately reviewed homework. This review helped the patient identify some useful learning implications from completing (or not completing) homework (e.g. examining outcomes in relation to prior predictions).
- If relevant, therapist helped the patient identify relevant emotional, cognitive or practical reasons for non-completion ***and*** discussed ways to overcome future 'blocks' to completion (e.g. asked "how could you overcome xx next time?").
- Therapist responded to completed homework with appropriate encouragement ***and*** to non-completion in a supportive manner.

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully engaged the patient in a comprehensive homework review. This review helped the patient identify useful and relevant learning implications from completing (or not completing) homework and to link this learning to their formulation / treatment goals (e.g. identified what it tells them about how best to understand or make progress with their difficulties).
- If relevant, therapist worked with the patient to identify relevant emotional, cognitive or practical reasons for non-completion ***and*** helped them identify specific ways to overcome future 'blocks' to completion (e.g. using a calendar for a patient who has difficulty remembering to complete homework, using problem solving to break homework down into less threatening components).
- Therapist responded to completed homework with optimal levels of encouragement and praise ***and*** to non-completion in a highly supportive, sensitive and open manner.

4.2. Choosing Suitable Homework

Ability to plan homework which is tailored to the therapeutic context and builds upon session material or previous homework.

1. Limited

- Therapist failed to plan further homework arising from the session or building on previous homework.

Or absence of skill or an ***inappropriate performance***:

- Homework was poorly suited to the patient's therapeutic context [1] (e.g. expecting the patient to carry out a highly feared exposure after first session or planned a difficult or large assignment for a patient with low motivation).

2. Basic

Major substantive problems in one or more of the following areas:

- Homework was largely appropriate given the patient's therapeutic context, but required fine tuning (e.g. overly ambitious, easy, challenging or complicated for the patient, 'standard' homework used without necessary adaptation).
- Homework was not related to issues (e.g. problems, cognitions, behaviours) dealt with in-session ***or*** previous homework, therefore restricting opportunities for relevant information gathering or generalising of concepts and skills learnt into daily life.
- Significant risk the patient would fail to benefit from homework or would have predictions confirmed in a way that was counter therapeutic.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Homework was appropriate given the patient's therapeutic context (e.g. amount and difficulty was achievable given the patient's motivation and cognitive functioning).
- Homework related to issues (e.g. problems, cognitions, behaviours) dealt with in-session ***or*** previous homework, therefore enabling some opportunities for information gathering and/or generalising of concepts and skills learnt into daily life. However, opportunities for more central homework tasks or further learning were missed.

4. Advanced

Consistently high level of skill in the following areas:

- Custom tailoring of homework meant it was optimally suited to the individual patient's therapeutic context (e.g. represented a moderate and realistically achievable challenge for the patient).
- Homework built upon important issues (e.g. problems, cognitions, behaviours) dealt with in-session ***or*** previous homework, therefore enabling opportunities for relevant information gathering and/or generalising of concepts and skills learnt into daily life.
- Therapist thoroughly considered the impact of likely outcomes of homework (e.g. set homework up as 'no lose' or worked with patient to make a plan for dealing with potential negative outcomes).

[1] Therapeutic context = patient's presenting problems, goals, stage of treatment, formulation, presentation etc.

4.3. Rationale for Homework

Ability to facilitate the patient's understanding of the importance and potential benefits of homework.

1. Limited

- Therapist failed to plan further homework arising from the session or building on previous homework.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to facilitate the patient's understanding of the purpose of homework (e.g. failed to develop a rationale for importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, these were ignored or disregarded ***or*** therapist responded in a negative manner (e.g. defensive, hostile).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist facilitated the patient's understanding of the purpose of homework with limited skill (e.g. developed a limited or confusing rationale for importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, therapist gave these limited consideration (e.g. acknowledging a question but failing to answer it).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist facilitated the patient's understanding of the purpose of homework with some degree of skill (e.g. developed a rationale for importance / potential benefits) but some minor problems were evident (e.g. did not relate rationale to the patient's presenting problems / treatment goals / formulation, conveyed the rationale with limited clarity).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, these were acknowledged and therapist made some attempt to discuss and resolve the issues.

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully facilitated the patient's understanding of the purpose of homework (e.g. worked with the patient to develop a clear and succinct rationale for the importance and potential benefits which aligned with the patient's presenting problems / treatment goals / formulation).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, therapist responded openly, fully discussed the issues and (where possible) resolved the issues (e.g. clarified any ambiguities).

4.4. Planning Homework

Ability to work with the patient to ensure they have a clear and detailed understanding of the homework task(s).

1. Limited

- Therapist failed to plan further homework arising from the session or building on previous homework.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Content or practicalities for homework were vague or unclear (e.g. experiment with going out).
- Therapist made little or no attempt to identify potential obstacles. ***Or*** if the patient raised potential obstacles, these were ignored or disregarded.

2. Basic

Major substantive problems in one or more of the following areas:

- Content and practicalities of homework were poorly specified (e.g. go into a specific situation and see if I get anxious).
- Therapist attempted to identify potential obstacles, with limited skill (e.g. asked complicated or vague questions such as "do you think you could complete the task?"). ***Or*** if the patient raised potential obstacles, therapist gave these limited consideration (e.g. acknowledging an obstacle but failing to address it).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Content and practicalities of homework were outlined, though some specific details were lacking (e.g. exactly what, where, when, how, outcome evaluation).
- Therapist made some attempt to prepare the patient for how to complete homework (e.g. talked through a homework example, made sure the task was written down).
- Therapist identified potential obstacles (e.g. asked relevant questions such as "What might get in the way of completing this task?"). ***And*** made some attempt to discuss ways to overcome these obstacles (e.g. asked "how might we overcome that?").

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully worked with the patient to agree a clear and detailed plan of exactly what homework involved (e.g. what, how, where, when, frequency, duration, outcome measurement).
- Therapist thoroughly prepared the patient for how to complete homework (e.g. worked through a homework example together in-session, used guided imagery techniques).
- Therapist worked with the patient to identify any potential obstacles, fully discussed these obstacles ***and*** (where possible) identified ways to overcome them (e.g. using problem solving to break homework into smaller tasks if the patient felt overwhelmed, discussing advantages and disadvantages of homework).

Domain 5 - Measuring Change

A range of methods can usefully be employed to measure change, these include standardised questionnaires, diaries, frequency counts, ratings of duration of event or experience, self-ratings of emotions and cognitions etc. These methods should be employed *across treatment* to measure change in symptoms and movement towards goals over the course of treatment as a whole. However, these methods should also be employed *within treatment sessions* in order to measure the impact of cognitive-behavioural interventions completed within session and for homework (e.g. belief ratings pre/post intervention).

NB In addition to viewing session recording, it is necessary to review any relevant written in-session materials in order to make an informed rating on this item (e.g. completed questionnaires, diaries, belief ratings, treatment goals, etc.).

5.1. Choosing Suitable Measures

Ability to select appropriate, clinically relevant methods for measuring change in symptoms and associated features (beliefs, behaviours, feelings) and movement towards goals.

1. Limited

- Therapist neglected to measure change in symptoms, associated features or movement towards goals.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist selected standardised measures with inadequate reliability, validity or usability, despite the availability of better measures. [1]
- Therapist created inadequate idiosyncratic measures with poor usability (e.g. too complicated for the patient to complete), completely unsuitable response scales (e.g. yes / no to measure symptom change), or undefined targets (e.g. rate your 'feelings').
- Measures were irrelevant and of no clinical use (e.g. depression questionnaire for patient with anxiety, belief ratings of irrelevant cognitions, frequency of irrelevant behaviours).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist selected standardised measures with poor validity, reliability or usability, despite the availability of better measures.
- Therapist created idiosyncratic measures which were difficult to use, had inappropriate response scales or had poorly defined targets.
- Measures were of limited relevance and clinical use (e.g. general anxiety questionnaire for patient with panic disorder, belief ratings of peripheral cognitions).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist selected standardised measures with adequate reliability, validity and usability, though better measures were available.
- Therapist created usable idiosyncratic measures with relevant response scales and outlined targets (e.g. 0 to100 emotional intensity ratings, 0 to 8 or 0 to 10 problem severity ratings).
- Measures were relevant and served a clinical purpose (e.g. panic cognitions questionnaire for patient with panic disorder, belief ratings for related cognitions, rating severity of relevant problem).

4. Advanced

Consistently high level of skill in the following areas:

- Therapist selected the most reliable, valid and usable standardised measures which were available.
- Therapist created idiosyncratic measures with clearly defined targets and highly appropriate response scales (e.g. 0 to 100 rating of clearly specified behaviour).
- Measures were highly relevant and served an important clinical function (e.g. belief ratings for key cognitions, diary provided useful insight into relationship between different aspects of the patient's problems).

[1] Therapists should select the most reliable, valid and usable standard measures which were *available* to them. Thus therapists should *not* be penalised if they were unable to access 'gold standard' questionnaire measures (e.g. due to service constraints, cost).

5.2. Implementing Measures

Ability to administer measures at suitable time points across and within session and to skilfully review, interpret and respond to the information gleaned.

1. Limited

- Therapist neglected to measure change in symptoms, associated features or movement towards goals.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Monitoring schedule was random, haphazard, inappropriate or not clinically useful (e.g. no baseline function to measure change against [e.g. thought record without pre-intervention belief rating], patient overly burdened by completing measures too frequently, progress not regularly checked).
- Therapist failed to review or conducted an extremely brief or cursory review of completed measures which did not provide useful clinical information (e.g. didn't compare post-intervention scores with pre-intervention scores, lack of understanding of completed measures).
- If relevant, therapist made little or no attempt to identify or explore barriers (e.g. difficulty understanding or completing measures, doubts about the purpose of monitoring) or responded to barriers in an unhelpful or negative manner (e.g. scolded patient for incorrectly completing measure).

2. Basic

Major substantive problems in one or more of the following areas:

- Monitoring schedule was unsuitable or of limited clinical use (e.g. measures not used regularly enough or used more frequently than necessary, measures implemented at inappropriate time points).
- Therapist conducted a rushed or superficial review of completed measures which provided little useful clinical information (e.g. therapist showed a limited ability to interpret completed measures).
- If relevant, therapist made some attempt to identify barriers to monitoring and responded to these in an understanding manner. However, therapist made little or no attempt to explore or overcome barriers.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Monitoring schedule was appropriate (e.g. at the beginning, middle and end of treatment, pre/post interventions) and served a useful clinical purpose (e.g. provided insight into changes in patient's symptoms or impact of interventions).
- Therapist adequately reviewed completed measures and attempted to use this information for clinical purposes but some minor problems were evident (e.g. struggled to relate the results to the individual patient).
- If relevant, therapist identified barriers to monitoring, responded in a supportive manner and made some attempt to explore or overcome them.

4. Advanced

Consistently high level of skill in the following areas:

- Monitoring schedule was systematic, appropriate and served an important clinical function (e.g. provided clear and relevant information as to effectiveness of interventions or treatment progress).
- Therapist comprehensively reviewed completed measures and used information gleaned for relevant clinical purposes (e.g. exploring improvements or deteriorations to better understand maintaining factors, highlighting progress to encourage motivation, making decisions about where best to focus therapeutic efforts).
- If relevant, therapist thoroughly identified and explored barriers to monitoring. Where possible therapist worked with the patient to resolve issues (e.g. working through measure with the patient, modifying or simplifying measures).

Domain 6 - Effective Use of Time

6.1. Pace

Ability to pace the session in a manner which is well suited to the therapeutic context and patient's capacity for learning.

1. Limited

Absence of skill or an ***inappropriate performance:***

- Pace consistently much too fast or slow for therapeutic context [1] or patient's optimal learning (e.g. therapist repeatedly belaboured a point long after the patient had grasped the message, important issues were rushed- making it difficult for the patient to grasp or retain key concepts, therapist switched from topic to topic far too rapidly).

2. Basic

Major substantive problems:

- There were a number of times when the pace was too quick or slow for the therapeutic context or patient's optimal learning (e.g. therapist slightly overstressed a point, moved on from emotionally difficult topics too quickly, switched topics before conclusions had been drawn, spoke to rapidly or too slowly).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- For most of the session therapist set a pace which was appropriate given the therapeutic context and patient's learning capacity. However, there were one or two occasions when the pace was a little too quick or slow to be optimally effective (e.g. therapist failed to slow the pace suitably where cognitive difficulties arose, moved the patient along before they had fully understood concepts).

4. Advanced

Consistently high level of skill:

- Therapist was responsive to and consistently matched the pace of the session to the therapeutic context and patient's learning capacity (e.g. gave more time and attention to issues where cognitive difficulties arose, allowed the patient to experience emotions rather than rushing onto an intervention / another topic).

[1] Therapeutic context = patient's presenting problems, goals, stage of treatment, formulation, presentation etc.

6.2. Time Management

Ability to manage time within the session in a balanced and efficient manner.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Inefficient, unbalanced allocation of time meant a number of important issues were not covered at all (e.g. not enough time left to plan homework, summarise session content, exchange feedback etc.).
- Omitted issues were given limited or no acknowledgement.
- Session ran very significantly over ***or*** under scheduled time without clinical justification [1] (e.g. 50 minute session lasted 70 minutes or ended after 30 minutes with no clinical justification).
- Overly rigid adherence to a schedule was unhelpful to the patient (e.g. comments such as "We said we'd move on after 10 minutes so we need to move on now").

2. Basic

Major substantive problems in one or more of the following areas:

- Less than optimal use of time meant some important issues were not covered ***or*** were inadequately covered (e.g. insufficient time to review learning from an intervention or to plan how to take the work forward).
- Omitted issues were acknowledged but were not rescheduled.
- Session ran a little over ***or*** under scheduled time without sufficient clinical justification (e.g. 50 minute session lasted for 60 minutes or ended after 40 minutes).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist's use of time was generally well balanced and efficient, meaning most important issues were adequately covered. However, to make optimal use of time, it would have been necessary to spend a little less or a little more time on some issues (e.g. not quite enough time left at end of session to plan homework thoroughly).
- If issues were not covered, this was somewhat appropriate given the therapeutic context (e.g. focusing the entire session on one agenda item which was more distressing / significant than anticipated).
- Omitted issues were rescheduled for a different time or format (e.g. next session, homework).
- Length of session was largely as scheduled ***or*** there was some clinical justification for deviating from planned schedule.

4. Advanced

Consistently high level of skill in the following areas:

- Consistently balanced and efficient allocation of time meant all important issues were thoroughly covered. ***Or*** if issues were not covered, this was necessary given the therapeutic context and this decision was made openly and explicitly with the patient.
- Omitted issues were rescheduled for a different time or format (e.g. next session, homework).
- Length of session was as scheduled ***or*** there was clear clinical justification for deviating from planned schedule.

[1] Examples of clinical justification for running over / under time = significant patient distress, expression of risk etc.

6.3. Maintained Focus

Ability to maintain focus on important issues, whilst demonstrating appropriate flexibility in response to unanticipated issues.

1. Limited

Absence of skill or an ***inappropriate performance***:

- Session lacked any clear focus and appeared aimless (e.g. patient dominated session completely and therapist simply followed the patient's lead- making little or no attempt to focus session back to appropriate topics, general unstructured 'chat' which did not appear to have any specified aims or reach any particular conclusions).
- ***Or*** overly rigid focus was unhelpful to the patient (e.g. therapist dismissed potentially important issues raised by the patient or disregarded patient's distress because it did not relate to the agenda / fit with the session plan).

2. Basic

Major substantive problems:

- Therapist showed limited skill in maintaining focus on appropriate topics, resulting in frequent repetitions, unproductive digressions or unjustified discussions of peripheral issues (e.g. therapist frequently switched topics without clear justification, was slow to respond to digressions, was abrupt or tactless when re-directing discussion to appropriate topics).
- ***Or*** therapist focused on established priorities in a rigid, inflexible or mechanistic manner (e.g. failed to acknowledge or respond to unanticipated issues).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies*** in the following areas:

- Therapist largely maintained focus on session priorities (i.e. agenda items). However, if repetitions, unproductive digressions or unnecessary discussions of peripheral issues arose, these were re-directed back to an appropriate topic in a timely manner and with some degree of sensitivity.
- ***And/or*** if unanticipated issues arose, therapist acknowledged these and showed appropriate flexibility (e.g. allowing the session to diverge from planned topics to make maximum therapeutic progress or maintain client engagement / therapeutic relationship).

4. Advanced

Consistently high level of skill in the following areas:

- Therapist consistently maintained focus on session priorities (i.e. agenda items). However, if repetitions, unproductive digressions or unnecessary discussions of peripheral issues arose, these were quickly acknowledged, sensitively evaluated and skilfully steered back to an appropriate topic (e.g. "We seem to be straying from our plan for the session- is this something that it is important that we focus on? Or would it be more useful to return to [planned focus]?").
- ***And*** if unanticipated issues arose, therapist acknowledged these and showed appropriate flexibility. Any decisions to change focus were negotiated openly and explicitly with the patient.

Domain 7 - Fostering Therapeutic Relationship

7.1. Interpersonal style

Ability to embody a positive interpersonal style which is congruent with the therapeutic context.

1. Limited

Absence of skill or an ***inappropriate performance***:

- Therapist consistently demonstrated a negative interpersonal style (e.g. impatient, aloof, insincere, patronising, distant, distracted, cold, uncaring, intimidating, reprimanding, hostile, demeaning etc.).
- ***Or*** therapist's interpersonal style was poorly suited to the patient's therapeutic context [1] (e.g. therapist praised patient excessively, demonstrated high level of warmth to a highly avoidant / socially anxious patient thus making them feel uncomfortable).

2. Basic

Major substantive problems:

- Therapist generally demonstrated a reasonable interpersonal style (e.g. was friendly, positive, constructive etc.) which was appropriate given the patient's therapeutic context. However, at times therapist showed limited flexibility in adapting their interpersonal style in response to therapeutic context (e.g. insensitive use of humour which invalidated the patient's feelings, inappropriate self-disclosure, inappropriate disclosure of ideas about the patient's problems [e.g. told patient with eating disorder idea for final weight too early in treatment]).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- Therapist generally demonstrated a positive interpersonal style (e.g. open, warm, respectful, genuine etc.) which was suited to individual patient's therapeutic context (e.g. friendly but not overly familiar, optimistic without being excessively positive).

4. Advanced

Consistently high level of skill:

- Therapist consistently demonstrated a positive interpersonal style (e.g. embodied and effectively communicated sincerity, open-minded curiosity, genuineness, honesty, sensitive concern, optimism, professionalism, encouragement, unconditional positive regard etc.) ***and*** seamlessly adapted their interpersonal style so that it was congruent with the therapeutic context (e.g. optimal levels of warmth, concern etc. given the individual patient's current needs).

[1] Therapeutic context = patient's interpersonal style, presenting problems, history, stage of treatment, formulation & presentation.

7.2. Empathic Understanding

Ability to accurately grasp the content and emotional tone of the patient's viewpoint (i.e. their understanding of themselves and the world around them) and to sensitively and appropriately conveying this understanding.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist frequently projected their own attitudes, conventional attitudes, or attitudes derived from particular theoretical systems onto the patient.
- Therapist repeatedly ignored or completely misinterpreted obvious elements of the patient's viewpoint.
- Therapist negated, undermined or was critical of the patient's viewpoint (e.g. appeared disapproving, disparaging or judgemental) or over-identified with patient's perspective (i.e. was not objective).
- Therapist seemed unwilling or unable to revise their understanding of the patient's viewpoint in response to new clinical information or in-session changes.

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist accurately reflected more obvious, verbal cues (i.e. what patient explicitly said), but repeatedly ignored or completely misinterpreted more subtle emotional or nonverbal cues (e.g. tone of voice, body language etc.) ***and/or*** responded inappropriately (e.g. excessive empathy).
- Therapist found it difficult or seemed reluctant to revise their understanding of the patient's viewpoint in response to new clinical information or in-session changes.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist picked up on most verbal and more subtle emotional or nonverbal cues to gain an understanding of the patient's viewpoint but at times therapist slightly misinterpreted patient or focused on peripheral aspects of what patient said.
- Therapist attempted to convey their understanding of the patient's viewpoint but with limited skill (e.g. restating patient's words unnecessarily, overly long summaries, unsuitable tone of voice, intellectualising issues).

4. Advanced

Consistently high level of skill in the following areas:

- Therapist consistently picked up on verbal, emotional and nonverbal cues to gain an accurate and intimate understanding of the patient's viewpoint.
- Therapist consistently used appropriate verbal (e.g. paraphrasing, summarising, reflecting) and non-verbal responses (e.g. tone of voice, facial expression) to convey their understanding of the patient's viewpoint. This helped the patient feel listened to and understood on an emotional and cognitive level throughout the session (e.g. therapist conveyed the importance of working on issues which were important to the patient when prioritising agenda items, acknowledged participants bravery in undertaking anxiety provoking tasks etc.)

7.3. Collaboration

Ability to encourage the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation.

1. Limited

Absence of skill or an ***inappropriate performance:***

- Therapist repeatedly made very little or no attempt to work with or involve the patient (e.g. therapist appeared to come to the session with pre-planned ideas that were not open to discussion, focused on issues that were not a priority for the patient, told the patient what to do without requesting any input, ignored or disregarded issues raised by the patient, did not elicit or value the patient's ideas, did not share information with the patient, tried to control or dominate the patient, seemed to be 'cross-examining' or arguing with the patient).
- ***Or*** therapist was frequently too passive and offered the patient too much responsibility (e.g. provided no input where the patient was not able to do so independently, inappropriately followed the patient's lead).

2. Basic

Major substantive problems:

- Therapist made some attempt to involve and work with the patient. However, on a number of occasions, the therapist offered the patient too much or too little responsibility given the stage of therapy or patient's presentation [1] (e.g. therapist expected the patient to take the lead in very early sessions, did not give the patient enough choice, debated with or lectured the patient, paid minimal attention to the patient's ideas, did not give the patient sufficient space and time to think, used overly complicated or vague questions such as "What shall we talk about today?", "What next?").

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- For most of the session, the therapist worked effectively with the patient and helped them to share responsibility in a manner appropriate to the stage of therapy and patient's presentation (e.g. therapist offered the patient choices, paid attention to the patient's ideas, gave the patient some time and space to think, asked relevant questions such as "What do you think the most important things are for us to cover today?", "How might we test that out?"). However, there were one or two occasions when the therapist and patient did not appear to be working as a team (e.g. therapist was too instructive or passive).

4. Advanced

Consistently high level of skill:

- Therapist consistently encouraged the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation (e.g. therapist offered the patient appropriate choices, elicited and incorporated the patient's ideas, gave the patient plenty of time and space to think, asked skilful questions such as "How might you take this forward next week?", "What's the most important thing you need to do to keep the progress going?", "You have mentioned X, Y and Z. What do you think is most important to cover today, given that we are nearing the end of therapy?").

[1] NB it may be appropriate for therapists to take more responsibility in initial sessions, but responsibility should be increasingly transferred to the patient as therapy progresses, depending on patient's presentation.

Domain 8 - Effective Two-way Communication

8.1. Patient Feedback

Ability to elicit, explore and respond to feedback about the patient's understanding of and reaction to all aspects of session.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to check the patient's understanding of and reaction to the session ***or*** discouraged the patient from giving feedback.
- Therapist repeatedly ignored or disregarded patient feedback ***or*** responded in a negative manner (e.g. became defensive or argumentative).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist inconsistently checked the patient's understanding of and reaction to the session, meaning that feedback was not elicited for a number of important aspects of the session (e.g. formulation, homework, interventions etc.).
- Therapist used overly complicated or vague questions to elicit patient feedback (e.g. asked "Is that OK?", "Does everything seem to be going well?").
- Therapist failed to ask for patient feedback about any negative reactions to session, despite clear need to do so (e.g. patient's comments indicated potential concerns).
- Therapist acknowledged patient feedback but made little or no attempt to discuss or respond to it.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist regularly checked the patient's understanding of and reaction to the session. However, feedback was not elicited for one or two important aspects of the session (e.g. formulation, homework, interventions etc.).
- Therapist used some appropriate questions to elicit patient feedback (e.g. asked "How do you think the session is going so far", "Does this homework seem manageable?", "Is the homework clear?", "How does the idea of doing X strike you?", "How do you feel about attempting X?").
- Therapist made some attempt to elicit potentially negative reactions to session (e.g. doubts, scepticism or concerns) as well as positive reactions (e.g. asked "Do you have any concerns about X", "You seem a little uncertain, is there anything you feel unsure about?").
- Therapist discussed patient feedback and made some attempt to respond to patient's feedback (e.g. answered questions, revised focus in response to patient's concerns).

5. Advanced

Consistently high level of skill in the following areas:

- Therapist consistently checked the patient's understanding of and reaction to all aspects of session (e.g. formulation, homework, interventions, agenda, measuring change, patient's response to therapist etc.).
- Therapist consistently used skilful questions to elicit patient feedback (e.g. asked "From your point of view, why is it important to do X?", "What might you learn from doing X?").
- Therapist actively encouraged the patient to express positive and negative reactions to all aspects of session in a warm and friendly manner which invited exploration (e.g. asked "What has been least helpful today?", "Is there anything I've said that didn't make sense?").
- Therapist thoroughly and openly discussed patient feedback in a non-judgemental, supportive and sensitive manner. ***And*** therapist responded appropriately to patient's feedback (e.g. adapted interventions, adjusted their own behaviour, clearly and fully answered questions).
- If relevant, therapist picked up on subtle cues (e.g. tone of voice, hesitation etc.) to identify difficulties giving authentic feedback (e.g. patient responded in accordance with what they thought therapist wished to hear) and openly discussed these with the patient.

8.2. Reflective Summaries

Ability to work with the patient to reflect upon and summarise salient session content in order to facilitate joint understanding of crucial therapeutic material and consolidate key learning.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to encourage the patient to review and reflect on session content.
- Therapist made little or no attempt to summarise important information or provided very minimal, highly inappropriate or inaccurate summaries.

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist prompted the patient to review and reflect on session content ***and/or*** provided summaries either too rarely (e.g. only once or twice) ***or*** too frequently (e.g. continually providing summaries or asking the patient to reflect on session content).
- Reflective summaries were of limited therapeutic benefit as they were superficial ***and/or*** focused on irrelevant information or peripheral learning points.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- At the beginning and end of session, and a number of times within session, therapist prompted the patient to review and reflect on session content (e.g. asked relevant questions such as “what’s the main thing you’d like to remember about this session") ***and*** provided accurate capsule summaries (e.g. “so to recap, at the beginning of the session we discussed X and the impact it had on your mood. Then we talked about Y etc.”).
- Reflective summaries focused on relevant information and learning points and thus were of some use in helping to reinforce understanding of important therapeutic material and consolidate key learning.

4. Advanced

Consistently high level of skill in the following areas:

- At the beginning and end of session, and at appropriately regular intervals within session, therapist used skilful questioning to elicit in-depth patient reflections on and summaries of session content ***and*** provided clear, succinct and meaningful capsule summaries.
- Reflective summaries skilfully drew together salient information and key learning points in a way which helped to reinforce and facilitate joint understanding of crucial therapeutic material, consolidate key learning and enabled the patient to gain new insights and make therapeutic shifts.

Global Performance Rating

How would you rate the therapist's overall performance within this session?

1. Limited

- Therapist demonstrates a significant absence of skill or an inappropriate performance which is likely to have negative therapeutic consequences.

2. Basic

- Therapist's performance is somewhat appropriate with some degree of skill evident.
- However, major substantive problems are evident.

3. Good

- Therapist demonstrates a good degree of skill with no major problems.
- However, minor problems or inconsistencies are evident in the therapist's performance.

4. Advanced

- Therapist consistently demonstrates a high level of skill with only very few and very minor problems.

Patient Complexity Rating

How complex do you feel the patient was to work with? Issues that may influence complexity include level of patient motivation, level of engagement, the presence of social or environmental problems, the number of problems, the severity of problems, and suitability for CBT.

1. Very straightforward

2. Somewhat Straightforward

3. Somewhat Complex

4. Very Complex

References

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. doi: 10.1017/S1352465801004040

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*(2), 218-244. doi: 10.1037/0033-2909.90.2.218

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012a). Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, 1-11. doi: 10.1080/10503307.2012.716528

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

Keen, A., & Freeston, M. (2008). Assessing competence in cognitive-behavioural therapy. *British Journal of Psychiatry, 193*(1), 60-64. doi: 10.1192/bjp.bp.107.038588

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

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognititve and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use (3rd ed.)*. Oxford: Oxford University Press.

Vallis, T., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54*(3), 381-385. doi: 10.1037/0022-006X.54.3.381

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), 620-630. doi: 10.1037/0022-006X.61.4.620

Weck, F., Bohn, C., Ginzburg, D. M., & Ulrich, S. (2011). Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire sessions. *Psychotherapy Research, 21*(6), 658-669. doi: 10.1080/10503307.2011.602751

Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *Journal of Nervous and Mental Disease, 199*(4), 276-279. doi: 10.1097/NMD.0b013e3182124617

Young, J., & Beck, A. T. (1980). Cognitive Therapy Scale Rating Manual. Retrieved from www.beckinstitute.org 16.10.2013

Assessment of Core CBT Skills (ACCS)

FEEDBACK FORM

Formative Feedback Form

Version 2- October 2013

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus and was conducted in affiliation with the University of Oxford's Department of Psychiatry and the Oxford Cognitive Therapy Centre (OCTC).

Therapist's Name: _____ **Assessor's Name:** _____

Date of Submission: _____

For each item, please assign a rating of:

1. Limited	2. Basic	3. Good	4. Advanced

Key strengths and learning needs:
For each domain, draw on the exemplar therapist behaviours provided and the session material being assessed to provide examples of what made some aspects of the therapist's performance particularly successful what made some aspects of the therapist's performance less successful. Suggest alternative, more skilful way of working within any weaker domains and highlight ways in which the therapist could further develop their skills in this area.

Domain 1 - Agenda Setting

1.1. Suitable Items	
1.2. Feasible Agenda	
Total Domain Score *(average item score)* *Click to Calculate*	

See user manual page 7 – 8

Key strengths and learning needs:

Domain 2 - Formulation

2.1. Coherent Formulation	
Total Domain Score *(average item score)* Click to Calculate	

See user manual page 9

Key strengths and learning needs:

Domain 3 – CBT Interventions

3.1. Appropriate Intervention Targets	
3.2. Choosing Suitable Interventions	
3.3. Rationale for Interventions	
3.4. Implementing Interventions	
3.5. Reviewing Interventions	
Total Domain Score *(average item score)* *Click to Calculate*	

See user manual page 10 – 14

Key strengths and learning needs:

Domain 4 – Homework

4.1. Reviewing Homework	
4.2. Choosing Suitable Homework	
4.3. Rationale for Homework	
4.4. Planning Homework	
Total Domain Score *(average item score)* *Click to Calculate*	

See user manual page 15 – 18

Key strengths and learning needs:

Domain 5 – Measuring Change

5.1. Choosing Suitable Measures	
5.2. Implementing Measures	
Total Domain Score *(average item score)* *Click to Calculate*	

See user manual page 19 – 20

Key strengths and learning needs:

Domain 6 – Effective Use of Time

6.1. Pace	
6.2. Time Management	
6.3. Maintained Focus	
Total Domain Score *(average item score)* *Click to Calculate*	

See user manual page 21 – 23

Key strengths and learning needs:

Domain 7 – Fostering Therapeutic Relationship

7.1. Interpersonal Style	
7.2. Empathic Understanding	
7.3. Collaboration	
Total Domain Score *(average item score)* *Click to Calculate*	

See user manual page 24 – 26

Key strengths and learning needs:

Domain 8 – Effective Two-way Communication

8.1. Patient Feedback	
8.2. Reflective Summaries	
Total Domain Score *(average item score)* *Click to Calculate*	

See user manual page 27 – 28

Key strengths and learning needs:

Overall Session Score

Total Score (*add total domain scores*)	*Click to Calculate*	*(range = 8 to 32)*

Patient Complexity Rating

How complex do you feel the patient was to work with?

1. Very Straightforward **2. Somewhat Straightforward** **3. Somewhat Complex** **4. Very Complex**

Global Performance Rating

Therapist's overall performance within the session was:

1. Limited **2. Basic** **3. Good** **4. Advanced**

Please explain *why* you chose this global performance rating:

Skills Profile

Therapist's Key Strengths

What were the therapist's key strengths within the session recording? How can the therapist continue to build upon these strengths in the future?

Therapist's Learning Needs

Based on the session recording, what skills does the therapist need to concentrate on further improving and developing? How could the therapist further develop their skills in these areas?

CHAPTER FIVE:

A pilot study examining psychometric properties of the Assessment of Core CBT Skills (ACCS) scale

Chapter Overview

The Assessment of Core CBT Skills (ACCS) is a novel rating scale for assessing the competence with which therapists deliver Cognitive Behavioural Therapy (CBT). Development of the ACCS involved three key stages: (1) scale development, (2) expert review, and (3) scale evaluation. The final phase of this process, scale evaluation, is further broken down into two discrete studies: a pilot study examining the psychometric properties of the ACCS and a focus group examining the usability and utility of the ACCS. This chapter presents results from the initial pilot study that investigates the psychometric properties of the ACCS rating scale in 'real world' CBT training and routine practice contexts. This pilot evaluation shows that both the self-rated and assessor-rated versions of the ACCS demonstrate good internal consistency, inter-rater reliability, and discriminant validity. Thus the ACCS is suitable for use in clinical practice, training settings and research studies and can be used as a self-rating tool as well as an assessor-rated tool. In addition, the ACCS was found to be correlated with, but distinct from the Revised Cognitive Therapy Scale (CTS-R) and was comparable to the CTS-R in terms of internal consistency and discriminant validity and may have advantages over the CTS-R in terms of inter-rater reliability. Thus the ACCS may provide an acceptable alternative to the CTS-R, with updated criteria and broader applicability.

Acknowledgements:

With many thanks to therapists and assessors within the Oxford Cognitive Therapy Centre and the First Step service for participating in the study and to David Slingo, Sarah Rakovshik, Richard Thwaites, Laura Lockhart, David Sandford, and Ryan Askey-Jones for supporting this research within their services.

Contents

Introduction 337

Method 337

Participating CBT Centres 337

Ethics 338

Participants 339

Patients 340

Rating Procedure 343

OCTC 343

First Step 345

Materials 345

Cognitive therapy scale-revised 345

ACCS 346

Session recordings 347

Data analysis 347

Descriptive statistics 347

Internal consistency 348

Inter-rater reliability 348

Discriminant validity 349

Convergent validity 349

Results 349

Descriptive statistics 349

Assessor-rated ACCS 349

Self-rated ACCS 355

Internal consistency 360

Correlation between domain items 360

Item-total correlation 361

Cronbach's alpha 363
Inter-rater reliability 365
Discriminant validity 368
Convergent validity 368
Discussion 368
Summary of findings 368
Limitations and future directions 371
Conclusions 372
References 373
Appendices 375
Appendix 1. Ethical Consent Approval 375
Appendix 2. Invitation to Participate 376
Appendix 3: Participant Information Sheet: Oxford Cognitive Therapy Centre 377
Appendix 4: Consent Form: Oxford Cognitive Therapy Centre 380
Appendix 5: Participant Information Sheet: First Step 381
Appendix 6: Cognitive Therapy Scale-Revised 384
Appendix 8: Supporting information cover sheet to accompany therapists' session recording 402

Introduction

Chapters three and four of this thesis outlined the initial development and refinement of a novel Cognitive Behavioural Therapy (CBT) competence rating scale: the Assessment of Core CBT Skills (ACCS). The ACCS provides a framework for assessors to deliver formative and summative feedback regarding therapists' performance within observed CBT treatment sessions and for therapists to rate and reflect on their own performance. The final phase of the scale development process involves an evaluation of the ACCS. This evaluation phase can be further broken down into two discrete studies: a pilot study examining the psychometric properties of the ACCS and a focus group examining the usability and utility of the ACCS. Chapter six outlines findings from the focus group, whilst this chapter reports results from the pilot study.

The pilot study was designed to investigate the psychometric properties of the ACCS rating scale in 'real world' CBT training and routine practice contexts. The primary aim of the study was to evaluate the reliability and validity of the assessor-rated and self-rated ACCS scale in order to determine whether the ACCS is suitable for use in clinical practice, training settings and research studies. Examining the psychometric properties of rating scales is also useful in highlighting any items which require modification or omission (DeVellis, 2012) and thus will also identify any further revisions that are required. The psychometric properties the study examines are: the ability of items to discriminate levels of competence (i.e. their ability to adequately capture and differentiate between different levels of competence), internal consistency (i.e. the degree to which items assess the same underlying construct), inter-rater reliability (i.e. the level of agreement between different assessors' ratings on the ACCS), discriminant validity (i.e. whether the ACCS is sensitive to improvements in competence), and convergent validity (i.e. whether the ACCS correlates well with a previous measure of competence: the Revised Cognitive Therapy Scale [CTS-R]). A secondary aim of the study was to compare the psychometric properties of the ACCS to the CTS-R (Blackburn et al., 2001), a currently used measures of CBT competence. Given that the ACCS aims to further develop and build upon the CTS-R it is expected that the ACCS would be at least comparable to the CTS-R in terms of internal consistency, inter-rater reliability and discriminant validity.

Method

Participating CBT Centres

Two centres participated: the Oxford Cognitive Therapy Centre (OCTC) and First Step. Within OCTC participants were recruited from the 2013/14 intake of the Postgraduate Diploma in

Cognitive Behavioural Therapy course (PGDip) run in collaboration with the University of Oxford's Department of Continuing Education (for a detailed description of the course see McManus, Westbrook, Vazquez-Montes, Fennell, & Kennerley, 2010). This well-established course provides training in the use of CBT with a variety of client groups and disorders. Participants were also recruited from the First Step service in Cumbria, an NHS Increasing Access to Psychological Therapies treatment service for people with mild to moderate depression and anxiety (see http://www.cumbriapartnership.nhs.uk/first-step.htm).

It was initially intended that ratings would be collected from a wider range of routine practice and CBT training centres. Sixteen CBT courses or treatment centres were approached about participating in the study. These courses/centres were approached because senior employees had participated in the ACCS evaluation study outlined in chapter four. Two courses/centres (12.5 %) agreed to participate (First Step and the OCTC). Eight courses/centres (50 %) did not respond to the invitation, three (18.75 %) declined to participate due to the increased burden on assessors and three (18.75 %) provided the opportunity for assessors to decide whether to participate or not, resulting in three assessors providing 12 ACCS ratings. Due to the poor response rate, and thus the very small amount of data obtained, this additional data was not included. This was primarily because the intention in collecting additional data was to enable factor analysis to be completed (requiring a sample of 110 ratings from *independent* therapists) and the additional data collected was not sufficient to allow this analysis to be completed. Thus, including these extra ratings would have been of no benefit.

Ethics

Ethical permission to collect data within OCTC was obtained by the University of Oxford Research Ethics Committee (see Appendix one). For the purposes of ethical approval, participants were considered to be the therapists enrolled in the OCTC PGDip CBT. All therapists were provided with a written invitation to participate (see Appendix two) and a copy of the information sheet (see Appendix three). In addition, the primary researcher (Kate Muse) spoke to the therapists as a group to explain what participation would involve and answer any questions they had. All therapists provided written informed consent prior to taking part (see Appendix four). In addition, recordings were subject to informed consent (i.e. patients agreed that the session could be recorded and listened to by an OCTC supervisor as part of the therapist's training).

The University of Oxford's clinical trials and research governance research services designated data collection within First Step as service evaluation, rather than research. This was because the recordings were submitted and rated within the First Step service as part of a service

level review, with only anonymised ratings of therapists' performance being submitted to the primary researcher in order to evaluate the ACCS. Consequently, the study did not require approval by an ethics committee. However, approval was given by Cumbria Partnership NHS Foundation Trust for the First Step service to collect and rate session recordings for service evaluation purposes and the recorded patients consented to the session being recorded and viewed by a member of the First Step service as part of their supervision / professional development. Additionally, both the First Step service lead and the clinical audit team for the local area (Copeland) provided approval for anonymised ACCS ratings to be submitted for the purposes of the ACCS pilot study. To ensure that all therapists within First Step were aware of the ACCS pilot study and had the opportunity to raise any questions, an information sheet was circulated to all staff members (see Appendix five).

Participants

Participants were: 1) therapists who completed self-ratings and submitted recordings for assessors to rate (herein referred to as 'therapists') and 2) supervisors or senior therapists who rated the submitted recordings (herein referred to as 'assessors'). Within First Step, 35 CBT practitioners participated as therapists and 12 CBT practitioners working in a managerial and/or supervisory capacity participated as assessors. These groups were not mutually exclusive. Thus, seven participants who worked in a supervisory capacity completed assessments for their supervisees as well as submitting a recording of one of their own treatment sessions to be rated by a senior psychotherapist (i.e. participated as both therapists and assessors).

Within OCTC the 23 therapists enrolled on the PGDip in CBT were invited to participate as therapists. They were separately asked to consent to (i) their supervisors rating their session recordings using the ACCS in addition to the CTS-R and (ii) to completing self-ratings using the ACCS in addition to the CTS-R separately. Thus, they could agree to participate in either both or only one aspect of the study. Twenty therapists (86.96%) agreed that their supervisors could rate their recordings using the ACCS and nineteen (82.60 %) agreed to complete self-ratings using the ACCS. Eleven senior CBT practitioners employed as supervisors on the PGDip by OCTC participated as assessors.

Demographic characteristics for therapists and assessors within OCTC and First Step are outlined in Table one. Within OCTC, therapists had been practicing CBT for significantly less time

than assessors (U = 0.00, p = <.001) [1] and the majority of therapists had not yet treated any CBT cases, whereas all assessors had treated more than 200 CBT cases. Within First Step, therapists had also been practicing CBT for significantly less time than assessors (U = 97.50, p = <.05) however, the majority of therapists and assessors had treated more than 200 CBT cases. Assessors within OCTC had been practicing CBT for significantly more time than assessors within First Step (U = 4.00, p = <.001). However, the majority of assessors in both sites had treated more than 200 CBT cases. Therapists within First Step had been practicing CBT for significantly more time than therapists within OCTC (U = 95.00, p = <.001) and the majority of therapists within First Step had treated over 200 CBT cases, whereas the majority of therapists within OCTC had not yet treated any CBT cases.

Patients

Due to patient confidentiality, little information about the patients in the submitted treatment sessions was available. However, patients' primary presenting problem(s) were identified by therapists within OCTC and by assessors within First Step. Assessors from both services also rated the complexity of patients in the recordings on a one to four scale (1 - very straightforward, 2 - somewhat straightforward, 3 - somewhat complex, 4 - very complex). This information is presented for recordings submitted in First Step and OCTC in Table two. Patients in both sites presented primarily with depression and anxiety disorders. There was no significant difference between the complexity of the patients in First Step and OCTC (U = 1209.00, p = .55) [1], with patients in both sites being rated as somewhat straightforward.

[1] As the data was not normally distributed, non-parametric Mann-Whitney U-tests were performed to examine the difference between the years of experience held by therapists and assessors in OCTC and First Step and between complexity of patients in OCTC and First Step.

Table 1.
Demographic characteristics for therapist and assessor participants in First Step and the Oxford Cognitive Therapy Centre.

	First Step		**OCTC** [1]	
	Therapists	**Assessors**	**Therapists**	**Assessors**
	n = 25 *(5 missing [2])*	*n* = 12	*n* = 17 [3] *(3 missing [2])*	*n* = 11
Gender - % female	80.00 %	75.00 %	76.00 %	100 %
Years practicing CBT Mean (*SD*) [range]	5.60 (2.69) [2 – 15]	8.13 (3.66) [5 – 15]	3.00 (1.95) [0 – 10]	19.17 (5.70) [13 – 30]
Number of CBT cases:				
% treated 0 cases	0.00 %	0.00 %	58.82 %	0.00 %
% treated 1 - 50 cases	3.33 %	0.00 %	29.41 %	0.00 %
% treated 50 – 200 cases	16.67 %	16.67 %	11.76 %	0.00 %
% treated > 200 cases	80.00 %	83.33 %	0.00 %	100 %
BABCP accredited CBT therapists %	100 %	100 %	0 %	100 %
Profession: %				
Clinical psychologist	0.00 %	8.33 %	15.00 %	54.55 %
Counselling psychologist	0.00 %	0.00 %	15.00 %	0.00 %
Psychotherapist	0.00 %	0.00 %	10.00 %	0.00 %
Nurse	23.33 %	16.67 %	30.00 %	27.27 %
Psychiatrist	0.00 %	0.00 %	5.00 %	9.09 %
CBT therapist	70.00 %	50.00 %	0.00 %	9.09 %
General practitioner	0.00 %	0.00 %	15.00 %	0.00 %
Speech and language therapist	0.00 %	0.00 %	5.00 %	0.00 %
Physiotherapist	0.00 %	0.00 %	5.00 %	0.00 %
Occupational Therapist	0.00 %	8.33 %	0.00 %	0.00 %
Social worker	4.29 %	16.67 %	0.00 %	0.00 %

[1] OCTC = Oxford Cognitive Therapy Centre.

[2] Five therapists within First Step and three therapists within OCTC did not complete demographics.

[3] One therapist did not consent to completing self-ratings using the ACCS but did consent to their supervisor completing ACCS ratings. For parsimony, this person is included here within the therapist sample.

Table 2: Characteristics for patients in the submitted session recordings in First Step and the Oxford Cognitive Therapy Centre.

	First Step n = 35	**OCTC** [1] *n* = 76
Main presenting problem: *n* (%)		
Depression	11 (31.43 %)	12 (15.79 %)
Anxiety	0 (0.00 %)	7 (9.21 %)
Depression / low mood and anxiety	2 (5.71 %)	11 (14.47 %)
Panic disorder with / without agoraphobia	1 (2.86 %)	10 (13.16 %)
Panic disorder with / without agoraphobia and depression	0 (0.00 %)	2 (2.63 %)
Social anxiety	7 (20.00 %)	5 (6.58 %)
Social anxiety and depression	2 (5.71 %)	3 (3.95 %)
Social anxiety and low self-esteem	0 (0.00 %)	1 (1.32 %)
Low self-esteem	0 (0.00 %)	3 (3.95 %)
Obsessive compulsive disorder	2 (5.71 %)	2 (2.63 %)
Obsessive compulsive disorder / health anxiety	0 (0.00 %)	1 (1.32 %)
Health anxiety / depression	1 (2.86 %)	0 (0.00 %)
Health anxiety	1 (2.86 %)	2 (2.63 %)
Specific phobia	0 (0.00 %)	5 (6.58 %)
Chronic pain	0 (0.00 %)	2 (2.63 %)
Post-traumatic stress disorder	1 (2.86 %)	3 (3.95 %)
Generalised anxiety disorder	6 (17.14 %)	3 (3.95 %)
Anger	1 (2.86 %)	0 (0.00 %)
Missing information	0 (0.00 %)	4 (5.25 %)
Supervisor complexity rating: Mean, *SD* [range]	2.23 (.70) [1 – 3.5]	2.14 (0.81) [1 – 4] [2]
Senior therapist complexity rating: Mean, *SD* [range]	2.09 (.56) [1 – 3]	N/A

[1] OCTC = Oxford Cognitive Therapy Centre

[2] *n* = 74 (2 ratings missing)

Rating Procedure

The rating procedures followed within OCTC and First Step and outlined below.

OCTC

As part of their training, therapists were required to submit six recordings of CBT sessions with patients, one at the beginning and one at the end of each of the three terms. Data was collected from the first two terms only (providing up to four recordings per therapist: see Figure one). This is because in the first two terms therapists were instructed to select 'straightforward' patients, whereas in the third term therapists were asked to select more complex cases. Previous research examining competence within the context of OCTC's PGDip in CBT has suggested that this increase in complexity of cases may cause competence ratings to decrease in term three of the course (McManus et al., 2010). Thus, term three ratings were not included in order to limit the potential confounding variable of changes in patient complexity. Recordings were selected by therapists and were rated using the Cognitive Therapy Scale-Revised (CTS-R: Blackburn et al., 2001) by their course supervisors. For the purposes of this study, assessors were asked to complete ACCS ratings in addition to (and prior to) CTS-R ratings. Eleven assessors completed these ratings (one assessor completed two ratings, two assessors completed four ratings, one assessor completed five ratings, two assessors completed six ratings, one assessor completed eight ratings, three assessors completed ten ratings, and one assessor completed 11 ratings). In addition, 20 session recordings (26.32% of the total sample) were selected at random and double marked. These sessions reflected a spread of competency (total ACCS scores assigned by the original assessor ranged from 14.62 to 27.75). The 20 recordings were originally rated by ten different assessors from OCTC and all of the second ACCS ratings were completed by one of the scale authors (Sarah Rakovshik), who was blind to the stage of training and did not have access to the first assessors' ratings.

Therapists received formative feedback about their performance both verbally and in writing, as well as numerical scores on the CTS-R and ACCS. Therapists also completed a CTS-R self-rating of their performance within the recorded session. For the purposes of this study, therapists were asked to complete ACCS self-ratings ratings in addition to (and prior to) CTS-R self-ratings ratings. These were submitted to supervisors alongside the recordings. However, supervisors were instructed to complete their ACCS and CTS-R ratings before viewing therapists' self-ratings. Assessors and therapists were provided with a copy of the ACCS manual (see Appendix six in chapter four). Assessors and therapists were also encouraged to contact the primary researcher (Kate Muse) if they had any questions or queries about how to use the ACCS.

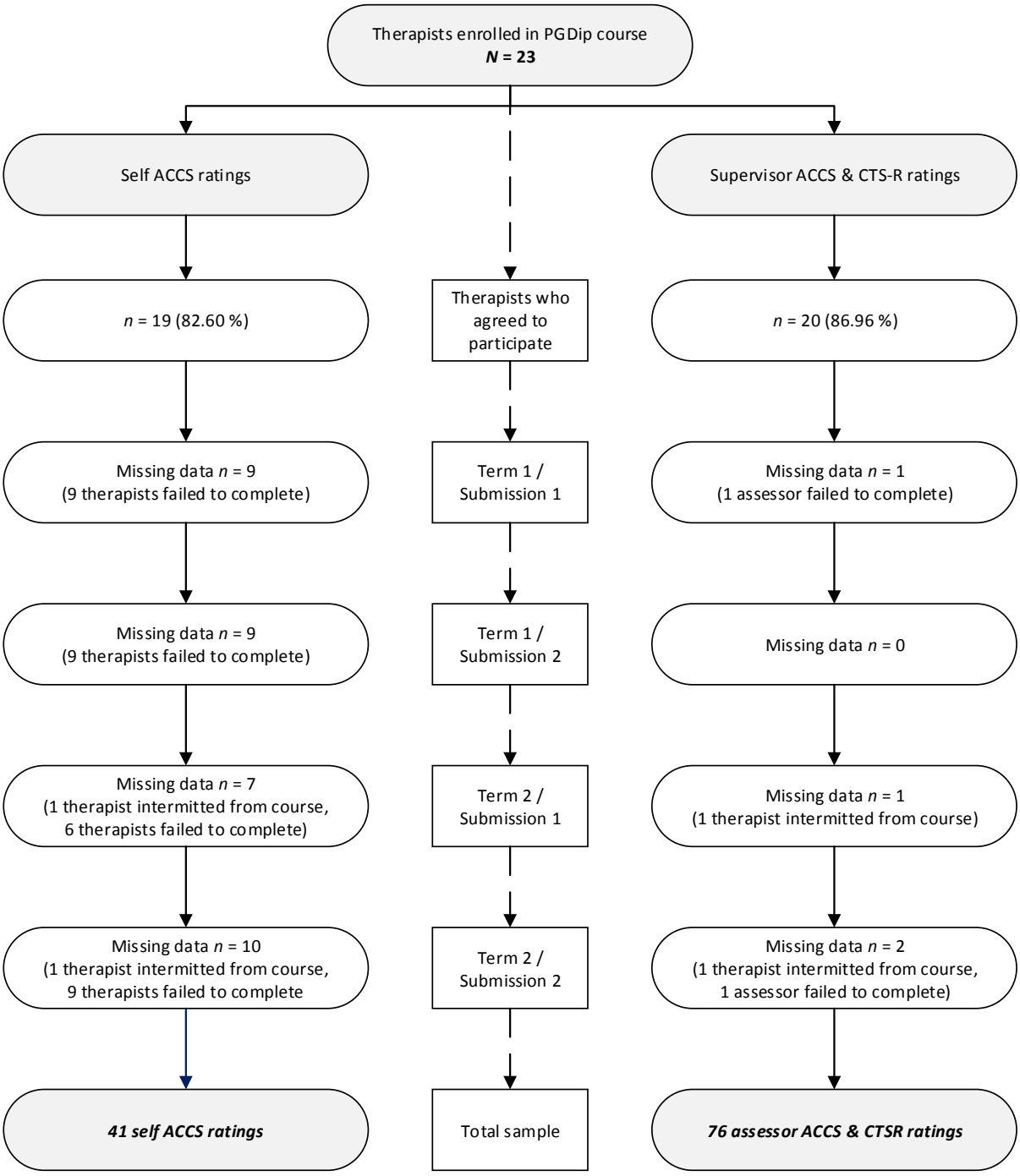


Figure 1: Flowchart outlining data collection at the Oxford Cognitive Therapy Centre Site

First Step

As part of their continuing professional development, therapists were routinely required to submit video recordings of CBT treatment sessions to their clinical supervisors for feedback on their performance. For the purposes of the current study, therapists selected one session recording for submission. This recording was viewed and rated using the ACCS scale by three people: the therapist's supervisor within First Step, a Senior Psychotherapist within First Step, and the therapist them self (i.e. self-rating) (see Figure two). Eleven supervisors completed ACCS ratings (one supervisor completed one rating, five supervisors completed two ratings, one supervisor completed three ratings, two supervisors completed four ratings, one supervisor completed five ratings, and one supervisor completed eight ratings) and four Senior Psychotherapists (SPs) completed ACCS ratings (one SP completed four ratings, one SP completed six ratings, one SP completed ten ratings, and one SP completed 15 ratings). Completed ACCS rating were sent directly to a member of the First Step administration team without being shared with any other parties. This meant that none of the other parties (i.e. therapists, supervisors, or senior psychotherapists) viewed ratings completed by the other members of staff. Assessors and therapists were provided with a copy of the ACCS manual (see Appendix six in chapter four). Assessors and therapists within First Step also attended a one-day training course in how to use the ACCS scale (see chapter seven).

Materials

Cognitive therapy scale-revised

Recordings submitted by therapists within OCTC were rated using the Cognitive Therapy Scale-Revised (CTS-R: Blackburn et al., 2001, see Appendix six). This is a 12-item scale that assesses general therapeutic skills (agenda, feedback, collaboration, pacing/use of time, and interpersonal effectiveness) and CBT specific skills (guided discovery, conceptualisation, identifying key cognitions, application of change methods, application of behavioural techniques, homework, and facilitation of emotional expression). All items are rated on a seven-point scale (0 – *incompetent/non-compliance* to 6- *expert: compliance + high skill*), with item-specific descriptions being given for each anchor. Total CTS-R scores range from 0 to 72, with higher scores representing a higher level of skill. The CTS-R demonstrates high internal consistency (α range= .75–.97 Blackburn et al., 2001; James, Blackburn, Milne, & Reichfelt, 2001; Reichelt, James, & Blackburn, 2003). Studies which investigate inter-rater reliability for the CTS-R report agreement ranging from fair to good for total scores (Blackburn et al., 2001; Gordon, 2006; Reichelt et al., 2003) and from poor to good for individual items (Reichelt et al., 2003). For further information about the CTS-R see the literature review in Chapter One.

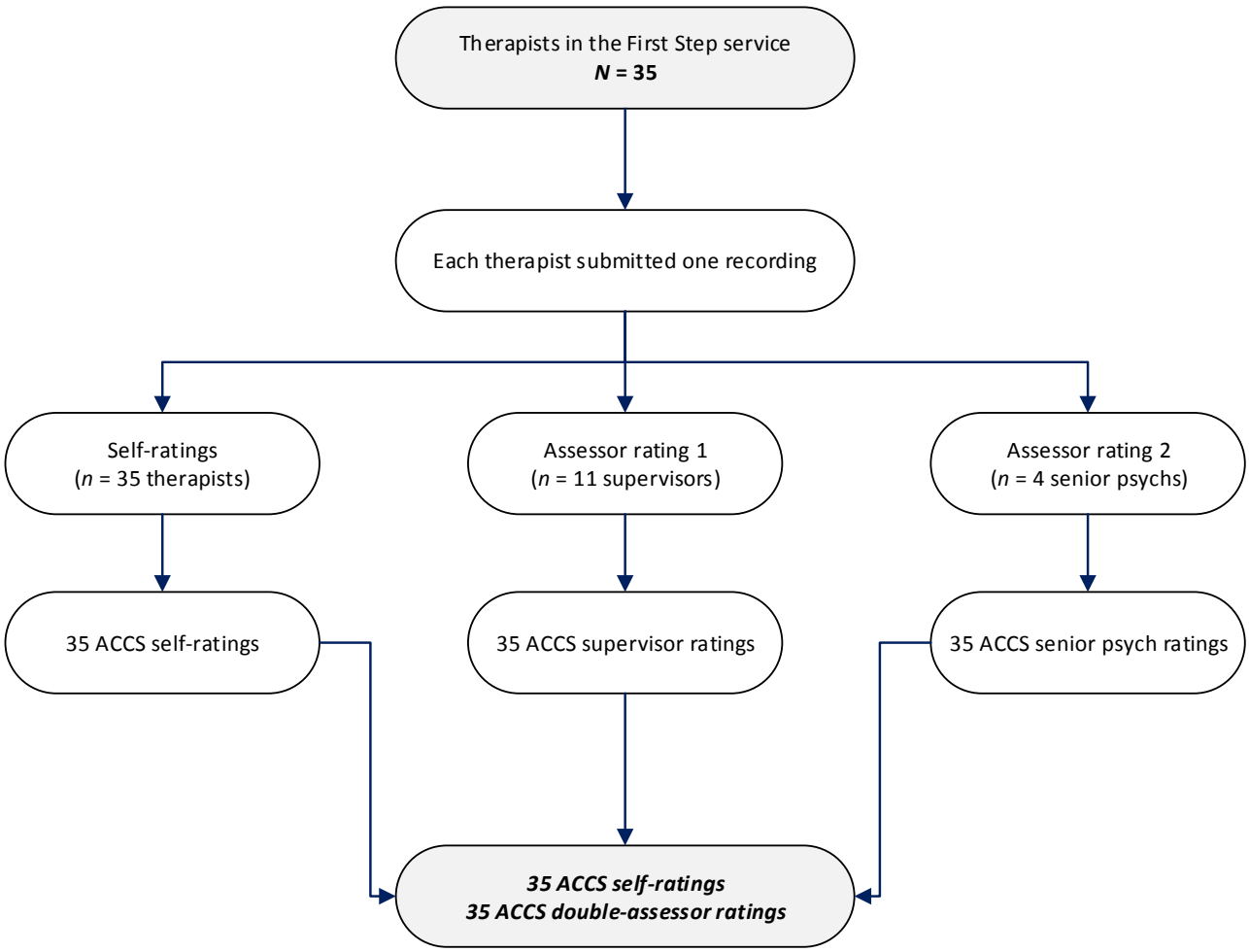


Figure 2: Flowchart outlining data collection at the First Step Site

ACCS

Recordings submitted by therapists within OCTC and First Step were rated using the Assessment of Core CBT Skills (ACCS: see chapter three). The ACCS consists of eight core competence domains (agenda setting, formulation, CBT interventions, homework, effective two-way communication, fostering therapeutic relationship, effective use of time, and assessing change), split into 22 individual items. All items are rated on a four-point scale (1- *limited*, 2- *basic*, 3- *good*, 4- *advanced*), with item-specific descriptions being given for each anchor. Total ACCS scores are calculated by summing the total domain scores (the average scores of items within the domain) and range from eight to 32, with higher scores representing a higher level of skill.

Assessors and therapists were provided with a copy of the ACCS manual (see Appendix six in chapter four). The manual provides guidance as to how to make judgements about the quality

or skilfulness of therapists' performance using the rating scale. As well as general guidelines, the manual provides clarification about the constructs measured by each scale item by detailing item-specific examples of behaviours consistent with each scale point. A numerical version of the ACCS feedback form (i.e. a feedback form with space for in-depth formative comments removed) was used to rate the therapists performance within the recorded treatment sessions (see Appendix seven).

Session recordings

The total number of session recordings available from OCTC and First Step are outlined in Figures one and two respectively. As different samples were used according to the analysis being conducted, more detail is provided within the data analysis section below. Therapists also completed a supporting information cover sheet (see Appendix eight). This was submitted alongside treatment session recordings in order to provide assessors with information about the therapeutic context of the recording (e.g. the session number, patient's presenting problem, treatment goals etc.).

Data analysis

Psychometric analysis was conducted for data collected within OCTC and First Step independently, as well as for the pooled data from both sites. Within First Step, all recordings were rated twice on the ACCS: once by a supervisor and once by a senior psychotherapist. Using an aggregate score of multiple assessors' ratings improves the reliability of the measurement for analyses (Vallis, Shaw, & Dobson, 1986). However, this approach also limits the generalisability of the results to studies and settings using only a single assessor. As most training and routine practice settings do not have the resources to use multiple assessors, it was therefore decided to use only the supervisor's ratings within the psychometric analysis (with the exception of inter-rater reliability).

Descriptive statistics

The mean, standard deviations and range of scores are provided for the ACCS total score, each of the eight ACCS domains, and the 22 individual ACCS items. Each item mean and variance was calculated, as was the percentage of individuals assigned the maximum and minimum values for each item. These descriptive statistics were examined to establish whether the items discriminated well (i.e. did any items have very low variance because the majority of assessors/therapists assigned the same response category?) whether any items demonstrate strong positive or negative skew (i.e. did the distribution of scores on the item concentrate on the

upper or lower range of the scale), and to check for any floor or ceiling effects (i.e. did any items fail to distinguish between therapists below or above the level of competence asked about in the response scale because extreme items are missing in the lower or upper end of the scale?). These descriptive statistics were completed independently for 76 self-ratings on the ACCS (41 from OCTC and 35 from First Step) and for 111 assessor-ratings on the ACCS (76 from OCTC and 35 from First Step).

Internal consistency

Three measures of internal consistency (the degree to which items assess the same underlying construct) were employed. First, Pearson's correlations between items within the same domain were examined to identify any items that were unrelated to or overlapped significantly with other items in the domain. Second, to examine how highly correlated each item was with the overall scale, corrected item-total correlations were calculated (i.e. the Pearson correlation between each item and the average scores of the remaining scale items). Third, the correlation amongst individual items in the measure was calculated using Cronbach's alpha (α), the average of all possible split-half reliabilities (the correlation between the total scores of two randomly divided halves of the sample). The α if item deleted was also examined for each item to investigate whether any items needed to be removed from the scale to increase internal consistency. Internal consistency was calculated independently for 76 self-ratings on the ACCS (41 from OCTC and 35 from First Step) and for the 111 assessor-ratings on the ACCS (76 from OCTC and 35 from First Step). To enable comparison, corrected item-total correlations and Cronbach's alpha was also calculated for 76 assessor-ratings and 41 self-ratings on the CTS-R (all from OCTC).

Inter-rater reliability

Because a pool of assessors was used within OCTC and First Step, the same two raters did not assess every tape. Thus, to enable comparison between first and second markers, assessors were allocated as either "rater 1" or "rater 2". Within OCTC supervisors were allocated as rater 1 and the second marker was allocated as rater 2. Within First Step, supervisors were allocated as rater 1 and senior psychotherapists were allocated as rater 2. Agreement between raters was examined for the total score and individual items by calculating intraclass correlation coefficients (ICC), treating the raters as random effects (Strout & Fleiss, 1979, Model 2, 1). ICC was used instead of Pearson's correlation because it is considered a more suitable measure of reliability, especially when several pairs of judges rate the same targets and where assignment to pairs is arbitrary (Lahey, Downey, & Saal, 1983; Shrout & Fleiss, 1979). ICC values was calculated for the

20 pairs of assessor-ratings on the ACCS completed in OCTC and for the 35 pairs of assessor-ratings on the ACCS completed in First Step.

Discriminant validity

The competence of trainee CBT therapists has been shown to increase as they improve their skills during the course of diploma-level training (McManus et al., 2010; Williams, Moorey, & Cobb, 1991). If the ACCS is a valid measure of CBT competence, it would be expected that trainees' competence would increase over time as they develop skills during training. A repeated measure ANOVA was therefore used to examine whether the ACCS total scores of the four session recordings increased significantly over the course of training and Bonferroni post-hoc pairwise comparisons were used to test for differences between all possible session recording pairs. To enable comparison, this analysis was also conducted for CTS-R ratings. Data was missing for three therapists (one therapist intermitted and two assessors failed to complete ratings). This analysis was therefore conducted for the 17 therapists for whom a full data set (i.e. four ACCS and CTS-R ratings) was available.

Progress of ACCS ratings across the course was not examined for self-ratings. This was in part because research has shown that self-ratings often decrease as therapists begin to gain competence (McManus, Rakovshik, Kennerley, Fennell, & Westbrook, 2011), possibly because they become more aware of their skills deficits as they move from the developmental stage of 'unconscious incompetence' to 'conscious incompetence' (Howell, 1982; Inskipp & Proctor, 1993). Additionally, a complete data set was only available for a very small number of therapists (n = 6).

Convergent validity

The correlation (Pearson's correlation coefficient) between ratings assessors made on the ACCS and those made on the CTS-R will be examined to explore the relationship between the scales. As both scales assess CBT competence, it would be expected that the total scores from the two scales would be positively correlated with each other. This analysis was conducted using the 76 ACCS and CTS-R ratings completed by assessors within OCTC.

Results

Descriptive statistics

Assessor-rated ACCS

Descriptive statistics for assessor-rated ACCS are presented in Table three. Examining the means of the items shows that most items fell close to the centre of the range of possible scores

(range: 1 – 4, mid-point 2.5). Specifically, the mean fell between two and three for the majority of items (within OCTC and First Step *n* = 20, 90.90 %). However, the mean fell just outside of this range for a small number of items (within OCTC two items [9.09 %] had a mean of 3.39 [7.1 interpersonal style] and 3.26 [7.2 empathic understanding]; within First Step two items [9.09 %] had a mean of 1.86 [5.2 implementing measures] and 3.16 [7.1 interpersonal style]).

For the majority of items assessors used either the full range of ACCS scores (i.e. one to four: within OCTC and First Step *n* = 15, 68.18 %) or very near to the full range of scores (i.e. 1 - 3.5 assigned within OCTC for *n* = 2, 9.09 %; 1.5 – 4 assigned within First Step for *n* = 5, 22.73 %). For a small number of items, assessors did not assign the upper limit of the scale (i.e. no scores above three assigned within OCTC *n* = 3, 13.64 % [items 5.1- choosing suitable measures, 5.2- implementing measures, & 6.2 time management]; within First Step *n* = 2, 9.09 % [items 5.1- choosing suitable measures & 5.2- implementing measures]) or the lower limit of the scale (i.e. no scores below two assigned within OCTC *n* = 2, 9.09 % [items 7.1- interpersonal style & 7.2- empathic understanding]).

Examining the percentage of participants assigned the maximum and minimum values for each item reveals a similar picture. For the majority of items, a small percentage (<20 %) of participants were assigned either a one or a four (First Step: *n* = 19 items, 86.36 %, OCTC: *n* = 18 items, 81.82 %). Within First Step and OCTC, between 28.1 % and 34.4 % of participants were assigned a one and none were assigned a four for items 5.1- choosing suitable measures and 5.2- implementing measures. For two items within OCTC (7.1- interpersonal style & 7.2- empathic understanding) no participants were assigned a one and over 20 % of participants were assigned a four. A similar pattern was seen for item 7.1 in First Step (2.9 % assigned a one and 22.9 % assigned a four) but not for item 7.2.

Taken together, the descriptive statistics indicate that most items in the assessor-rated ACCS: did not demonstrate a strong positive or negative skew, were able to capture different levels of competence, and were not limited by floor or ceiling effects. However, items in the measuring change domain (5.1 and 5.2) were clustered around the lower range of the scale, with many participants being assigned the lower limit of the scale and no participants being assigned the upper limit of the scale. Items 7.1- interpersonal style and 7.2- empathic understanding were also clustered around the upper range of the scale within OCTC, although this was only replicated for interpersonal style within First Step.

Table 3: Descriptive statistics for assessor-rated Assessment of Core CBT Skills (ACCS) in the Oxford Cognitive Therapy Centre (OCTC) and First Step.

	OCTC *n = 76*				**First Step** *n = 35*				**Pooled Data** *n = 111*			
Item / domain [1]	**Mean (*SD*)**	**[range]** [2]	**% min score (1)**	**% max score (4)**	**Mean (*SD*)**	**[range]**	**% min score (1)**	**% max score (4)**	**Mean (*SD*)**	**[range]**	**% min score (1)**	**% max score (4)**
Item 1.1	2.77 (0.56)	1 – 4	2.6	3.9	2.60 (0.64)	1 – 4	2.9	2.9	2.72 (0.59)	1 - 4	2.6	3.5
Item 1.2	2.84 (0.54)	1 – 4	2.6	5.3	2.66 (0.68)	1 – 4	5.7	5.7	2.78 (0.59)	1 – 4	3.5	5.3
Domain 1	2.80 (0.50)	1 – 4	2.6	2.6	2.63 (0.62)	1 – 4	2.9	2.9	2.75 (0.54)	1 – 4	2.6	2.6
Item 2.1	2.72 (0.77)	1 – 4	9.2	7.9	2.56 (0.69)	1 – 4	5.7	5.7	2.67 (0.75)	1 - 4	7.9	7.0
Domain 2	2.72 (0.77)	1 – 4	9.2	7.9	2.56 (0.69)	1 – 4	5.7	5.7	2.67 (0.75)	1 - 4	7.9	7.0
Item 3.1	2.70 (0.67)	1 – 4	3.9	5.3	2.81 (0.61)	1.5 – 4	2.9	8.6	2.73 (0.65)	1 - 4	2.6	6.1
Item 3.2	2.72 (0.53)	1 – 4	1.3	1.3	2.79 (.63)	1 – 4	2.9	8.6	2.74 (0.57)	1 - 4	1.8	3.5

Item 3.3	2.77 (0.61)	1 – 4	2.6	5.3	2.70 (.61)	1 – 4	2.9	5.7	2.75 (0.60)	1 - 4	2.6	5.3
Item 3.4	2.57 (0.54)	1 – 3.5	1.3	0.0	2.76 (.72)	1 – 4	2.9	14.3	2.63 (0.61)	1 - 4	1.8	4.4
Item 3.5	2.36 (0.70)	1 – 3.5	10.5	7.9	2.67 (.80)	1 – 4	8.6	11.4	2.46 (0.75)	1 - 4	9.6	3.5
Domain 3	2.62 (0.51)	1.20 – 3.70	0.0	0.0	2.75 (.58)	1.10 – 4	0.0	5.7	2.66 (0.53)	1.10 – 4	0.0	1.8
Item 4.1	2.42 (0.76)	1 – 4	15.8	1.3	2.74 (.76)	1 – 4	5.7	14.3	2.52 (0.77)	1 - 4	12.3	5.3
Item 4.2	2.67 (0.52)	1 – 4	1.3	1.3	2.70 (.71)	1 – 4	5.7	8.6	2.68 (0.58)	1 - 4	2.6	3.5
Item 4.3	2.65 (0.59)	1 – 4	3.9	2.6	2.50 (.77)	1 – 4	8.6	8.6	2.60 (0.65)	1 - 4	5.3	4.4
Item 4.4	2.35 (0.63)	1 – 4	7.9	1.3	2.37 (.81)	1 – 4	11.4	8.6	2.36 (0.69)	1 - 4	8.8	3.5
Domain 4	2.52 (0.49)	1.25 – 3.75	0.0	0.0	2.58 (.63)	1.25 – 4	0.0	2.9	2.54 (0.53)	1.25 - 4	1.8	0.9

Item 5.1	2.13 (0.85)	1 – 3	28.9	0.0	2.01 (0.78)	1 – 3	28.6	0.0	2.09 (0.82)	1 – 3	28.1	0.0
Item 5.2	2.03 (0.86)	1 – 3	34.2	0.0	1.86 (0.73)	1 – 3	34.4	0.0	1.97 (0.82)	1 - 3	33.3	0.0
Domain 5	2.08 (0.81)	1 – 3	26.3	0.0	1.94 (0.72)	1 – 3	25.7	0.0	2.03 (0.78)	1 – 3	25.4	0.0
Item 6.1	2.68 (0.58)	1 – 4	1.3	3.9	2.83 (0.62)	1.50 – 4	2.9	11.4	2.73 (0.59)	1 - 4	0.9	6.2
Item 6.2	2.55 (0.62)	1 – 3	6.6	0.0	2.60 (0.65)	1.50 – 4	2.9	8.6	2.56 (0.63)	1 - 4	4.4	2.6
Item 6.3	2.59 (0.62)	1 – 4	5.3	1.3	2.79 (0.69)	1.50 – 4	2.9	14.3	2.65 (0.64)	1 - 4	3.5	5.3
Domain 6	2.61 (0.50)	1.33 – 3.33	0.0	0.0	2.74 (0.59)	1.50 – 4	0.0	8.6	2.65 (0.53)	1.33 – 4.00	0.0	2.6
Item 7.1	3.39 (0.50)	2 – 4	0.0	36.8	3.16 (0.60)	1.50 – 4	2.9	22.9	3.32 (0.54)	1.50 – 4	0.9	31.6
Item 7.2	3.26 (0.55)	2 – 4	0.0	28.9	2.97 (0.69)	1 – 4	2.9	17.1	3.17 (0.61)	1 - 4	0.9	24.6

Item 7.3	2.95 (0.66)	1 – 4	1.3	14.5	2.89 (0.75)	1 – 4	2.9	17.1	2.93 (0.69)	1 - 4	1.8	15.3
Domain 7	3.20 (0.47)	2.33 – 4	0.0	9.2	3.00 (0.61)	1.17 – 4	0.0	8.6	3.14 (0.52)	1.17 - 4	0.0	8.8
Item 8.1	2.78 (0.67)	1 – 4	3.9	9.2	2.63 (0.68)	1 – 4	2.9	8.6	2.73 (0.67)	1 - 4	3.5	8.8
Item 8.2	2.70 (0.76)	1 – 4	5.3	11.8	2.56 (0.83)	1 – 4	8.6	11.4	2.66 (0.78)	1 - 4	6.1	11.4
Domain 8	2.74 (0.63)	1 – 4	2.6	3.9	2.59 (0.67)	1 – 4	2.9	5.7	2.70 (0.65)	1 - 4	2.6	4.4
Total Score	21.12 (3.06)	13.72 – 27.75	-	-	20.79 (3.78)	9.77 – 26.87	-	-	21.02 (3.29)	9.77 – 27.75	-	-

[1] Total domain scores are the average scores of items within the domain. It is worth noting that domain two only includes one item- thus the total domain score for domain two is the same as the item score.

[2] Individual item / domain range = 1 – 4. Total ACCS score range = 8 – 32.

OCTC = Oxford Cognitive Therapy Centre

Self-rated ACCS

Descriptive statistics for self-rated ACCS scores are presented in Table four. Examining the means of the items shows that most items fell closely within the centre of the range of possible scores (range: 1 – 4, mid-point 2.5). Specifically, the mean fell between two and three for the majority of items (within OCTC and First Step *n* = 20, 90.91 %). However, the mean fell just outside of this range for a small number of items (within OCTC two items: [9.09 %] had a mean of 1.94 [5.1 choosing suitable measures] and 1.90 [2.2 implementing measures]; within First Step two items [9.09% %] had a mean of 3.07 [7.1 interpersonal style] and 3.09 [7.2 empathic understanding]).

Within OCTC the range for only a small number of items fell at or near the full range of ACCS scores (i.e. 1 - 3.5 assigned for *n* = 4, 18.18 %). For the majority of items the upper limit of the scale was not used (i.e. no scores above three assigned for *n* = 17, 77.27 %), with the lower limit of the scale not being used (i.e. no scores below two) for only one item (4.55 %). Within First Step, the full range of ACCS scores (i.e. 1 – 4) were used for eight items (36.36 %). For the majority of items the lower limit of the scale was not used (i.e. no scores below two for *n* = 11, 50.00 %), with the upper limit of the scale not being used (i.e. no scores above three) for only one item (4.55 %). Additionally, two items (9.09 %), did not use the upper or lower limit (i.e. 2 – 3 assigned).

Examining the percentage of participants assigned the maximum and minimum values for each item shows that for the majority of items, a small percentage (<20 %) of participants were assigned either a one or a four (First Step: *n* = 21 items, 95.45 %, OCTC: *n* = 20 items, 90.90 %). Within OCTC, between 31.7 % and 36.6 % of participants were assigned a one and none were assigned a four for two items (5.1- choosing suitable measures and 5.2- implementing measures). For one item within First Step (7.2- empathic understanding) 2.9 % of participants were assigned a one and just over 20 % of participants were assigned a four.

Taken together, the descriptive statistics indicate that most items in the self-rated ACCS: did not demonstrate a strong positive or negative skew, were able to capture different levels of competence, and were not limited by floor or ceiling effects. However, all of the items on the self-rated ACCS clustered around the centre point of the scale, with therapists in OCTC tending not to assign the upper limit of the scale and therapists in First Step tending not to assign the lower limit of the scale. Within OCTC, items in the measuring change domain (5.1 and 5.2) were clustered around the lower range of the scale, with many participants being assigned the lower limit of the scale and no participants being assigned the upper limit of the scale. However, this finding was not replicated in First Step. Item 7.2- empathic understanding was also clustered around the upper range of the scale within First Step, although this was not replicated within OCTC.

Table 4: Descriptive statistics for self-rated Assessment of Core CBT Skills (ACCS) in the Oxford Cognitive Therapy Centre (OCTC) and First Step.

	OCTC				**First Step**				**Pooled Data**			
	n = 41				***n = 35***				***n = 76***			
Item / domain [1]	**Mean (*SD*)**	**[range]** [2]	**% min score (1)**	**% max score (4)**	**Mean (*SD*)**	**[range]**	**% min score (1)**	**% max score (4)**	**Mean (*SD*)**	**[range]**	**% min score (1)**	**% max score (4)**
Item 1.1	2.48 (0.53)	1 – 3	2.4	0.0	2.60 (0.60)	1 – 4	2.9	2.9	2.55 (0.56)	1 – 4	2.6	1.3
Item 1.2	2.55 (0.54)	1 – 3	2.4	0.0	2.63 (0.59)	1 – 4	2.9	2.9	2.60 (0.56)	1 – 4	2.6	1.3
Domain 1	2.51 (0.42)	1.50 – 3.00	0.0	0.0	2.62 (0.57)	1 – 4	2.9	2.9	2.57 (0.49)	1 – 4	1.3	1.3
Item 2.1	2.38 (0.58)	1 – 3	4.9	0.0	2.74 (0.62)	1 – 4	2.9	5.7	2.55 (0.62)	1 – 4	3.9	2.6
Domain 2	2.38 (0.58)	1 – 3	4.9	0.0	2.74 (0.62)	1 – 4	2.9	5.7	2.55 (0.62)	1 – 4	3.9	2.6
Item 3.1	2.47 (0.54)	1 – 3	2.4	0.0	2.84 (0.50)	2 – 4	0.0	5.7	2.65 (0.55)	1 – 4	1.3	2.6
Item 3.2	2.37 (0.54)	1 – 3	2.4	0.0	2.66 (0.47)	2 – 3	0.0	0.0	2.51 (0.53)	1 – 3	1.3	0.0

Item 3.3	2.44 (0.63)	1 – 3	7.3	0.0	2.65 (0.49)	2 – 3	0.0	0.0	2.53 (0.58)	1 – 3	3.9	0.0
Item 3.4	2.23 (0.46)	1 – 3	2.4	0.0	2.49 (0.56)	1 – 3	2.9	0.0	2.35 (0.52)	1 – 3	2.6	0.0
Item 3.5	2.23 (0.55)	1 – 3	7.3	0.0	2.56 (0.65)	1 – 4	2.9	5.7	2.38 (0.62)	1 – 4	5.3	2.6
Domain 3	2.33 (0.45)	1 – 3	2.4	0.0	2.75 (0.58)	1.10 – 4	0.0	5.7	2.54 (0.54)	1 – 4	1.3	2.6
Item 4.1	2.27 (0.76)	1 – 4	12.2	2.4	2.59 (0.78)	1 – 4	11.4	5.7	2.43 (0.77)	1 – 4	11.8	3.9
Item 4.2	2.54 (0.55)	1 – 3	2.4	0.0	2.77 (0.51)	2 – 4	0.0	2.9	2.65 (0.54)	1 – 4	1.3	1.3
Item 4.3	2.51 (0.67)	1 – 4	4.9	2.4	2.71 (0.52)	2 – 4	0.0	2.9	2.63 (0.58)	1 – 4	2.6	2.6
Item 4.4	2.13 (0.54)	1 – 3	9.8	0.0	2.44 (0.59)	2 – 4	0.0	5.7	2.28 (0.59)	1 – 4	5.3	2.6
Domain 4	2.36 (0.45)	1.50 – 3.38	0.0	0.0	2.58 (0.63)	1.25 – 4	0.0	2.9	2.47 (0.54)	1.25 – 4	0.0	1.3
Item 5.1	1.94 (0.85)	1 – 3	36.6	0.0	2.40 (0.74)	1 – 4	11.4	2.9	2.17 (0.82)	1 – 4	25.0	1.3

Item 5.2	1.90 (0.75)	1 – 3	31.7	0.0	2.27 (0.78)	1 – 4	14.3	5.7	2.09 (0.78)	1 – 4	23.7	2.6
Domain 5	1.91 (0.79)	1 – 3	35.5	0.0	2.34 (0.72)	1 – 4	11.4	2.9	2.12 (0.78)	1 – 4	23.7	1.3
Item 6.1	2.38 (0.57)	1 – 3	2.4	0.0	2.71 (0.49)	2 – 4	0.0	2.9	2.55 (0.53)	1 – 4	1.3	1.3
Item 6.2	2.39 (0.61)	1 – 3	4.9	0.0	2.79 (0.47)	2 – 4	0.0	2.9	2.59 (0.56)	1 – 4	2.6	1.3
Item 6.3	2.39 (0.65)	1 – 3	7.3	0.0	2.80 (0.53)	2 – 4	0.0	5.7	2.60 (0.61)	1 – 4	3.9	2.6
Domain 6	2.39 (0.48)	1 – 3	0.0	0.0	2.77 (0.39)	2 – 3.67	0.0	0.0	2.58 (0.45)	1.33 – 3.67	0.0	0.0
Item 7.1	2.88 (0.51)	2.00 – 4.00	0.0	7.3	3.07 (0.56)	2 – 4	0.0	17.1	2.98 (0.53)	2 – 4	0.0	11.8
Item 7.2	2.69 (0.59	1 – 3.5	2.4	0.0	3.09 (0.66)	1 – 4	2.9	22.9	2.89 (0.62)	1 – 4	2.6	10.5
Item 7.3	2.52 (0.65)	1 – 3.5	4.9	0.0	2.99 (0.59)	2 – 4	0.0	14.3	2.75 (0.64)	1 – 4	2.6	6.6
Domain 7	2.70 (0.49)	1.33 – 3.67	0.0	0.0	3.05 (0.48)	1.67 – 4	0.0	5.7	2.87 (0.49)	1.33 – 4	0.0	2.6

Item 8.1	2.41 (0.66)	1 – 3	7.3	0.0	2.66 (0.58)	2 – 4	0.0	5.7	2.55 (0.61)	1 – 4	3.9	2.6
Item 8.2	2.23 (0.70)	1 – 3	14.6	0.0	2.46 (0.55)	2 – 4	0.0	2.9	2.34 (0.63)	1 – 4	7.9	1.3
Domain 8	2.32 (0.60)	1 – 3	4.9	0.0	2.56 (0.50)	2 – 4	0.0	2.9	2.44 (0.55)	1 – 4	2.6	1.3
Total Score	18.85 (2.80)	11.23 – 24.01	-	-	19.61 (3.98)	8.33 – 28.47	-	-	19.34 (3.27)	8.33 – 28.47	-	-

[1] Total domain scores are the average scores of items within the domain. It is worth noting that domain two only includes one item- thus the total domain score for domain two is the same as the item score.

[2] Individual item / domain range = 1 – 4. Total ACCS score range = 8 – 32.

OCTC = Oxford Cognitive Therapy Centre

Internal consistency

Correlation between domain items

As the ACCS breaks down a number of aspects of CBT competence into smaller constituent parts, it is possible that there is too much overlap between items within the domain and thus that some items could be removed or amalgamated. Similarly, as ACCS items are clustered within domains according to the theoretical relationship between the items, it is possible that the items within the domain are not related in practice. As can be seen from the correlations between domain items in Table five, neither of these issues seemed to be problematic for the ACCS scale. For the assessor-rated ACCS, all items demonstrated a moderate or strong correlation with other items in the same domain. For the self-rated ACCS, 24 of the 25 item pairs (96.00 %) from the OCTC site and 16 of the 25 item pairs (64.00 %) from the First Step demonstrated a moderate or strong correlation with other items in the same domain. The only item pairing that had weak correlations across both sites for the self-rated version were 4.1- reviewing homework and 4.4- planning homework. However, these items were strongly correlated across both sites for the assessor-rated version. One item pairing (5.1- choosing suitable measures and 5.2-implementing measures) was very strongly correlated across both sites for both the self- and assessor-rated versions.

Table 5: Correlations between domain items on the Assessment of Core CBT Skills (ACCS)

Domain	Items	Pearson's Correlation			
		Assessor-rated ACCS		Self-rated ACCS	
		OCTC *N* = 76	First Step *N* = 35	OCTC *N* = 41	First Step *N* = 35
One	1.1 vs 1.2	.64 **	.64 **	.52 **	.85 **
Two	n/a (one item)				
Three	3.1 vs 3.2	.69 **	.70 **	.62 **	.31
	3.1 vs 3.3	.72**	.72 **	.38 **	.20
	3.1 vs 3.4	.55**	.55 **	.56 **	.29
	3.1 vs 3.5	.58 **	.58 **	.63 **	.46 **
	3.2 vs 3.3	.48 **	.48 **	.51 **	.26
	3.2 vs 3.4	.58 **	.58 **	.58 **	.64 **

	3.2 vs 3.5	.48 **	.48 **	.56 **	.24
	3.3 vs 3.4	.62 **	.62 **	.37 **	.32
	3.3 vs 3.5	.62 **	.62 **	.52 **	.34
	3.4 vs 3.5	.66 **	.66 **	.60 **	.44 **
Four	4.1 vs 4.2	.30 **	.30 **	.40	.22
	4.1 vs 4.3	.47 **	.47 **	.34	.21
	4.1 vs 4.4	.51 **	.51 **	.23 *	.25
	4.2 vs 4.3	.60 **	.60 **	.59 **	.53 **
	4.1 vs 4.4	.49 **	.49 **	.63 **	.45 **
	4.3 vs 4.4	.50 **	.50 **	.65 **	.52 **
Five	5.1 vs 5.2	.80 **	.80 **	.86 **	.81 **
Six	6.1 vs 6.2	.48 **	.48 **	.45 **	.65 **
	6.1 vs 6.3	.48 **	.48 **	.38 **	.34 *
	6.2 vs 6.3	.67 **	.67 **	.41 **	.29
Seven	7.1 vs 7.2	.62 **	.62 **	.48 **	.66 **
	7.1 vs 7.3	.57 **	.57 **	.51 **	.25
	7.2 vs 7.3	.41 **	.41 **	.42 **	.42 *
Eight	8.1 vs 8.2	.59 **	.59 **	.54 **	.58 **

*p < .05; **p < .01; ***p < .001.

Item-total correlation

Corrected item-total correlations for the self- and assessor-rated versions of the ACCS are outlined in Table six. The range of item-total correlations considered ‘acceptable’ is 0.30 to 0.80 (Loewenthal, 2001; Streiner & Norman, 2003). Items below this range [<.3] indicate that the item is not measuring the same construct as other items in the scale and items above this range [>.8] indicate item overlap and thus may be redundant. Nearly all of the items (86.36 %) fell within the acceptable range for both the assessor and self-rated versions of the ACCS across First Step and

OCTC sites. Two items fell just below this range (4.1- reviewing homework and 3.3- rationale for interventions) and one item fell just above this range (7.3- collaboration). However, these item only narrowly missed the threshold, these results were not consistent across sites, and all of the item-total correlations for the pooled data fell within the acceptable range (assessor ratings = .33 - .74, self-ratings = .37 - .61). Thus, no items were removed from the scale. Corrected item-total correlations for the assessor-rated CTS-R ranged from .41 to.72 (*n* = 76 from OCTC), and for the self-rated CTS-R ranged from .55 to .82 (*n* = 41 from OCTC), and thus also fell within the acceptable range.

Table 6: Corrected item-total correlations for the self- and assessor-rated versions of the Assessment of Core CBT Skills (ACCS) in the Oxford Cognitive Therapy Centre (OCTC) and First Step.

	Assessor ACCS Ratings			**Self ACCS Ratings**		
Item	**OCTC** ***n = 76***	**First Step** ***n = 35***	**Pooled Data** ***n = 111***	**OCTC** ***n = 41***	**First Step** ***n = 35***	**Pooled Data** ***n = 76***
Item 1.1	0.50	0.49	0.49	0.37	0.35	0.37
Item 1.2	0.47	0.33	0.41	0.36	0.42	0.38
Item 2.1	0.50	0.78	0.59	0.31	0.61	0.50
Item 3.1	0.73	0.74	0.72	0.53	0.51	0.58
Item 3.2	0.56	0.78	0.65	0.60	0.31	0.52
Item 3.3	0.70	0.80	0.73	0.63	0.16	0.46
Item 3.4	0.68	0.71	0.68	0.51	0.53	0.56
Item 3.5	0.72	0.81	0.74	0.74	0.70	0.74
Item 4.1	0.46	0.74	0.54	0.28	0.54	0.43
Item 4.2	0.42	0.50	0.46	0.38	0.37	0.43
Item 4.3	0.55	0.60	0.57	0.43	0.55	0.50
Item 4.4	0.66	0.56	0.62	0.49	0.56	0.57
Item 5.1	0.35	0.33	0.33	0.48	0.36	0.48
Item 5.2	0.45	0.40	0.42	0.50	0.38	0.48

Item 6.1	0.50	0.65	0.55	0.49	0.54	0.55
Item 6.2	0.61	0.69	0.64	0.43	0.44	0.50
Item 6.3	0.65	0.77	0.69	0.54	0.59	0.61
Item 7.1	0.50	0.71	0.58	0.42	0.47	0.45
Item 7.2	0.32	0.65	0.46	0.44	0.50	0.52
Item 7.3	0.41	0.83	0.58	0.52	0.40	0.53
Item 8.1	0.39	0.60	0.47	0.40	0.46	0.45
Item 8.2	0.49	0.72	0.58	0.33	0.52	0.42

OCTC = Oxford Cognitive Therapy Centre

Cronbach's alpha

Cronbach's alpha (α) ranges from 0 = *items independent* to 1 = *items identical*. There is a balance to be struck between low correlation (indicated that the items do not assess the same underlying construct) and very high correlation (indicating item overlap/redundancy). Hence an α of .80 to .85 would generally represent satisfactory agreement (Bland & Altman, 1997). However, for scales which provide individual summative assessment, such as the current scale, an α as high as >.90 may be necessary (DeVellis, 2012). It is also advisable to strive for higher alphas during scale development as this allows for alphas to deteriorate somewhat when used in a new context (DeVellis, 2012). Cronbach's alpha for the assessor-rated version of the ACCS was .90 in OCTC and .94 in First Step which is comparable to the Cronbach's alpha for the assessor-rated CTS-R in this sample (α = .90, OCTC data only, *n* = 76). For the self-rated version of the ACCS Cronbach's alpha was .88 in OCTC and .88 in First Step, which is comparable to the Cronbach's alpha for the self-rated CTS-R in this sample (α = .90, OCTC data only, *n* = 41). Taken together these results indicate that there was more than satisfactory agreement between scale items for the self- and assessor-rated version of the ACCS.

Table seven outlines the Cronbach's alpha (α) if the item was deleted for each item in the self-rated and assessor-rated versions of the ACCS. The α if item deleted for the assessor-rated version of the ACCS ranged from .89 to .90 in OCTC and from .94 to .95 in First Step. For the self-rated version of the ACCS the α if item deleted ranged from 0.86 to 0.88 in OCTC and from .86 to .88 in First Step. These results indicate that none of the ACCS items would significantly increase the scale α if they were deleted and thus none were removed from the scale to increase internal consistency.

Table 7: Cronbach's alpha (α) if item deleted for the self- and assessor-rated versions of the Assessment of Core CBT Skills (ACCS) in Oxford Cognitive Therapy Centre (OCTC) and First Step.

	ACCS Assessor Ratings			**ACCS Self -ratings**		
Item	**OCTC** ***n = 76*** α = .90	**First Step** ***n = 35*** α = .94	**Pooled Data** ***n = 116*** α = .91	**OCTC** ***n = 41*** α = .88	**First Step** ***n = 35*** α = .88	**Pooled Data** ***n = 76*** α = .89
Item 1.1	0.90	0.94	0.92	0.87	0.87	0.89
Item 1.2	0.90	0.95	0.92	0.87	0.87	0.89
Item 2.1	0.90	0.94	0.92	0.87	0.87	0.89
Item 3.1	0.89	0.94	0.91	0.87	0.87	0.89
Item 3.2	0.90	0.94	0.91	0.86	0.88	0.89
Item 3.3	0.89	0.94	0.91	0.86	0.88	0.89
Item 3.4	0.89	0.94	0.91	0.87	0.87	0.89
Item 3.5	0.89	0.94	0.91	0.86	0.86	0.88
Item 4.1	0.90	0.94	0.92	0.88	0.87	0.89
Item 4.2	0.90	0.94	0.92	0.87	0.87	0.89
Item 4.3	0.90	0.94	0.92	0.87	0.87	0.89
Item 4.4	0.89	0.94	0.91	0.87	0.87	0.89
Item 5.1	0.90	0.95	0.92	0.87	0.88	0.89
Item 5.2	0.90	0.94	0.92	0.87	0.88	0.89
Item 6.1	0.90	0.94	0.92	0.87	0.87	0.89
Item 6.2	0.90	0.94	0.91	0.87	0.87	0.89
Item 6.3	0.90	0.94	0.91	0.87	0.87	0.89
Item 7.1	0.90	0.94	0.92	0.87	0.87	0.89
Item 7.2	0.90	0.94	0.92	0.87	0.87	0.89
Item 7.3	0.90	0.94	0.92	0.87	0.87	0.89
Item 8.1	0.90	0.94	0.92	0.87	0.87	0.89
Item 8.2	0.90	0.94	0.92	0.87	0.87	0.89

OCTC = Oxford Cognitive Therapy Centre

Inter-rater reliability

Intra-class correlations between raters are outlined in Table eight below. The following benchmarking scale was used to interpret agreement coefficients: < 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). Agreement between raters for the ACCS total score was good in OCTC (ICC = .74) and in First Step (ICC = .73). ICCs for individual items ranged from .79, $p < .001$ to .27, NS in OCTC and from .83, $p < .001$ to .28, NS in First Step. Within OCTC none of the individual items fell in the range of poor agreement, five (22.73 %) fell in the range of fair agreement, seven (31.82 %) fell in the range of moderate agreement, ten (45.45 %) fell in the range of good agreement, and none fell in the range of very good agreement. Within First Step none of the individual items fell in the range of poor agreement, four (18.18 %) fell in the range of fair agreement, four (18.18 %) fell in the range of moderate agreement, 12 (54.55 %) fell in the range of good agreement, and two (9.09 %) fell in the range of very good agreement. None of the items demonstrated fair agreement within First Step and OCTC.

Table 8: Intra-class correlations between raters for the Assessment of Core CBT Skills (ACCS) in the Oxford Cognitive Therapy Centre (OCTC) and First Step.

Item [1]	OCTC *n = 20*	First Step *n = 35*	Pooled Data *n = 55*
Item 1.1	.69 *	.66 ***	.62 ***
Item 1.2	.37	.75 ***	.70 ***
Domain 1	.31	.74 ***	.68 ***
Item 2.1	.27	.82 ***	.72 ***
Domain 2	.27	.82 ***	.72 ***
Item 3.1	.54	.67 ***	.61 ***
Item 3.2	.71**	.68 ***	.61 ***
Item 3.3	.78 **	.71 ***	.70 ***
Item 3.4	.79 **	.68 ***	.70 ***
Item 3.5	.71 **	.66 ***	.67 ***
Domain 3	.82 ***	.76 ***	.77 ***

Item 4.1	.71 **	.47 *	.45 *
Item 4.2	.46 *	.28	.36 *
Item 4.3	.67*	.37	.39 *
Item 4.4	.61*	.32	.40 *
Domain 4	.68 *	.41	.48 **
Item 5.1	.46	.78***	.69 ***
Item 5.2	.54 *	.75***	.68 ***
Domain 5	.48	.83 ***	.73 ***
Item 6.1	.52	.37	0.35
Item 6.2	.62 *	.42	.50 **
Item 6.3	.64 *	.50 *	.55 **
Domain 6	.64 *	.32	.45 *
Item 7.1	.67 **	.69 ***	.69 ***
Item 7.2	.40	.83 ***	.74 ***
Item 7.3	.40	.71 ***	.64 ***
Domain 7	.62 *	.79 ***	.75 ***
Item 8.1	.38	.57 **	.51 **
Item 8.2	.54*	.61 ***	.59 **
Domain 8	.52	.66 **	.62 ***
Total Score	.74 **	.73 ***	.70 ***

*p < .05; **p < .01; ***p < .001.

Colour coding for agreement coefficients: < 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.*

[1] Total domain scores are the average scores of items within the domain. It is worth noting that domain two only includes one item- thus the total domain score for domain two is the same as the item score.

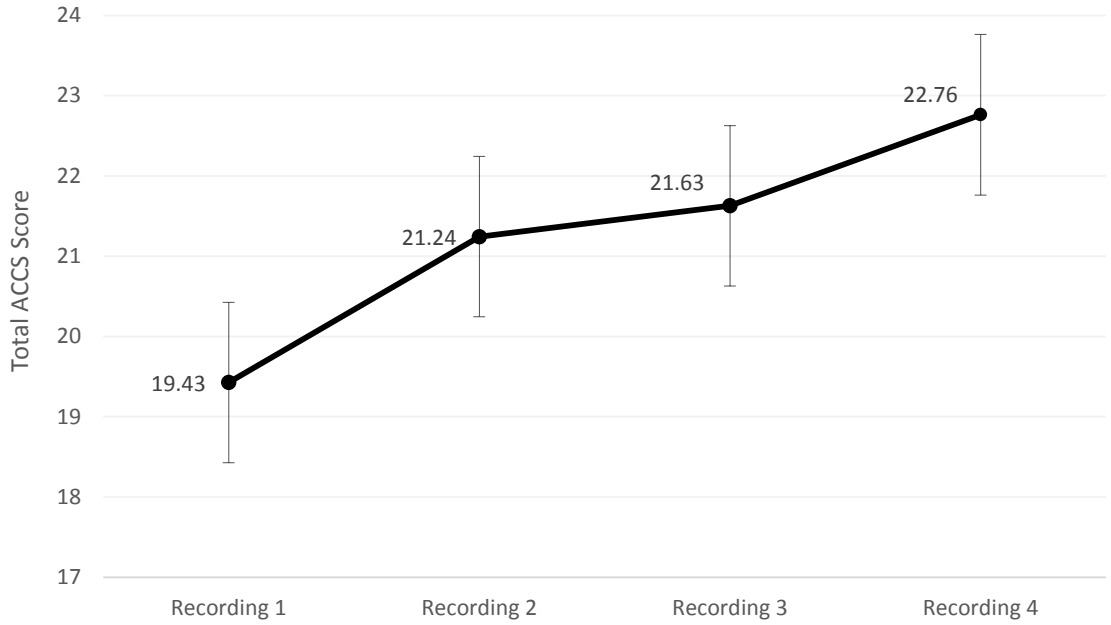


Figure 3: Means and S.E.s for total Assessment of Core CBT Skills (ACCS) scores for recordings submitted in term one and two of the Oxford Cognitive Therapy Centre Post Graduate Diploma in CBT (ACCS total score range = 8 - 32), *N* = 17.

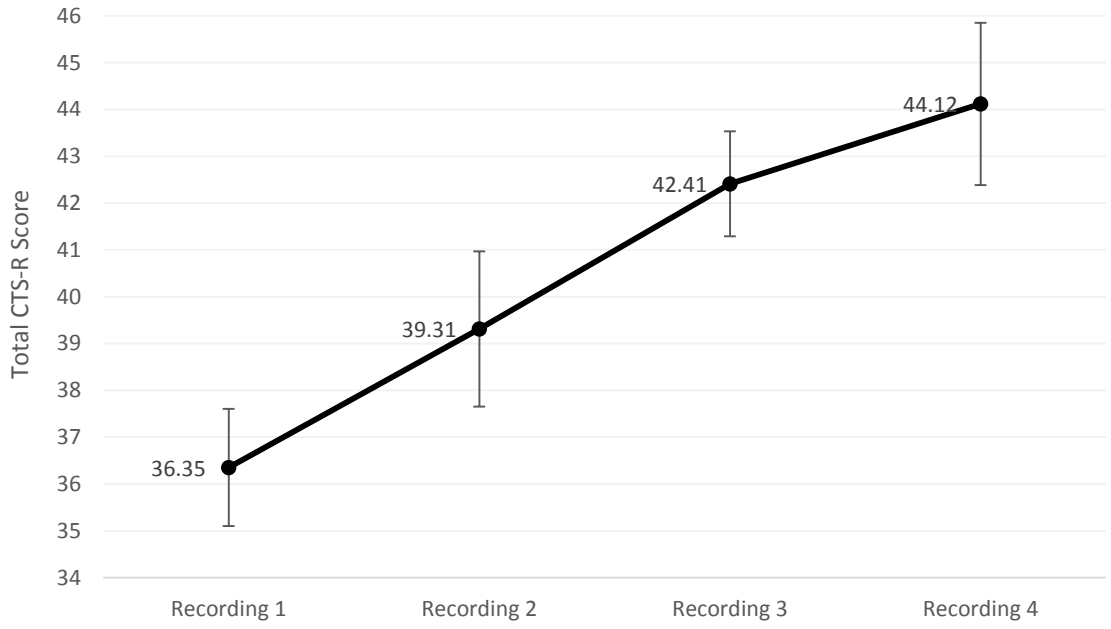


Figure 4: Means and S.E.s for total Revised Cognitive Therapy Scale (CTS-R) scores for recordings submitted in term one and two of the Oxford Cognitive Therapy Centre Post Graduate Diploma in CBT (CTS-R total score range = 0 – 72), *N* = 17.

Discriminant validity

The mean total scores for the four recordings submitted in term one and two of OCTC'S PGDip in CBT (*N* = 17) are presented for the ACCS in Figure three and for the CTS-R in Figure four. A repeated measures ANOVA indicated a significant increase in ACCS total scores over time (F[3,48] = 4.05, *p <.05*). Thus, as therapists progressed through the course, their level of competence on the ACCS improved. Bonferroni post-hoc pairwise comparisons showed that there was a significant increase in ACCS scores assigned from recording one to recording four (*p* >.05) and no significant difference between the ACCS scores assigned for any other recordings. There was also a significant increase in CTS-R total scores over time (F[3,48] = 6.35, *p <.01*). Bonferroni post-hoc pairwise comparisons showed that there was a significant increase in ACCS scores assigned from recording one to recording three (*p* >.05) and from recording one to recording four (*p* >.01).

Convergent validity

There was a strong positive relationship (*r* = .65, *p* = >.001) between total scores on the ACCS and the CTS-R assigned by assessors within OCTC (*n* = 76) and between total self-rated scores on the ACCS and CTS-R (*r* = .59, *p* = >.001, *n* = 40 within OCTC). These results indicate that, whilst the scales include distinct content, they do measure the same underlying construct (CBT competence).

Discussion

Summary of findings

The results of the pilot study indicate that the assessor-rated and self-rated versions of the Assessment of Core CBT Skills (ACCS) rating scale demonstrate good reliability and validity. Descriptive statistics revealed that the majority of ACCS items did not demonstrate a strong positive or negative skew, were able to capture different levels of competence, and were not limited by floor or ceiling effects, both in the assessor- and self-rated samples. However, items in the measuring change domain (5.1- choosing suitable measures and 5.2- implementing measures) clustered around the lower range of the scale, possibly reflecting an inability of the ACCS to adequately capture lower levels of competence in measuring change. However, this could also be explained by a lack of variability among the sample (i.e. a general lack of competence in this domain). Furthermore, examining the submission cover sheets revelaed that therapists often failed to provide information relating to idiosyncratic or standarised measures that were used within or across sessions and assessors in the focus group (see chapter six) reported uncertanty about whether they should use this documentation. Hence it is also possible that those completing the ACCS did not have access to or failed to take into consideration this supporting information. In

order to address this issue, further clarfication was added to the ACCS about the use of supporting documentation (see chapter six for further details).

In contrast, items 7.1- interpersonal style and 7.2 empathic understanding clustered around the upper range of the scale, in the assessor-rated sample. This could indicate that the ACCS does not adequately capture upper levels of competence in these areas. Alternatively, these findings could be due to a lack of variability among the sample. This is plausible given that the therapist participants were predominantly NHS professionals (e.g. nurses, clinical psychologists etc.) who had a number of years of clinical training and experience outside of a CBT framework. It is, therefore, likely that these individuals would have developed a high degree of skill in generic aspects of therapeutic competence, such as empathy and interpersonal skills. Descriptive statistics were broadly comparable across sites. The exception to this was the distribution of the self-rated ACCS scores: therapists in OCTC tended not to assign the upper limits of the scale, whilst therapists in First Step tended not to assign the lower limits of the scale. This may reflect a difference in therapist confidence due to the differing contexts as therapists in OCTC were novices recently embarked on training in CBT, whilst therapists in First Step were trained and experienced CBT therapists who had been working in routine practice for a number of years.

Measures of internal consistency indicate that items in the self-rated and assessor-rated ACCS are highly intercorrelated and thus that the ACCS measures a single phenomenon (pooled item-total correlations: assessor-ratings = .33 to .74, self-ratings = .37 to .61 and pooled α: assessor-ratings = .91 assessor, self-ratings = .89). However, these alpha values did not indicate excessive item overlap or redundancy (DeVellis, 2012). This is important given that the ACCS breaks down aspects of CBT competence into a number of smaller components, thus resulting in a large number of items. Internal consistency ratings for the assessor-rated ACCS are comparable to those reported elsewhere for assessor-ratings on the Revised Cognitive Therapy Scale (CTS-R: α range = .75 - .97; Blackburn et al., 2001; James et al., 2001; Reichelt et al., 2003), as well as within the current sample (α = .90; corrected item-total correlations = .41 to.72).

The majority of items demonstrated a moderate or strong correlation with other items in the same domain for the self-rated and assessor-rated ACCS, indicating that the items are related but still distinct. Thus, whilst the items are clustered within domains according to the theoretical relationship between the items, results indicate that the items are related in practice. However, correlations between one item pairing (5.1- choosing suitable measures and 5.2-implementing measures) were very strong, indicating item overlap. This is likely to be because a number of therapists did not employ any methods for measuring change, meaning that their rating for both items was ‘1- limited’ (because they failed to include the feature outlined).

Agreement between assessors for the ACCS total score was good (pooled ICC = .70). This is comparable or higher than inter-rater reliability achieved with the CTS-R (ICC range = .40 to .87, 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]). These results are encouraging given that good inter-rater reliability is notoriously difficult to achieve when assessing CBT competence (see chapter one). 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). Within the current study, OCTC assessors received no training, whilst assessors in First Step attended a one-day training session in how to use the ACCS. It was, therefore, surprising that inter-rater reliability for the total score was good within both sites (OCTC = .74, First Step = .73). Although this suggests that assessors may not require training in order to achieve good inter-rater reliability on the ACCS, it is important to recognise that assessors in OCTC had a great deal of prior experience and training in how to assess competence using other rating scales and that the second rater was involved in development of scale. Thus, the sample within OCTC may not be representative of assessors who would typically use the scale. Further research is therefore needed to establish whether adequate inter-rater agreement can be achieved on the ACCS without assessor training.

Inter-rater reliability for individual items ranged from fair to good (ICC for pooled data ranged from .35 to .74). Thus none of the individual items fell in the range of poor agreement. This is an improvement on the inter-rater reliability reported for individual CTS-R items, which shows poor agreement for a number of items (ICC = -.14 to .84 [Blackburn et al., 2001] pre-training *r* = .07 to .59, post-training *r* = .26 - .62 [Reichelt et al., 2003]). However, agreement between raters fell in the fair range (ICC = .21 - 0.40, pooled data set) for four items. These were the three items relating to the provision of homework (4.2- choosing suitable homework, 4.3- rationale for homework, and 4.4- planning homework) and item 6.1- pace. It is also worth noting that inter-rater reliability within First Step mirrored the findings outlined in the pooled data. However, within OCTC, five different items fell in the fair range (1.2- feasible agenda, 2.1- coherent formulation, 7.2- empathic understanding, 7.3- collaboration, and 8.1- patient feedback).

As therapists progressed through the Post Graduate Diploma in CBT, their level of competence on the ACCS improved significantly as did their level of competence on the CTSR. Thus, the ACCS appears to compare well with the CTSR in terms of discriminant validity and could provide a useful measure for measuring therapists' progress within CBT training programmes.

Limitations and future directions

The sample size employed in this pilot study is comparable to or larger than those used to develop other competence rating scales (range = nine to 26 therapists and 21 to 102 session recordings for the CATACS, CTS, CTS-R, CATS-Psy and YACS; Barber, Liese, & Abrams, 2003; Blackburn et al., 2001; Carroll et al., 2000; Dobson, Shaw, & Vallis, 1985; Gordon, 2006; Haddock et al., 2001; Vallis et al., 1986). However, finding therapists and assessors who were both willing and able to complete time consuming competence assessments was challenging. Consequently, the sample size is still relatively small and as such is a limitation of this study. Additionally, as only a small number of the CBT training courses and treatment centres approached were able to participate, there may be some degree of selection bias in the sample used. There are several risks to using a small sample to evaluate 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. A small sample also increases the likelihood that the sample may not be representative of the population for which the scale is intended. The sample may, for example, possess a narrower range of competence than would be expected of the population or the meaning ascribed to items in the sample may be atypical of the meaning of those items in the larger population of assessors and therapists. To try and minimise any such idiosyncrasies, two very different sample populations were recruited: novice CBT therapists taking part in a CBT training course and trained individuals delivering CBT within a routine practice setting. An additional benefit of this evaluation strategy is that it means that the evaluation of the scale did not only take place in the centre in which the scale was developed (OCTC). Instead, the ACCS was also piloted in a treatment centre with little prior experience of assessing CBT competence and no prior knowledge of or affiliation with the ACCS scale (First Step). Furthermore, analysis of the scale was carried out for both centres independently, as well as for the pooled data, with results being broadly comparable across both centres.

Another drawback to the relatively 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 ACCS has 22 items, meaning that between 110 and 220 ratings from *independent* therapists are required. Thus, in order to determine how many latent variables underlie the items, it will be necessary to collect a much larger sample size and to use this to carry out a factor analysis. It will also be important to determine the predictive validity of the ACCS scale- i.e. to examine whether the measure predicts improvements in patients'

symptoms. Although all of the aspects of competence included within the ACCS are theoretically important in delivering CBT, it is not actually known whether all of these aspects of competence are, in practice, necessary to achieve good patient outcomes. Further examination of this issue is vital given that the ultimate goal of delivering competent CBT is to alleviate patients' symptoms.

Ratings were completed by individuals who knew the therapists, which is realistic given that many training courses and routine practice settings use supervisors to rate competence. However, assessors' prior knowledge of the therapists may have influenced their ratings. It was possible to mitigate against some potential biases in First Step by ensuring that the ACCS ratings remained anonymous (i.e. were not shared with therapists). However, this was not possible within OCTC because the assessments formed part of a formative assessment process. Assessors also rated the same therapist more than once, meaning that rater confounds may have influenced the study results. Additionally, assessors within OCTC rated therapists on both the ACCS and the CTS-R. Although assessors were asked to complete ACCS ratings after CTS-R ratings, this instruction may not have been followed.

Patients in the recordings rated in the pilot study primarily presented with anxiety and/or depression and were largely judged to be somewhat straightforward cases. It is, therefore, not possible to draw conclusions regarding the validity or reliability of the ACCS when assessing the delivery of cognitive behavioural therapy to individuals with more complex, severe problems. It has been argued that generic rating scales are not suitable for assessing CBT competence within the context of complex cases (e.g. psychosis, borderline personality disorder, eating disorders) where differences in the nature of psychological processes impact on the way CBT is delivered (Gordon, 2006; Haddock et al., 2001). Thus, examination of the psychometric properties of the ACCS within the context of more complex patient presentations will be an important extension of the current study.

Conclusions

Despite its limitations, this initial evaluation shows that both the self-rated and assessor-rated ACCS demonstrate good internal consistency, inter-rater reliability, and discriminant validity. Hence the ACCS appears to be suitable for use in clinical practice, training settings and research studies and can be used as a self-rating tool as well as an assessor-rated tool. In addition, the ACCS was found to be correlated with, but distinct from the CTS-R and was comparable to the CTS-R in terms of internal consistency and discriminant validity and may have advantages over the CTS-R in terms of inter-rater reliability. Thus the ACCS appears to provide an acceptable alternative to the CTS-R.

References

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

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. doi: 10.1017/S1352465801004040

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal, 314*, 572. doi: 10.1136/bmj.314.7080.572

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

DeVellis, R. F. (2012). *Scale development: Theory and Applications*. Thousand Oask, CA: Sage.

Dobson, K. S., Shaw, B. F., & Vallis, T. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24*(4), 295-300. doi: 10.1111/j.2044-8260.1985.tb00662.x

Gordon, P. K. (2006). A comparison of two versions of the cognitive therapy scale. *Behavioural and Cognitive Psychotherapy, 35*(3), 343-353. doi: 10.1017/S1352465806003390

Gwet, K. L. (2010). *Handbook of inter-rater reliability (2nd ed.): The definitive guide to measuring the extent of agreement among multiple raters*. Gaithersburg: Advanced Analytics.

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. doi: 10.1017/S1352465801002089

Howell, W. S. (1982). *The empathic communicator*. Belmont, CA: Wadsworth.

Inskipp, F., & Proctor, B. (1993). *Making the most of supervision: professional development for counsellors, psychotherapists, supervisors and trainees*. London: Cascade.

James, I. A., Blackburn, I. M., Milne, D. L., & Reichfelt, F. K. (2001). Moderators of trainee therapists' competence in cognitive therapy. *British Journal of Clinical Psychology, 40*, 131-141. doi: 10.1348/014466501163580

Lahey, M. A., Downey, R. E., & Saal, F. E. (1983). Intraclass correlations: There's more than meets the eye. *Psychological Bulletin, 93,* 586 - 595.

Loewenthal, K. M. (2001). *An introduction to psychological tests and scales (2nd ed.)*. Philadelphia: Taylor & Francis Inc.

McManus, F., Rakovshik, S., Kennerley, H., Fennell, M., & Westbrook, D. (2011). An investigation of the accuracy of therapists' self-assessment of cognitive-behaviour therapy skills. *British Journal of Clinical Psychology, 51,* 292 - 306. doi: 10.1111/j.2044-8260.2011.02028.x

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. doi: 10.1016/j.brat.2010.08.002

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research and Evaluation, 9*(11). Retrieved August 13, 2014 from http://PAREonline.net/getvn.asp?v=9&n=11 .

Reichelt, F., 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. doi: 10.1016/S0005-7916(03)00022-3

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

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use (3rd ed.)*. Oxford: Oxford University Press.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Use of factor analysis in counselling psychology research. *Journal of counseling psychology, 34*, 414 - 424.

Vallis, T., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54*(3), 381-385. doi: 10.1037/0022-006X.54.3.381

Williams, R. M., Moorey, S., & Cobb, J. (1991). Training in cognitive-behaviour therapy: Pilot evaluation of a training course using the cognitive therapy scale. *Behavioural and Cognitive Psychotherapy, 19*(04), 373-376. doi: 10.1017/S0141347300014075

Appendices

Appendix 1. Ethical Consent Approval

MEDICAL SCIENCES INTER DIVISIONAL RESEARCH ETHICS COMMITTEE
Medical Sciences Divisional Office

University of Oxford, Level 3, John Radcliffe Hospital, Headington, Oxford, OX3 9DU
Tel: +44(0)1865 851046 Fax: +44(0)1865 750750
ethics@medsci.ox.ac.uk http://www.medsci.ox.ac.uk



CONFIDENTIAL

Ref: MSD-IDREC-C1-2013-173

Miss Kate Muse
Powic Building
Dept of Psychiatry
Warneford Hospital

19th November 2013

OX3 7JX

Dear Miss Muse

CUREC checklist

I am writing to acknowledge receipt of your CUREC-1 form for your project: **A psychometric review of a novel Cognitive Behavioural Therapy (CBT) competence rating scale.**

On the basis of the information you have provided this has now been approved by the Medical Sciences IDREC **subject to:**

a) **it is your responsibility to comply with the requirements for administering any tests or questionnaires and if in doubt to contact the publisher of those tests or questionnaires.**

The reference number for this project is **MSD-IDREC-C1-2013-173** and is valid for a period of 11 months from CUREC 1 approval date, 8th October 2013. Please may I remind you that your project may be reviewed at some stage during an annual audit of projects.

Amendments

Should you at some stage alter some of the techniques or procedures then you should first undertake a checklist (CUREC-1) to see whether these changes alter the ethics of the research. If these remain the same then the committee will require notification of the changes to lodge with the project. If they do not remain the same then you may need to complete a CUREC-2 form and undergo further scrutiny by the committee.

Please do not hesitate to contact me if you have any queries about this.

Yours sincerely

Gill Halstead
Acting Secretary to Medical Sciences IDREC

Appendix 2. Invitation to Participate

Template Written Invitation to Participate

An invitation to participate in a psychometric review of a novel Cognitive Behavioural Therapy (CBT) competence rating scale

We have developed a new Cognitive Behavioural Therapy (CBT) competence rating scale which can be used to provide formative and summative feedback on therapists' CBT skills . We would like to investigate the psychometric properties of the rating scale in order to evaluate and further refine the scale. We are asking all students on the Postgraduate Diploma in Cognitive Behavioural Therapy course (PGDip) run by the Oxford Cognitive Therapy Centre (OCTC) whether they would be willing to participate in the study.

Taking part would involve two key elements. First, your supervisor would rate the session recordings you routinely submit for evaluation during the course using our rating scale as well as the Cognitive Therapy Scale (revised - CTS-R). Second, we would also ask you to rate your own performance in these sessions using our scale. I have attached an information sheet which provides further details about why the study is being conducted and what your involvement would be. If you have any questions or quieries about the study then please feel free to ask us.

Thank you very much for taking the time to consider participating.

Best wishes,

Kate Muse, Freda McManus & Sarah Rakovshik

Appendix 3: Participant Information Sheet: Oxford Cognitive Therapy Centre

UNIVERSITY OF OXFORD
DEPARTMENT OF PSYCHIATRY

WARNEFORD HOSPITAL
OXFORD
OX3 7JX

TEL: 07834817388
EMAIL: kate.muse@psych.ox.ac.uk

PARTICIPANT INFORMATION SHEET

A psychometric review of a novel Cognitive Behavioural Therapy (CBT) competence rating scale

Principal Researcher: Kate Muse, Supervisors: Dr Freda McManus & Dr. Sarah Rakovshik

You are being invited to take part in a research study that aims to evaluate the psychometric properties (e.g., reliability, validity, usability) of a new rating scale for assessing CBT competence. In order to decide whether you would like to take part, please read 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 you decide whether to participate. Please contact the principal researcher (Kate Muse) using the contact details outlined on page 3 with any questions.

Why have I been contacted?

You have contacted because we are seeking to recruit students on the Postgraduate Diploma in Cognitive Behavioural Therapy course (PGDip) run by the Oxford Cognitive Therapy Centre (OCTC).

What is the purpose of the study?

There is a need for improved methods of assessing CBT competence, particularly those which assess whether therapists can demonstrate the skills necessary to effectively deliver CBT. We have developed a new CBT competence rating scale (Assessment of Core CBT Skills: ACCS) which can be used to provide formative and summative feedback on therapists' CBT skills within observed treatment sessions. The central aim in developing this scale is to provide a tool which improves upon currently available scales (particularly the Revised Cognitive Therapy Scale: CTS-R) in terms of validity, reliability and usability. The current 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 study is being conducted by Kate Muse as a part of a DPhil at the Department of Psychiatry, University of Oxford, under supervision of Dr Freda McManus and Dr Sarah Rakovshik. The research has been funded by the British Association of Behavioural and Cognitive Psychotherapists (BABCP). 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 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 also our intention publish the results of the study in 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?

There are two key aspects to participation in this study:

1. As you are aware, as part of your training and evaluation on the PGDip in CBT you are required to submit recordings of six CBT sessions. These sessions are rated by your supervisor using the CTS-R. If you agree to take part, we would ask your supervisor to rate your performance within these sessions on our new ACCS scale as well as the CTS-R. You would then be provided with feedback on your performance using both the ACCS and the CTS-R. **Only your CTS-R rating will contribute towards your grading within the PGDip course.** You will not be asked to submit any additional session recordings for the purposes of this study. As both ratings will be carried out by your supervisor, your session recordings will **not** be viewed by any individuals other than your supervisor who will be listening to them in any case to make the ratings on the CTS-R.

2. In addition, as part of your training on the PG Dip. in CBT, you are asked to complete self-ratings of your performance (using the CTS-R) for each of the six recordings of CBT sessions that you submit to your supervisor for evaluation on the CTS-R. If you agree to participate in this study, we will ask you to provide self-ratings on the ACCS scale in addition to the CTS-R. It is anticipated that participation would involve spending approximately one hour reading through and familiarising yourself with the ACCS scale. After this, each self-rating which you complete on the ACCS is expected to take approximately ten minutes.

If you agree to participate, we will ask you to provide consent for your supervisor to rate your sessions using the ACCS in addition to the CTS-R and for you to provide self-ratings of your sessions using the ACCS in addition to the CTS-R separately. Thus you can agree to participate in only one aspect of the study. For both aspects of the study, we will also ask you to complete a short demographics form which will take approximately five minutes to complete.

All information collected (supervisor and self-ratings, demographics form) will be identified by a unique ID number, rather that your name, and the ACCS and CTS-R ratings will be separated from the 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. It is important to note that there is no academic penalty for non-participation or withdrawal from this study.

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. performance ratings) 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 the University of Oxford 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?

The information you provide will be beneficial in evaluating the reliability, validity and usability of the ACCS scale and will also help to highlighting areas where the scale requires refinement. Thus, your participation in this study will help to improve current methods of assessing CBT competence. By taking part in this study you will also be provided with additional feedback about your performance which may be of use in helping to guide your future learning.

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 demographic 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 Research Ethics Committee at the University of Oxford (ethics@medsci.ox.ac.uk; Medical Sciences Inter-Divisional Research Ethics Committee, Medical Sciences Divisional Office, Level 3 John Radcliffe Hospital, Oxford, OX3 9DU, UK).

Who has reviewed this study?

The study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number MSD-IDREC-C1-2013-173).

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:	**Supervisor:**	**Supervisor:**
Kate Muse	Dr. Freda McManus	Dr. Sarah Rakovshik
Department of Psychiatry	Oxford Cognitive Therapy Centre	Oxford Cognitive Therapy Centre
Warneford Hospital	Warneford Hospital	Warneford Hospital
Oxford	Oxford	Oxford
OX3 7JX	OX3 7JX	OX3 7JX
Tel: 078348 17388	Tel: 01865 738816	Tel: 01865 738816
Email:	Email:	Email:
kate.muse@psych.ox.ac.uk	freda.mcmanus@psych.ox.ac.uk	rakovsh@post.harvard.edu

Appendix 4: Consent Form: Oxford Cognitive Therapy Centre

UNIVERSITY OF OXFORD
DEPARTMENT OF PSYCHIATRY

WARNEFORD HOSPITAL
OXFORD
OX3 7JX

TEL: 07834817388
EMAIL: kate.muse@psych.ox.ac.uk

CONSENT FORM

A psychometric review of a novel Cognitive Behavioural Therapy (CBT) competence rating scale

Principal Researcher: Kate Muse, Supervisor: Dr Freda McManus & Dr. Sarah Rakovshik

Study Purpose: The study aims to investigate the psychometric properties of the rating scale in order to evaluate and further refine the scale.

Please initial box

I confirm that I have read and understand the information sheet dated 07.08.13 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 the University of Oxford and other regulatory authorities, for the purpose of 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.	☐
In understand that the study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number MSD-IDREC-C1-2013-173).	☐
I understand how to raise a concern or make a complaint.	☐
I agree that my supervisor can rate my performance within treatment sessions submitted for evaluation on the PGDip in CBT on the ACCS scale as well as the CTS-R and that these anonymised ratings can be used for purposes of scale evaluation.	☐
I agree to submit self-ratings of my performance on the ACCS in addition to the CTS-R to my supervisor and agree that these anonymised ratings can be used for purposes of scale evaluation.	☐
I agree to complete a demographics form.	☐

Name of Participant	Date	Signature
Researcher	Date	Signature

Appendix 5: Participant Information Sheet: First Step

UNIVERSITY OF OXFORD

DEPARTMENT OF PSYCHIATRY

WARNEFORD HOSPITAL
OXFORD
OX3 7JX

TEL: 07834817388
EMAIL: kate.muse@psych.ox.ac.uk

PARTICIPANT INFORMATION SHEET

A psychometric review of a novel Cognitive Behavioural Therapy (CBT) competence rating scale- First Step Cumbria

Principal Researcher: Kate Muse, Supervisors: Dr Freda McManus & Dr. Sarah Rakovshik

You are being invited to take part in a study that aims to evaluate the psychometric properties (e.g., reliability, validity, usability) of a new rating scale for assessing CBT competence. 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 (Kate Muse) using the contact details outlined on page 3.

Why have I been contacted?
You have contacted because we are seeking to recruit CBT therapists within the First Step service.

What is the purpose of the study?
There is a need for improved methods of assessing CBT competence, particularly those which assess whether therapists can demonstrate the skills necessary to effectively deliver CBT. We have developed a new CBT competence rating scale (Assessment of Core CBT Skills: ACCS) which can be used to provide formative and summative feedback on therapists' CBT skills within observed treatment sessions. The central aim in developing this scale is to provide a tool which improves upon currently available scales (particularly the Revised Cognitive Therapy Scale: CTS-R) in terms of validity, reliability and usability. The current 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 study is being conducted by Kate Muse as a part of a DPhil at the Department of Psychiatry, University of Oxford, under supervision of Dr Freda McManus and Dr Sarah Rakovshik. The research has been funded by the British Association of Behavioural and Cognitive Psychotherapists (BABCP). It is our intention publish the results of the study in 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?
As you are aware, as part of your continuing professional development (CPD), you will be asked to submit recordings of CBT treatment sessions to your clinical supervisors for feedback on your performance. Participation in the current study would involve one of these recordings being viewed and rated using the ACCS scale by three people:

1. Your supervisor within the First Step service
2. A Senior Psychotherapist within the First Step service
3. Yourself (i.e. a self-rating)

It is anticipated that completing a self-rating will involve spending approximately one hour reading through and familiarising yourself with the ACCS scale and then completing the rating scale, which involved rating your own performance on eight aspects of CBT competence. We will also ask you 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 Cumbria Partnership NHS Foundation Trust staff employed within the First Step service. The ACCS ratings completed by therapists and assessors will be sent directly to Emma Cowperthwaite (personal assistant to Dr Richard Thwaites- Clinical Lead for Non-Psychosis Pathways at First Step) without being shared with any other parties. This means that none of the other parties (i.e. your supervisor or senior psychologist within the service) will see your self-ratings or the ratings completed by the other member of staff.

Only anonymised demographics forms and rating sheets (i.e. paperwork identified by a unique ID number rather than your name) will be sent to Kate Muse- principal investigator- at the Department of Psychiatry, University of Oxford to ensure anonymity.

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. performance ratings) 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 the University of Oxford 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?

The information you provide will be beneficial in evaluating the reliability, validity and usability of the ACCS scale and will also help to highlighting areas where the scale requires refinement. Thus, your participation in this study will help to improve current methods of assessing CBT competence. By taking part in this study you will also have the opportunity to reflect upon your performance with the session recording, which may be of use in helping to guide your future learning.

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 demographic questionnaire happens to include any questions which, for whatever reason, you do not wish to answer then the question can be omitted.

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:	**Supervisor:**	**Supervisor:**
Kate Muse	Dr. Freda McManus	Dr. Sarah Rakovshik
Department of Psychiatry	Oxford Cognitive Therapy Centre	Oxford Cognitive Therapy Centre
Warneford Hospital	Warneford Hospital	Warneford Hospital
Oxford	Oxford	Oxford
OX3 7JX	OX3 7JX	OX3 7JX
Tel: 078348 17388	Tel: 01865 738816	Tel: 01865 738816
Email:	Email:	Email:
kate.muse@psych.ox.ac.uk	freda.mcmanus@psych.ox.ac.uk	rakovsh@post.harvard.edu

Appendix 6: Cognitive Therapy Scale-Revised

THE REVISED COGNTIVE THERAPY SCALE AUGUST 2000

Scorer:..... Date:.....

COGNITIVE THERAPY SCALE - REVISED (CTS-R)

The rating of the scale

The present seven point scale (i.e. a 0-6 Likert scale) extends from (0) where the therapist did not adhere to that aspect of therapy (non-adherence) to (6) where there is adherence and very high skill. Thus the scale assesses both adherence to therapy method and skill of the therapist. To aid with the rating of items of the scale, an outline of the key features of each item is provided at the top of each section. A description of the various rating criteria is given in the right hand margin - see example below in Figure 1. Further details are provided in the accompanying manual.

The examples are intended to be used as useful guidelines only. They are not meant to be used as prescriptive scoring criteria, rather providing both illustrative anchor points and guides.

Adjusting the scale in the presence of patient difficulties

The scale's dimensions were devised for patients assessed as being well/moderately suited for cognitive therapy (Safran & Segal, 1990). As such, adjustments may need to be made when patient difficulties are evident (e.g. excessive avoidance). Indeed, with problematic patients it is sometimes difficult to apply CT methods successfully; that is, with desirable change. In such circumstances the rater needs to assess the therapist's therapeutic skills in the application of the methods. Thus even though the therapist may be unsuccessful at promoting change, credit should be given for demonstrations of appropriate skilful therapy.

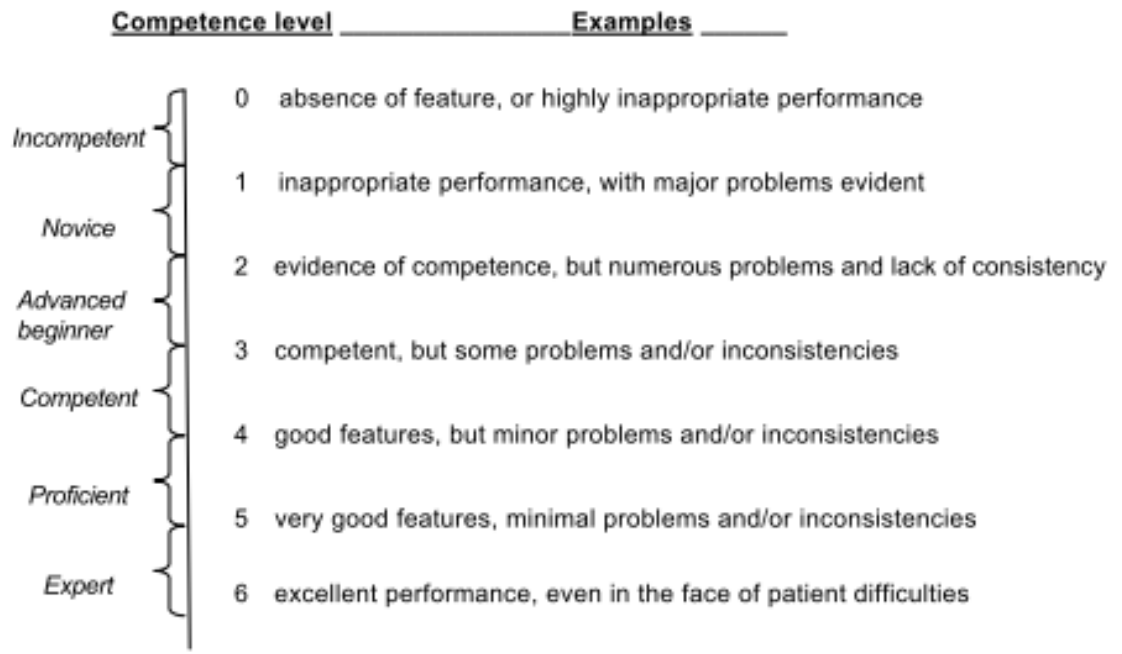
Safran, J.D. & Segal, Z.V. (1990) *Interpersonal* processes *in* cognitive *therapy.* New York, Basic Books.

Figure 1: Example of the scoring layout

Key features: this is an operationalised description of the item (see examples within the CTS-R).

Mark with an 'X' on the vertical line, using whole and half numbers, the level to which you think the therapist has fulfilled the core function. The descriptive features on the right are designed to guide your decision.

N.B. When rating, take into consideration the appropriateness of therapeutic interventions for stage of therapy and perceived patient difficulty.



* The present scale has incorporated the Dreyfus system (Dreyfus, 1989) for denoting competence, which is described fully in the manual. Please note that the 'top marks (i.e. near the 'expert' end of the continuum) are reserved for those therapists demonstrating highly effective skills, particularly in the face of difficulties (i.e. highly aggressive or avoidant patients; high levels of emotional discharge from the patients; and various situational factors).

The `**Key Features'** describe the important features that need to be considered when scoring each item. When rating the item, you must first identify whether some of the features are present. You must then consider whether the therapist should be regarded as competent with the features. 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 very highly.

The `**Examples'** are only guidelines and should not be regarded as absolute rating criteria.

Scoring Distribution

It is important to remember that the scoring profile for this scale should approximate to a normal distribution (i.e. mid-point 3), with relatively few therapists scoring at the extremes.

Dreyfus, H. L. (1989). The Dreyfus model of skill acquisition. *In* J. Burke (ed.) *Competency based education and training*. London: Falmer Press.

THE REVISED COGNTIVE THERAPY SCALE AUGUST 2000

ITEM 1 - AGENDA SETTING & ADHERENCE

Key features: To address adequately topics that have been agreed and set in an appropriate way. This involves the setting of discrete and realistic targets collaboratively.. The format for setting the agenda may vary according to the stage of therapy - see manual.

Three features need to be considered when scoring this item:

(i) presence/absence of an agenda which is explicit, agreed and prioritised, and feasible in the time available;

(ii) appropriateness of the contents of the agenda (to stage of therapy, current concerns etc.), a standing item being a review of the homework set previously;

(iii) appropriate adherence to the agenda.

Mark with an 'X' on the vertical line, the level to which you think the therapist has fulfilled the core function. The descriptive features on the right are designed to guide your decision.

NB: Agenda setting requires collaboration and credit for this should be given here, and here alone. Collaboration occurring at any other phase of the session should be scored under Item 3 (Collaboration).

Competence level	Examples NB: Score according to features, not examples!
0	No agenda set, highly inappropriate agenda set, or agenda not adhered to.
1	Inappropriate agenda set (e.g. lack of focus, unrealistic, no account of patient's presentation, homework not reviewed).
2	An attempt at an agenda made, but major difficulties evident (e.g. unilaterally set). Poor adherence.
3	Appropriate agenda, which was set well, but some difficulties evident (e.g. poor collaboration). Some adherence.
4	Appropriate agenda, minor difficulties evident (e.g. no prioritisation), but appropriate features covered (e.g. review of homework). Moderate adherence.
5	Appropriate agenda set with discrete and prioritised targets - review at the end. Agenda adhered to. Minimal problems.
6	Excellent agenda set, or highly effective agenda set in the face of difficulties.

THE REVISED COGNTIVE THERAPY SCALE AUGUST 2000

ITEM 2 - FEEDBACK

Key features: The patient's and therapist's understanding of key issues should be helped through the use of two-way feedback: The two major forms of feeding back information are through general summary and chunking of important units of information. The use of appropriate feedback helps both the therapist to understand the patient's situation, and the patient to synthesise material enabling him/her to gain major insight and make therapeutic shifts. It also helps to keep the patient focused.

Three features need to be considered when scoring this item:

(i) presence and frequency, or absence, of feedback. Feedback should be given/elicited throughout the therapy - with major summaries both at the beginning (review of week) and end (session summary), while topic reviews (i.e. chunking) should occur throughout the session;
(ii) appropriateness of the contents of the feedback;
(iii) manner of its delivery and elicitation (NB: can be written).

Competence level	Examples NB: Score according to features, not examples!
—	Absence of feedback or highly inappropriate feedback.
0	Minimal appropriate feedback (verbal and/or written).
1	Appropriate feedback, but not given frequently enough by therapist, with insufficient attempts to elicit and give feedback, e.g. feedback too vague to provide opportunities for understanding and change.
2	Appropriate feedback given and elicited frequently, although some difficulties evident in terms of content or method of delivery.
3	Appropriate feedback given and elicited frequently, facilitating moderate therapeutic gains. Minor problems evident (eg. inconsistent).
4	Highly appropriate feedback given and elicited regularly, facilitating shared understanding and enabling significant therapeutic gains. Minimal problems.
5	Excellent use of feedback, or highly effective feedback given and elicited regularly in the face of difficulties.
6	

ITEM 3 - COLLABORATION

Key features: The patient should be encouraged to be active in the session. There must be clear evidence of productive teamwork, with the therapist skillfully encouraging the patient to participate fully (e.g. through questioning techniques, shared problem solving and decision making) and take responsibility. However, the therapist must not allow the patient to ramble in an unstructured way.

Three features need to be considered: the therapist style should encourage effective teamwork through his/her use of:

(i) verbal skills (e.g. non-hectoring);
(ii) non-verbal skills (e.g. attention and use of joint activities);
(iii) sharing of written summaries.

NB: Questioning is a central feature with regard to this item, but questions designed to facilitate reflections and self discovery should be scored under Item 9 (Guided Discovery).

Competence level	Examples NB: Score according to features, not examples!
0	Patient is actively prevented or discouraged from being collaborative.
1	The therapist is too controlling, dominating, or passive.
2	Some occasional attempt at collaboration, but didactic style or passivity of therapist encourages passivity or other problems in the therapeutic relationship.
3	Teamwork evident, but some problems with collaborative set (e.g. not enough time allowed for the patient to reflect and participate actively).
4	Effective teamwork is evident, but not consistent. Minor problems evident.
5	Effective teamwork evident throughout most of the session, both in terms of verbal content and use of written summaries. Minimal problems.
6	Excellent teamwork, or highly effective teamwork in the face of difficulties.

ITEM 4 - PACING AND EFFICIENT USE OF TIME

Key features: The session should be well 'time managed' in relation to the agenda, with the session flowing smoothly through discrete start, middle, and concluding phases. The work must be paced well in relation to the patient's needs, and while important issues need to be followed, unproductive digressions should be dealt with smoothly. The session should not go over time, without good reason.

Three features need to be considered:

(i) the degree to which the session flows smoothly through the discrete phases;
(ii) the appropriateness of the pacing throughout the session;
(iii) the degree of fit to the learning speed of the patient.

Competence level	Examples NB: Score according to features, not examples!
0	Poor time management leads either to an aimless or overly rigid session.
1	The session is too slow or too fast for the current needs and capacity of the patient.
2	Reasonable pacing, but digression or repetitions from therapist and/or patient lead to inefficient use of time; unbalanced allocation of time, over time.
3	Good pacing evident some of the time, but diffuse at times. Some problems evident.
4	Balanced allocation of time with discrete start, middle and concluding phases evident. Minor problems evident.
5	Good time management skills evident, session running smoothly. Therapist working effectively in controlling the flow within the session. Minimal problems.
6	Excellent time management, or highly effective management evident in the face of difficulties.

ITEM 5 - INTERPERSONAL EFFECTIVENESS

Key features: The patient is put at ease by the therapist's verbal and non-verbal (e.g. listening skills) behaviour. The patient should feel that the core conditions (i.e. warmth, genuineness, empathy and understanding) are present. However, it is important to keep professional boundaries. In situations where the therapist is extremely interpersonally effective, he/she is creative, insightful and inspirational.

Three features need to be considered:

(i) empathy - the therapist is able to understand and enter the patient's feelings imaginatively and uses this understanding to promote change;
(ii) genuineness - the therapist has established a trusting working relationship;
(iii) warmth - the patient seems to feel liked and accepted by the therapist.

Competence level	Examples NB: Score according to features, not examples!
0	Therapist's manner and interventions make the patient disengage and become distrustful and/or hostile (absence of/or excessive i, ii, iii).
1	Difficulty in showing empathy, genuineness and warmth.
2	Therapist's style (e.g. intellectualisation) at times impedes his/her empathic understanding of the patient's communications.
3	The therapist is able to understand explicit meanings of patient's communications, resulting in some trust developing. Some evidence of inconsistencies in sustaining relationship.
4	The therapist is able to understand the implicit, as well as the explicit meanings of the patient's communications and demonstrates it in his/ her manner. Minor problems evident (e.g. inconsistent).
5	The therapist demonstrates very good interpersonal effectiveness. Patient appears confident that he/she is being understood, which facilitates self-disclosure. Minimal problems.
6	Highly interpersonally effective, even in the face of difficulties.

ITEM 6 — ELICITING OF APPROPRIATE EMOTIONAL EXPRESSION

Key features: The therapist facilitates the processing of appropriate levels of emotion by the patient. Emotional levels that are too high or too low are likely to interfere with therapy. The therapist must also be able to deal effectively with emotional issues which interfere with effective change (e.g. hostility, anxiety, excessive anger). Effective facilitation will enable the patient to access and express his/her emotions in a way that facilitates change.

Three features have to be considered:

(i) facilitation of access to a range of emotions;
(ii) appropriate use and containment of emotional expression;
(iii) facilitation of emotional expression; encouraging appropriate access and differentiation of emotions.

Competence Level	Examples NB. Score according to features, not examples!
0	Patient is under- or over stimulated (e,g, his her feelings are ignored or dismissed or allowed to reach an unmanaged pitch). Or the therapist's own mood or strategies (e.e. intellectualization) adversely influences the session.
1	Failure to facilitate access to, and expression of, appropriate emotional expression.
2	Facilitation of appropriate emotional expression evident, but many relevant opportunities missed.
3	Some effective facilitation of appropriate emotional expression, created and/or maintained. Patient enabled to become slightly more aware.
4	Effective facilitation of appropriate emotional expression leading to the patient becoming more aware of relevant emotions. Minor problems evident.
5	Very effective facilitation of emotional expression, optimally arousing the patient's motivation and awareness. Good expression of relevant emotions evident- done in an effective manner. Minimal problems.
6	Excellent facilitation of appropriate emotional expression, or effective facilitation in the face of difficulties

ITEM.7 - ELICITING KEY COGNITIONS

Key features: To help the patient gain access to his/her cognitions (thoughts, assumptions and beliefs) and to understand the relationship between these and their distressing emotions. This can be done through the use of questioning, diaries and monitoring procedures.

Three features need to be considered:

(i) eliciting cognitions that are associated with distressing emotions (i.e. selecting key cognitions or hot thoughts);
(ii) the skillfulness and breadth of the methods used (i.e. Socratic questioning; appropriate monitoring, downward arrowing, imagery, role-plays, etc.);
(iii) choosing the appropriate level of work for the stage of therapy (i.e. automatic thoughts, assumptions, or core beliefs).

NB: This item is concerned with the general work done with eliciting cognitions. If any specific cognitive or behavioural change methods are used, they should be scored under item 11 (change methods).

Competence level	Examples NB: Score according to features, not examples!
0	Therapist fails to elicit relevant cognitions.
1	Inappropriate cognitions and emotions selected, or key cognitions/emotions ignored.
2	Some cognitions/emotions (or one key cognition, e.g. core belief) elicited, but links between cognitions and emotions not made clear to patient.
3	Some cognitions/emotions (or one key cognition) elicited in a competent way, although some problems evident.
4	A number of cognitions and emotions (or one key cognition) elicited in verbal or written form, leading to a new understanding of their relationship. Minor problems evident.
5	Effective eliciting and selection of a number of cognitions/emotions (or one key cognition), which are generally dealt with appropriately. Minimal problems.
6	Excellent work done on key cognition(s) and emotion(s), even in the face of difficulties.

ITEM 8 - ELICITING AND PLANNING BEHAVIOURS

Key features: To help the patient gain insight into the effect of his/her behaviours with respect to the problems. This can be done through the use of questioning; diaries and monitoring procedures. The therapist works with the patient to plan strategies either to overcome or disrupt dysfunctional behavioural patterns.

Two features need to be considered:

(i) eliciting behaviours and plans that are associated with distressing emotions;
(ii) the skilfulness and breadth of the methods used (i.e. Socratic questioning; appropriate monitoring, downward arrowing, imagery, role-plays, etc.);

NB: This item is concerned with the general work done with eliciting behaviours and plans. If any specific cognitive or behavioural change methods are used, they should be scored under item 1 1 (change methods).

Competence level	Examples NB: Score according to features, not examples!
0	Therapist fails to elicit relevant behaviours and plans.
1	Inappropriate behaviours focused on and/or plans generated.
2	Some behaviours and plans elicited, but links between behaviours, cognitions and emotions not made clear to patient.
3	Some behaviours and plans elicited in a competant way, although some problems evident.
4	A number of behaviours and plans elicited in verbal or written form, leading to a new understanding of their importance in maintaining problems. Minor difficulties evident.
5	Effective eliciting and selection of a number of behaviours and plans, which are generally dealt with appropriately. Minimal problems.
6	Excellent work done on behaviours and plans, even in the face of difficulties.

ITEM 9 - GUIDED DISCOVERY

Key features: The patient should be helped -to -develop hypotheses regarding his/her current situation and to generate potential solutions' for him/herself. The patient is helped to develop a range of perspectives regarding his/her experience. Effective guided discovery will create doubt where previously there was certainty, thus providing the opportunity for re-evaluation and new learning to occur.

Two elements need to be considered:

(i) the style of the therapist - this should be.open and inquisitive;
(ii) the effective use of questioning techniques (e.g. Socratic questions) should encourage the patient to discover useful information that can be used to help him/her to gain a better level of understanding.

Competence level	Examples NB: Score according to features, not examples!
0	No attempt at guided discovery (e.g. hectoring and lecturing).
1	Little opportunity for discovery by patient. Persuasion and debate used excessively.
2	Minimal opportunity for discovery. Some use of questioning, but unhelpful in assisting the patient to gain access to his/her thoughts or emotions or to make connections between themes.
3	Some reflection evident. Therapist uses primarily a questioning style which is following a productive line of discovery.
4	Moderate degree of discovery evident. Therapist uses a questioning style with skill, and this leads to some synthesis. Minor problems evident.
5	Effective reflection evident. Therapist uses skilful questioning style leading to reflection, discovery, and synthesis. Minimal problems.
6	Excellent guided discovery leading to a deep patient understanding. Highly effective discovery produced in the face of difficulties, with evidence of a deeper understanding having been developed.

THE REVISED COGNTIVE THERAPY SCALE AUGUST 2000

ITEM 10 - CONCEPTUAL INTEGRATION

Key features: The patient should be helped to gain an appreciation of the history, triggers and maintaining features of his/her problem in order to bring about change in the present and future. The therapist should help the patient to gain an understanding of how his/her perceptions and interpretations, beliefs, attitudes and rules relate to his/her problem. A good conceptualisation will examine previous cognitions and coping strategies as well as current ones. This theory-based understanding should be well integrated and used to guide the therapy forward.

Two features need to be considered:

(i) the presence/absence of an appropriate conceptualisation which is in line with goals of therapy;
(ii) the manner in which the conceptualisation is used (e.g. used as the platform for interventions, homework etc.).

NB: This item is to do with therapeutic integration (using theory to link present, past and future). If the therapist deals specifically with cognitions and emotions, this should be scored under Items 6 (Facilitation of Emotional Expression) and 7 (Eliciting Key of Cognitions)

Competence level	Examples NB: Score according to features, not examples!
0	The absence of an appropriate conceptualisation.
1	The lack, or inappropriateness or misapplication of a conceptualisation leads to a neutral impact (e.g. interferes with progress or leads to aimless application of procedures).
2	Some rudimentary conceptualisation arrived at, but not well integrated with goals of therapy. Does not lead to a clear rationale for interventions.
3	Cognitive conceptualisation partially developed with some integration, but some difficulties evident (e.g. in synthesising and in sharing it with the patient). Leads to coherent interventions.
4	Cognitive conceptualisation is moderately developed and integrated within the therapy. Minor problems evident.
5	Cognitive conceptualisation is very well developed and integrated within the therapy - there is a credible cognitive understanding leading to major therapeutic shifts. Minimal problems.
6	Excellent development and integration evident, or highly effective in the face of difficulties.

THE REVISED COGNTIVE THERAPY SCALE AUGUST 2000

ITEM 11- APPLICATION OF CHANGE METHODS

Key features: Therapist skillfully uses, and helps the patient to use, appropriate cognitive and behavioural techniques in line with the formulation. The therapist helps the patient devise appropriate cognitive methods to evaluate the key cognitions associated with distressing emotions, leading to major new perspectives and shifts in emotions. The therapist also helps the patient to apply behavioural techniques in line with the formulation. The therapist helps the patient to identify potential difficulties and think through the cognitive rationales for performing the tasks. The methods provide useful ways for the patient to test-out cognitions practically and gain experience in dealing with high levels of emotion. The methods also allow the therapist to obtain feedback regarding the patient's level of understanding of prospective practical assignments (i.e. by the patient performing the task in- session).

Three features need to be considered:

(i) the appropriateness and range of both cognitive methods (e.g. cognitive change diaries, continua, distancing, responsibility charts, evaluating alternatives, examining pros and cons, determining meanings, imagery restructuring, etc.) and behavioural methods (e.g. behavioural diaries, behavioural tests, role play, graded task assignments, response prevention, reinforcement of patient's work, modeling, applied relaxation, controlled breathing, etc.);
(ii) the skill in the application of the methods - however, skills such as feedback, interpersonal effectiveness, etc. should be rated separately under their appropriate items;
(iii) the suitability of the methods for the needs of the patient (i.e. neither too difficult nor complex).

NB: This item is not concerned with accessing or identifying thoughts, rather with their re-evaluation.

Competence level	Examples NB: Score according to features, not examples!
0	Therapist fails to use or misuses appropriate cognitive and behavioural methods.
1	Therapist applies either insufficient or inappropriate methods, and/or with limited skill or flexibility.
2	Therapist applies appropriate methods, but major difficulties evident.
3	Therapist applies a number of methods in competent ways, although some problems evident (e.g. the interventions are incomplete).
4	Therapist applies a range of methods with skill and flexibility, enabling the patient to develop new perspectives. Minor problems evident.
5	Therapist systematically applies an appropriate range of methods in a creative, resourceful and effective manner. Minimal problems.
6	Excellent range and application, or successful application in the face of difficulties.

THE REVISED COGNTIVE THERAPY SCALE AUGUST 2000

ITEM 12 - HOMEWORK SETTING

Key features: This aspect concerns the setting of an appropriate homework task, one with clear and precise goals. The aims should be to negotiate an appropriate task for the stage of therapy in line with the conceptualisation; to ensure the patient understands the rationale for undertaking the task; to test out ideas, try new experiences, predict and deal with potential obstacles, and experiment with new ways of responding.

There are three aspects to this item:

(i) presence/absence of a homework task in which clear and precise goals have been set;
(ii) the task should be derived from material discussed in the session, such that there is a clear understanding of what will be learnt from performing the task;
(iii) the homework task should be set jointly, and sufficient time should be allowed for it to be explained clearly (i.e. explain, discuss relevance, predict obstacles, etc.).

NB: Review of homework from the previous session should be rated in Item 1(Agenda Setting)

Competence level	Examples NB: Score according to features, not examples!
0	Therapist fails to set homework, or sets inappropriate homework.
1	Therapist does not negotiate homework. Insufficient time allotted for adequate explanation, leading to ineffectual task being set.
2	Therapist negotiates homework unilaterally and in a routine fashion, without explaining the rationale for new homework.
3	Therapist has set an appropriate new homework task, but some problems evident (e.g. not explained sufficiently and/or not developed jointly).
4	Appropriate new homework jointly negotiated with a clear goals and rationales. However, minor problems evident.
5	Appropriate homework negotiated jointly and explained well, including an exploration of potential obstacles. Minimal problems.
6	Excellent homework negotiated, or appropriate one set in the face of difficulties.

Appendix 7: Numerical ACCS Feedback form used to rate the therapist's performance within recorded treatment sessions.

Assessment of Core CBT Skills (ACCS)

FEEDBACK FORM

Numerical Form

Version 2- October 2013

Oxford Cognitive Therapy Centre

UNIVERSITY OF OXFORD

BABCP www.babcp.com

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus and was conducted in affiliation with the University of Oxford's Department of Psychiatry and the Oxford Cognitive Therapy Centre (OCTC).

ACCS Feedback Form

Therapist's Name: _____ **Assessor's Name:** _____

Date of Submission: _____

For each item, please assign a rating of:

1. Limited **2. Basic** **3. Good** **4. Advanced**

Domain 1 - Agenda Setting

1.1. Suitable Items	
1.2. Feasible Agenda	
Total Domain Score *(average item score)*	

See user manual page 7 – 8

Domain 2 - Formulation

2.1. Coherent Formulation	
Total Domain Score *(average item score)*	

See user manual page 9

Domain 3 – CBT Interventions

3.1. Appropriate Intervention Targets	
3.2. Choosing Suitable Interventions	
3.3. Rationale for Interventions	
3.4. Implementing Interventions	
3.5. Reviewing Interventions	
Total Domain Score *(average item score)*	

See user manual page 10 – 14

Domain 4 – Homework

4.1. Reviewing Homework	
4.2. Choosing Suitable Homework	
4.3. Rationale for Homework	
4.4. Planning Homework	
Total Domain Score *(average item score)*	

See user manual page 15 – 18

Domain 5 – Measuring Change

5.1. Choosing Suitable Measures	
5.2. Implementing Measures	
Total Domain Score *(average item score)*	

See user manual page 19 – 20

Domain 6 – Effective Use of Time

6.1. Pace	
6.2. Time Management	
6.3. Maintained Focus	
Total Domain Score *(average item score)*	

See user manual page 21 – 23

ACCS Feedback Form

Domain 7 – Fostering Therapeutic Relationship

7.1. Interpersonal Style	
7.2. Empathic Understanding	
7.3. Collaboration	
Total Domain Score *(average item score)*	

See user manual page 24 – 26

Domain 8 – Effective Two-way Communication

8.1. Patient Feedback	
8.2. Reflective Summaries	
Total Domain Score *(average item score)*	

See user manual page 27 – 28

Overall Session Score

Total Score (*add total domain scores*)

Global Performance Rating

Therapist's overall performance within the session was:

1. Limited **2. Basic** **3. Good** **4. Advanced**

Patient Complexity Rating

How complex do you feel the patient was to work with?

1. **Very Straightforward** **2. Somewhat Straightforward** **3. Somewhat Complex** **4. Very Complex**

Appendix 8: Supporting information cover sheet to accompany therapists' session recording

ACCS Submission Cover Sheet

Therapist's Name: _____

Date of Submission: _____

Assessor's Name: _____

Supporting Material. The recording should be accompanied by any materials completed in the session or for homework (e.g. diaries, thought records etc.) and any completed questionnaires.

Formulation. Please attach a copy of the formulation (no more than one A4 sheet). The formulation should demonstrate the therapist's understanding - at that point in time - of how the patient's problems developed (background experiences, beliefs and assumptions, critical incidents) and what is maintaining them (vicious circles).

Session Summary:

- Stage of therapy / session number

- Patient's presenting problem(s)

- Patient's treatment goals

- Agenda

- Homework review (What were the assignments? Outcome?)

- Main topic(s) – What were the main things you worked on? What did you do (e.g. worked through problem on thought record, carried out behavioural experiment, looked for alternative explanations)? What is the rationale for this work?

- New homework (What are the assignments? What is the rationale for them?)

- Patient's feedback on the session (reactions, take home messages)

CHAPTER SIX:

A focus group evaluating the usability and utility of the Assessment of Core CBT Skills (ACCS) scale

Chapter Overview

This thesis presents the development and evaluation of the Assessment of Core CBT Skills (ACCS), a novel self- or assessor-rated scale that can be used to rate therapists' performance within Cognitive Behavioural Therapy (CBT) treatment sessions. Development of the ACCS involved three key stages: (1) scale development, (2) expert review, and (3) scale evaluation. The final phase of this process, scale evaluation, is further broken down into two discrete studies: a pilot study examining the psychometric properties of the ACCS (outlined in Chapter five) and a focus group examining the usability and utility of the ACCS (outlined in this chapter). The focus group was used to obtain in-depth qualitative feedback about the usability of the ACCS from expert assessors (*N* = 9) with experience of using the ACCS scale. Participants highlighted several strengths of the ACCS, including the relevance, clarity and specificity of scale items, the user-friendly design, and the capacity to provide useful formative feedback. A number of areas where the ACCS required further refinement also emerged from the group discussion, prompting minor changes to the content of the ACCS (i.e. revised wording for item descriptors), the provision of further rating guidance on the feedback form, further clarification in the manual rating guidelines, and simplification of the scoring system.

Acknowledgements:

With thanks to all of the participants at the Oxford Cognitive Therapy Centre who gave up their valuable time to take part in this study and without whom this research would not have been possible.

Contents

Introduction 406

Method 407

Participants 407

Data Collection 407

Data Analysis 406

Results 412

Key strengths 412

Limitations and Improvements 413

Revisions to the ACCS feedback form. 415

Improvements to rating guidelines in the ACCS manual 416

Modifications to the ACCS scoring system. 416

Changes to content of the ACCS scale. 418

Discussion 420

Summary of Findings 420

Limitations 421

References 423

Appendices 425

Appendix 1. Annotated Notes from the Assessment of Core CBT Skills (ACCS) Focus Group ... 425

Appendix 2. Initial Thematic Framework from the Assessment of Core CBT Skills (ACCS) Focus Group 430

Appendix 3. Revised Assessment of Core CBT Skills (ACCS) Feedback Form, Manual and Cover Sheet (version 3) 436

Appendix 4. Text relating to Socratic Enquiry 477

Introduction

This thesis focuses on the development of a novel Cognitive Behavioural Therapy (CBT) competence rating scale (Assessment of Core CBT Skills: ACCS). The ACCS provides a framework for assessors to provide formative and summative feedback regarding therapists' performance within observed CBT treatment sessions and for therapists to rate and reflect on their own performance. Development of the ACCS followed a classical measurement theory approach involving three key stages: (1) scale development, (2) expert review, and (3) scale evaluation (DeVellis, 2012). Chapter three outlined the initial process of scale development and Chapter four provided an 'expert' review of the ACCS scale. The final phase of this process, scale evaluation, is further broken down into two discrete studies: a pilot study examining the psychometric properties of the ACCS was outlined in Chapter five, whilst this chapter moves on to outline results from a focus group examining the usability and utility of the ACCS.

The development of a rating scale should be an iterative process that employs a range of qualitative and quantitative techniques to construct and test scale items (Brewer & Hunter, 2005; DeVellis, 2012). Although publications often focus on quantitative evaluation of rating scales, qualitative evaluation provides an important supplement to psychometric testing. In particular, qualitative feedback provides a greater depth of information and can highlight novel and unexpected benefits and limitations of a rating scale. Chapter four provides detailed feedback about the ACCS from a range of novice and expert CBT therapists. The majority of participants found the scale to be relevant, clear and easy to use and provided positive qualitative feedback about the usability, clarity, formative function, and comprehensive nature of the ACCS. Participants also highlighted areas where the ACCS required further improvement and therefore helped to inform a number of refinements to the ACCS. There are, however, two main limitations to this study. First, feedback was provided on the basis of reading the ACCS manual and feedback form, rather than experience of using the scale in practice. Second, the qualitative feedback was limited in that participants provided only short written statements within a questionnaire. Hence this study seeks to build upon the feedback study outlined in Chapter four by obtaining in-depth qualitative feedback from assessors with experience of using the ACCS. Focus groups offer the opportunity for collective discussion and hence provide a deeper understanding of how a group of people feel about an issue (Krueger & Casey, 2009; Vaughn, Schumm, & Sinagub, 1996). A focus group was therefore used to obtain assessors' feedback on the usability of the ACCS scale, with the intention of identifying what worked well in practice, what did not work so well in practice (i.e. barriers and problems encountered) and any areas which warrant further improvement or refinement.

Method

Participants

Participants in the focus group were supervisors who assessed therapists within the 2013/14 intake of the Postgraduate Diploma in Cognitive Behavioural Therapy course (PGDip) run by the Oxford Cognitive Therapy Centre (OCTC) in collaboration with the University of Oxford's Department of Continuing Education (for a detailed description of the course please see McManus, Westbrook, Vazquez-Montes, Fennell, & Kennerley, 2010). These participants were chosen because they all had significant knowledge and experience delivering CBT, had routinely assessed the competence of the trainee CBT therapists for a number of years (using a variety of different methods) and had experience of using the ACCS rating scale during term one and/or two of the PGDip. Eleven assessors were invited to participate, nine of whom (81.82 %) were able to attend. Assessors were all BABCP accredited CBT therapists who had been practicing CBT for between 13 and 30 years (M = 20.22, SD = 6.24) and all of whom had treated over 200 CBT cases. Four assessors were clinical psychologists, three were nurses, one was a psychiatrist, and one was a CBT therapist.

Data Collection

A focus group was used to obtain assessors' feedback on the ACCS scale, rather than a series of individual interviews. It was felt that the role of group dynamics would be beneficial in better understanding assessors' views about the ACCS scale, as group members could respond to ideas and comments of others (Krueger & Casey, 2009; Vaughn et al., 1996). In addition, as participants were already known to each other and were used to discussing issues such as how best to assess the competence of trainees on the course, it was felt that group dynamics (e.g. pressures of group conformity) were unlikely to unduly hinder the discussion (Stewart, Shamdasani, & Rook, 2007). Furthermore, the neutral nature of the topic meant that demand characteristics were likely to be the same in either an individual or group setting (Stewart et al., 2007).

A semi-structured interview schedule outlined in Table one was used to guide the discussion. Within the schedule, emphasis was placed on facilitating discussion of both positive and negative responses to the ACCS. Where negative responses arose, 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,

with participants being encouraged to explore issues as they arose within the discussion. Discussions were facilitated by the primary researcher (Kate Muse). Although this meant that the facilitator was not independent, it was decided that this would be the most appropriate choice given the researcher's in-depth knowledge and understanding of the ACCS. The focus group lasted for 55 minutes and 33 second. The group discussion was audio recorded. This recording was used as a basis for subsequent analysis. However, the discussion was not transcribed verbatim. This is because the study was completed as part of a routine course evaluation meeting. Thus, although participants gave consent for their views and opinions to be included within the project, they did not provide consent for direct quotations to be used for research purposes.

Data Analysis

Qualitative analysis of the focus group comprised of the 'framework technique' (Ritchie & Spencer, 2002). This simple approach was chosen as the intention of the analysis was not to provide a detailed exploration of personal experience (as is the case in approaches such as interpretative phenomenological analysis) or to generate a theory of a phenomena grounded in the data (as is the case in approaches such as grounded theory), but to offer a simple framework which described the key advantages, disadvantages and potential revisions to the ACCS which were highlighted by participants in the focus group (Braun & Clarke, 2006).

Analysis involved four key stages (Ritchie & Spencer, 2002). First, emergent themes from the interview and a priori themes from the interview schedule were used to identify an initial thematic framework. Second, the thematic framework was systematically applied to the data. As discussed previously, a transcript of the focus group was not used in the current study. Instead, the lead author listened to the recording and took detailed notes about all aspects of the content of the recording. These notes were then annotated according to the initial thematic framework (see Appendix one). Third, the data was rearranged within a chart according to themes (i.e. the content of the recording notes was distilled into a summary and entered into a chart of key themes- see Appendix two). Fourth, a final 'map' of key themes was created by aggregating patterns of data, weighing up the importance and dynamics of issues, searching for an overall structure in the data and synthesising the findings (see Figures one and two). The primary researcher (Kate Muse) took the lead in the analysis. However, Dr Sarah Rakovshik and Freda McManus met regularly with Kate Muse to discuss emerging findings and refine the emerging framework. Several methods were used to reduce the impact of researcher bias, including awareness of preconceptions; sharing preconceptions between researchers and ensuring that all interpretations were supported by participant-derived data.

Table one:

Summary of interview schedule

Open ended questions	**Example Prompts**
What has been your experience of using the ACCS? How have you found it?	Probing questions: • 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? Comparative questions: • Did anybody else in the group experience this? • Did anybody else in the group have a different experience?
What did you most like about the ACCS scale?	Probing questions: • 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? Comparative questions: • 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 ACCS scale?
What did you least like about the ACCS scale?	Probing questions: • 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*? • Do you have any examples of that which you could tell me about?

	Development questions: • 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? Comparative questions: • 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 ACCS scale?
How easy or difficult was it to use the ACCS?	Probing questions: • 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 ACCS? • Do you have any examples of that which you could tell me about? Development questions: • What would make the ACCS 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? Comparative questions: • 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 ACCS?
How clear did you find the items in the scale? *More specific probing questions[1]:* Were any items unclear or confusing? Did you feel the formulation item was clear?	Probing questions: • Could you tell me what you mean by that? • Could you say something more about that? • Why was it that *xxx* was unclear or confusing? • Do you have any examples of that which you could tell me about?

Did you feel the measuring change item was clear?	<u>Development questions:</u> • 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? <u>Comparative questions:</u> • 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?
How appropriate did you find the scoring system? *More specific probing questions:* Did you like the generic 4-point scale, with item-specific anchor descriptions? How useful was the 'global' rating you provided for therapists at the end of the scale? How useful was the 'overall session score' at the end of the scale? Would you rather have seen a mean item score? Did you use ½ marks? Do you think this is necessary/important? Why?	<u>Probing questions:</u> • 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? <u>Development questions:</u> • 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? <u>Comparative questions:</u> • 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?
How relevant did you think the ACCS was to the assessment of CBT competence? *More specific probing questions:*	<u>Probing questions:</u> • Could you tell me what you mean by that? • Could you say something more about that? • Why was it that *xxx* was not relevant?

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? Did you feel the formulation item was relevant? Did you feel the measuring change item was relevant?	• Do you have any examples of that which you could tell me about? <u>Development questions:</u> • 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? <u>Comparative questions:</u> • 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?
Did you feel that there was anything missing from the scale? *More specific probing questions:* Are there any key competences which the scale neglects? Would you liked to have seen reference to the use of protocols at all? Would you liked to have seen reference to use of emotional expression in the scale?	<u>Probing questions:</u> • 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? <u>Development questions:</u> • Could you tell me more about how you would like to see *xxx* incorporated within the scale? <u>Comparative questions:</u> • 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?
What did you think about the overall style, appearance and layout of the scale?	<u>Probing questions:</u> • 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? <u>Development questions:</u> • 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? <u>Comparative questions:</u> • 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?

Note- At the end of the interview, all experts were also asked whether there was "anything we haven't spoken about which you feel would be relevant or important?".

[1] More specific probing questions were based on initial feedback and questions which arose to the researcher during the scale development process. In order to avoid unduly influencing the group discussion, these questions were reserved until participants had fully discussed their answers to the broader, open-ended questions.

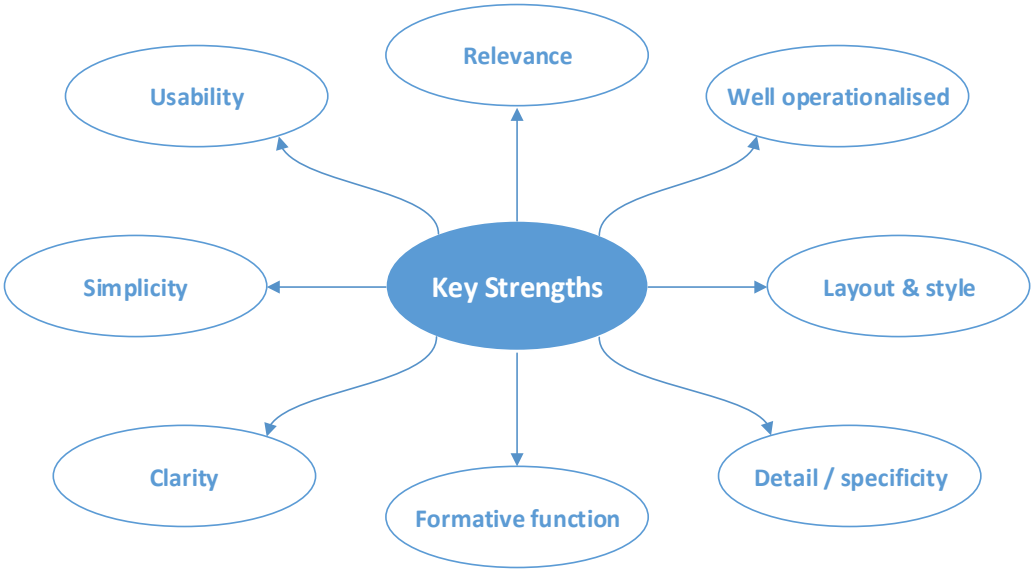
Results

Two key themes emerged from the analysis of the focus group: key strengths of the ACCS and limitations and suggestions for improvements to the ACCS. These are presented below.[1]

Key strengths

Eight key strengths were outlined by participants in the focus group (see Figure one). Participants felt the ACCS was very relevant to CBT competence, had a clear instruction manual (in terms of both general and item-specific guidance), was attractive in appearance (particularly in terms of layout, style and use of colour coding), and clearly and simply defined and operationalised the competences included in the scale. They also liked the detail and specificity of the ACCS, i.e. the way the ACCS expanded upon the core competences required to deliver CBT and broke down these competences into smaller components. This was particularly highlighted as being a strength within the homework, agenda, interventions, and communication domains. All of these strengths contributed to participants feeling that the ACCS was easy to use. Participants also commented that the increased detail and specificity of the ACCS provided them with a useful template for providing more in-depth feedback to therapists and meant that the ACCS served as a useful good practice guide for therapists, helping them to understand and remember the skills necessary to deliver CBT. Interestingly, participants did not feel that this increase specificity meant that the ACCS was too lengthy or took too long to complete, especially once they had become familiar with the scale and had opportunity to practice using it.

Figure 1: Key strengths of the Assessment of Core CBT Skills (ACCS) identified in the focus group.

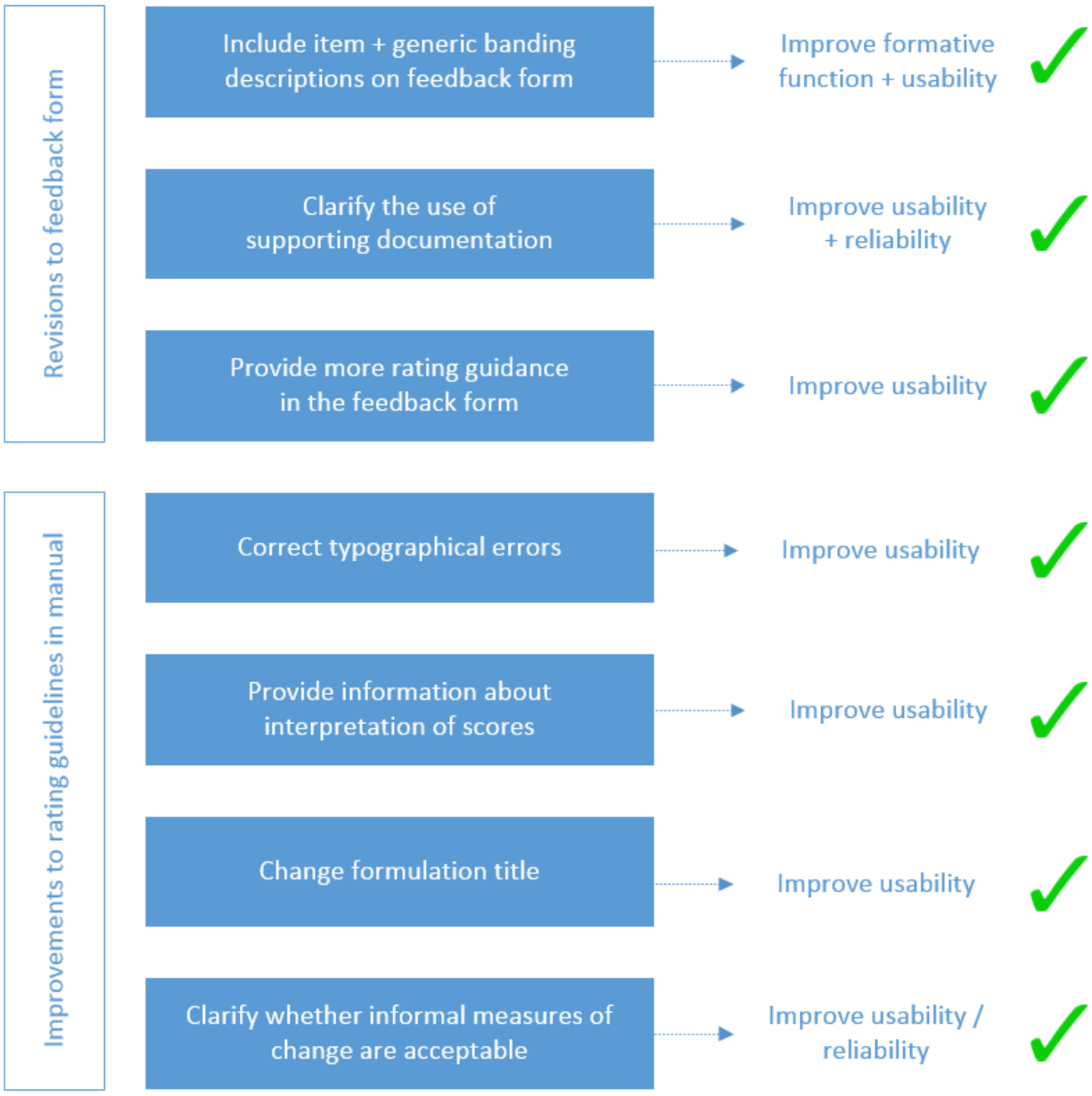


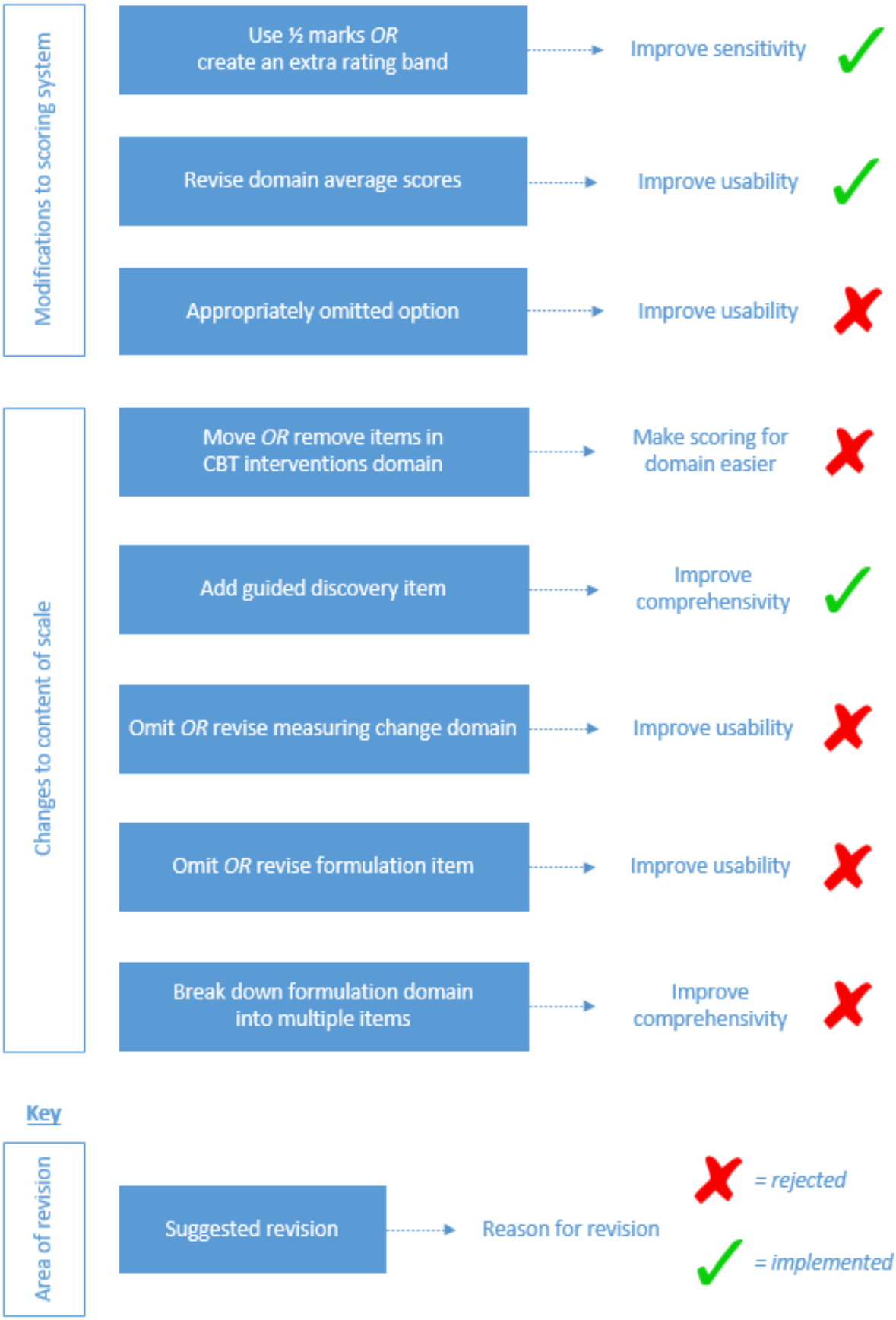
[1] Direct participant quotations are not provided to support the key themes identified in the text. This is because participants did not provide consent for direct quotations to be used for research purposes.

Limitations and Improvements

A number of limitations and potential methods for overcoming these difficulties emerged within the focus group. These are organised in relation to four main areas. An overview of these areas of revision is provided in Figure two and further detail about the reasons for the suggestions and the specific revisions which were made to the ACCS are provided below. Revised versions of the ACCS manual, feedback forms and submission cover sheet can be found in Appendices three to five.

Figure 2: Limitations and potential improvements to the Assessment of Core CBT Skills (ACCS) identified in the focus group.




Modifications to scoring system
Use ½ marks OR create an extra rating band
Improve sensitivity
Revise domain average scores
Improve usability
Appropriately omitted option
Improve usability
Changes to content of scale
Move OR remove items in CBT interventions domain
Make scoring for domain easier
Add guided discovery item
Improve comprehensivity
Omit OR revise measuring change domain
Improve usability
Omit OR revise formulation item
Improve usability
Break down formulation domain into multiple items
Improve comprehensivity
Key
Area of revision
Suggested revision
Reason for revision
= rejected
= implemented

Revisions to the ACCS feedback form.

A number of areas of confusion and misunderstandings arose within the focus group. Upon discussion about why these arose, a number of assessors revealed that they felt they had only had time to briefly read the ACCS manual and then quickly completed ACCS ratings without referring back to the manual. Consequently, assessors drew their own conclusions about what the items included (based on the item headings) and completing ratings on the basis of the banding headings alone (i.e. 1- limited, 2- basic, 3- good, 4- advanced), rather than the overarching or item-specific definitions provided in the manual. This problem was further exacerbated by the fact that assessors only provided numeric feedback and thus did not have to provide in-depth qualitative feedback to support and justify their ratings. Additionally, most participants had spent many years using the CTS and/or CTS-R, which lead to them becoming confused between these scales and the ACCS. In line with participants' suggestions, overarching item descriptors and generic banding descriptors were added to the feedback form so assessors would not need to refer to the manual so frequently. It was also felt that this would improve therapists' understanding of the feedback they received. Further guidance about how to use the scale was also added to the beginning of the feedback form in case assessors did not use the manual. However, guidance was also added advising assessors to spend some time reading and familiarising themselves with the manual before conducting an assessment and to refer to the manual for guidance about the type of performance which is consistent with each point on the rating scale for each item (see Appendix three). Another suggestion for overcoming this difficulty was to provide assessors with training in how to use the ACCS scale, as was used for the CTS-R (Reichelt, James, & Blackburn, 2003). The need for assessor training is, however, beyond the scope of the current study and is instead discussed further in Chapter seven.

There was some uncertainty amongst participants about whether and when to use supporting documentation in addition to viewing session recordings. This was particularly highlighted for the formulation and measuring change domains (i.e. participants weren't sure whether they should be using written formulations and information about questionnaire/idiosyncratic measures). Consequently, some used this additional material, whilst others relied on the recording alone. To overcome this problem clarification was added (1) to the brief introductory guidelines in the feedback form and (2) to the feedback form for the formulation and measuring change domains (this information was already in the manual) (see Appendix three). Additionally, the need to include information about idiosyncratic and standardised questionnaires was highlighted in the ACCS submission cover sheet to ensure that therapists submitted this material with their recording (see Appendix three).

Improvements to rating guidelines in the ACCS manual.

Three improvements were made to the rating guidelines in the ACCS manual. First, some participants were unsure whether it was acceptable for therapists to use informal measures of change, believing that the measuring change domain referred only to formal, standardised questionnaire measures. This is important as informal measures (e.g. simple visual analogue scales, frequency counts) are often sufficient for assessing change. To overcome this difficulty further clarification was provided in the manual, the description of the item was updated, and the heading of the domain was changed to "assessing change" (see Appendix three). Second, the 'coherent formulation' item refers to whether the formulation was actively used and updated. However, it was felt that this was not clear in the title of the item. Two alternative titles were proposed: "coherent and dynamic formulation" and "in session formulating", with the former being used. Finally, two small typographical errors were corrected.

Modifications to the ACCS scoring system.

The number of points on the ACCS scale was debated within the focus group. Some participants felt there was too much of a jump from basic to good and thus wanted more than four scale points. Some also commented that a larger range of scale points would improve the sensitivity of the scale. The first version of the ACCS did use a five-point scale. However, this was reduced to a four-point scale because the majority of participants in the feedback study (see Chapter four) felt that there was a lack of discrimination between the anchor descriptions, especially at the upper levels. Furthermore, many participants in the focus group felt the four-point scale was suitable, if they were able to provide ½ marks to those who fell between two of the descriptors (i.e. 1.5, 2.5, or 3.5), effectively increasing the scale to a seven-point scale. Thus, it was decided that the four point response scale would be retained. However, guidance was provided within the manual and feedback form, advising assessors that they could provide ½ marks to those who fell between two of the descriptors (see Appendix three).

A number of participants found the scoring system confusing, particularly the use of mean scores for each domain. Indeed, many therapists who completed self-ratings (and some assessors) provided total scores for each domain rather than mean scores. Although a button was created in the feedback form which automatically calculated the mean domain score, few participants were able to make the button work and others preferred to work with pen and paper in any case. Participants suggested that this could be overcome by removing the mean domain score, as the small number of items in each domain makes it relatively easy to identify therapists' performance within the domain without a mean or total domain score. Although version one of

the ACCS included total domain scores, these were changed to mean domain scores in response to feedback that this scoring system lead to uneven item weighting in the final total score (due to the different number of items in the domains: see Chapter four). Removing the mean domain score would re-introduce this problem, as each item would be provided with an equal weighting in the total score (meaning that the domains with more items would be more strongly represented). However, as it is not currently known which aspects of CBT skill outlined in the ACCS are more important or more predictive of patient outcome than others, it is not known whether providing each domain with the same weighting in the total score (by using mean domain scores) provides a more accurate reflection of the relative importance of the items in the total score. Thus, neither approach is without its problems. Given the difficulties assessors' faced in using mean scores and the current lack of knowledge about the relative importance of specific items and domains in the ACCS, the total/mean domain scores were removed, thus providing each item with equal priority in the total score. Rating guidelines were updated in line with this new scoring system (see Appendix three).

Some assessors found it difficult to interpret the total ACCS score (i.e. they were not sure whether the total score assigned indicated an acceptable level of competence). It was suggested that further guidance should be provided in the manual to aid interpretation. A minimum score above which a therapist can be considered competent was not provided in previous versions of the ACCS manual and was not added in this revision. This is because further research is required to establish an empirically proven 'tipping point' for competence using the ACCS (although a cut-off score is provided for the CTS-R this is somewhat arbitrary as it was derived from the value that was one standard deviation below the mean score of a group of certified CBT therapists within a research trial [Shaw, 1984; Shaw et al., 1999]). Furthermore, the level of competence and thus the score on the ACCS which is required may vary according to the context and purpose of the assessment (e.g. a lower level of performance may be acceptable for determining suitability for inclusion on a CBT training course, whereas those seeking accreditation may be required to demonstrate a higher level of performance). However, to provide those using the scale with a framework to help them interpret and understand the total score, information was provided within the manual about the total score which would be achieved if a therapists performance was consistently rated as 1-limited, 2-basic, 3-good, or 4-advanced (see Appendix three). Additionally, a mean item score option was provided in the feedback form to indicate the average performance of the therapist (see Appendix three).

The possibility of including an 'appropriately omitted' option was discussed. Some participants felt this would be helpful for the measuring change items (if no idiosyncratic or

standardised questionnaires were used / reviewed) and the formulation item (if the formulation wasn't explicitly referred to in-session). However, others felt it should always be possible to rate these items, providing the supporting documentation was referred to. Given that further guidance was provided to highlight the need to use supporting documentation to rate these items, and that the appropriately omitted option was not thought to be helpful for any other scale items, it was decided not to include this rating option.

Changes to content of the ACCS scale.

Some participants wondered whether the number of items in the CBT interventions domain caused confusion when calculating the mean domain score. Although the removal of some of these items was raised as a possibility, none were identified as warranting removal. Another possibility for reducing the length of this domain was to move the 'choosing suitable intervention targets' item to the formulation domain. However, this suggestion was not followed because, although the item refers to whether the selection of intervention targets was guided by the patient's idiosyncratic formulation, it also refers to whether the targets were skilfully defined, clarified, and specified. Additionally, the mean domain score was removed in response to other feedback from the group. Participants also commented that the formulation domain could be further broken down into a number of discrete constituent parts (e.g. is it coherent, suitable for stage of treatment, shared with the patient, revised in light of new information etc.). However, the formulation item was not broken down into multiple items as, upon discussion, it became clear that these additional components would often not be evident with a session.

There was some debate amongst participants about the inclusion of the formulation domain within the ACCS. Some participants questioned whether the formulation is, or indeed should be evident in every session, and thus whether it is appropriate to include this domain. In contrast, many participants felt that the formulation should be evident in each session and that viewing the written formulation alongside the session recording allows assessors to determine whether the approach taken in the session was in line with the broader written formulation (e.g. does a simple maintenance cycle discussed in-session fit in with the overarching formulation?). In line with this view, it has been argued within the CBT literature that a formulation should evolve as further clinical information arises and recurrent patterns and themes emerge and should also be regularly referred to in order to facilitate patients' understanding of their problems within a CBT framework and to highlight potential treatment strategies (Butler, 1998; Kuyken, Padesky, & Dudley, 2011). The view was therefore taken that a therapist's ability to conceptualise a patient's

problems should be evident throughout the course of treatment and the formulation domain was therefore retained.

A similar discussion was held about whether it was always appropriate to measure change in symptoms, associated features (e.g. beliefs, behaviours, feelings), and movement towards goals and thus the possibility of altering or removing the measuring change domain was considered. Some participant felt that measuring change is an important aspect of delivering CBT and one that should be evident in every session, although it was recognised that at times this issue is somewhat neglected by supervisors and therapists alike. Conversely, other participants felt it was not necessary to measure change every session. To some extent the debate seemed to reflect a somewhat controversial shift in the field towards more formalised and regular monitoring of progress (e.g. Clark, 2011). However, two areas of confusion also seemed to underlie the discussion: 1) whether supporting documentation could be used, and 2) whether this domain referred only to standardised, formal measures. Most participants felt that the domain could be rated for every session if supporting documentation was used and if informal, idiosyncratic measures (e.g. a traffic light system or thermometer scale) were considered appropriate. As the manual and feedback form were adapted to address both of these issues (see discussion above), these domains were retained.

A number of participants felt that Socratic enquiry was not evident in the ACCS. Socratic enquiry is a teaching method that uses a progressive series of questions designed to facilitate independent thinking and promote active learning (Kennerley, 2007; Overholser, 2010). Socratic enquiry can be used to (1) explore and unpack individuals experiences and identify idiosyncratic "meanings", (2) help patients widen their cognitive and behavioural horizons in order to shift unhelpful affect, (3) help patients develop independence in learning and retaining relevant CBT skills and knowledge, and (4) promote a climate of shared responsibility and expertise (Beck, 1995; Kennerley, 2007; Overholser, 2010; Padesky, 1993; Roth & Pilling, 2007; Tee & Kazantzis, 2011). Thus, the use of Socratic enquiry should be evident throughout all aspects of a CBT session and should pervade the use of other vehicles for promoting a change of perspective (e.g. debriefing of behavioural experiments, review of thought records etc.). Given the pervasive and recurrent nature of these aspects of competence, they were initially embedded within the rating framework of other specific competences (e.g. conceptualisation, agenda setting, delivering CBT interventions etc.) and these themes were further accentuated after six experts who participated in the feedback study identified that the first version of the ACCS scale did not feature guided discovery and Socratic enquiry prominently enough (see Chapter four). However, it was clear from discussion with participants in the focus group that Socratic enquiry was still not given

enough prominence in version two of the ACCS. Two potential solutions were identified. First, some participants suggested adding an extra item within the ACCS. Second, participants recognised that Socratic enquiry was evident to some degree in the collaboration, reviewing homework, and reviewing interventions items and some felt that it could be drawn out further within these items. Following consultation and guidance from Dr Helen Kennerley from the Oxford Cognitive Therapy Centre, an expert in Socratic methods (Kennerley, 2007; Westbrook, Kennerley, & Kirk, 2007), it was decided that it would not be appropriate to add an extra item to the scale because this item would overlap considerably with other items in the scale. Instead, the items which included reference to Socratic enquiry were revised to ensure that the references to this method of questioning were more explicit. The relevant text in items 3.5- reviewing interventions, 4.1- reviewing homework, and 7.3- collaboration is highlighted in yellow in Appendix four).

Discussion

Summary of Findings

The present study sought to obtain feedback from assessors with experience of using the Assessment of Core CBT Skills (ACCS) scale in order to examine assessors' views about the usability and utility of the ACCS. A number of strengths of the ACCS scale were outlined by participants in the focus group. In particular participants felt that the items in the scale were relevant, well operationalised, detailed and specific. Participants also commented that the ACCS had an attractive layout and style, performed a useful formative function, and was clear, simple and easy to use. A number of limitations and problems were also highlighted by participants, which is to be expected given that this was the first time the ACCS scale had actually been used in practice to rate therapists' competence. A number of resolutions to these problems were also identified by participants, many of which were implemented. Key revisions to the ACCS fell within four areas: (1) the provision of further rating guidance on the feedback form (e.g. the addition of item and generic banding descriptions and clarification about the use of supporting documentation), (2) improvements to the manual rating guidelines (e.g. clarification regarding the use of informal measures of change, amending the formulation item title, and providing further information about how to interpret ACCS scores), (3) modifications to the scoring system (e.g. providing guidance about the use of ½ marks, removing total domain scores, and adding an mean item score), and (4) making minor revisions to the wording of the item descriptors.

A number of confusions arose in the focus group because assessors only had time to briefly read the ACCS manual and then quickly completed therapist's ACCS ratings without

referring back to the manual. This problem may be specific to participants in this study (e.g. ACCS ratings were completed in addition to CTS-R ratings therefore increasing demands on already busy schedules and participants only completed numeric feedback and thus did not have to justify their ratings). However, the findings from this study may be representative in that busy assessors within research, training and routine practice settings alike may find it difficult to find the time or motivation to read the ACCS manual. One solution to this problem may be to 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 in this study and the feedback study (see Chapter four) consistently asked for detailed item-specific rating guidelines. Another solution highlighted by participants in this study was to add further guidance to the feedback form. This advice was followed to some degree, with further generic and item-specific guidance being added. However, adding all of the detail included in the manual would have defeated the purpose of the exercise. An alternative solution, and one that participants in this study widely endorsed, is to offer assessors training in how to use the ACCS. Although attending training is also time-consuming, there may be less barriers to attending training than there are to finding time in a busy schedule to read a manual. Thus, a useful avenue for further exploration may be the development of training in how to use the ACCS. It may also be worth considering the development of an online scoring system that presents assessors with relevant guidelines from the manual for each item that assessors are required to view before they enter their numerical score.

Limitations

It is important to note that all three individuals involved in conducting the analysis (Kate Muse, Freda McManus, and Sarah Rakovshik) were also involved in the development of the ACCS scale, which may have influenced their interpretation of the data. This potential limitation could be overcome by carrying out independent validation of the analytic process. Specifically, an individual who had no involvement in the development of the ACCS could review the paperwork trail that outlines the process of analysis, including the original notes from the focus group, initial annotation of the notes, the initial thematic framework, and the final thematic framework. This would ensure that the analysis was systematically achieved and that the themes which emerged were supported by the data.

Perhaps the most significant limitation to the current study is the sample of expert therapists who participated in the study. The only experts who were invited to participate were those from the Oxford Cognitive Therapy Centre who had completed assessments using the ACCS for students enrolled in a Post Graduate Diploma in CBT. These participants were selected because

they had significant knowledge and experience delivering CBT and of assessing the competence of the trainee CBT therapists (using a variety of different methods, including the ACCS). However, this recruitment process may have resulted in a somewhat biased and relatively homogenous sample. It is also possible that demand characteristics influenced the feedback given in the group, especially as the primary author of the ACCS lead the group. In addition, the sample size is relatively small (*n* = 9). Although this is an appropriate sample size for conducting in-depth focus groups (Stewart et al., 2007), it does further limit the generalisability of the results. That being said, it was not the intention of this study to provide insight into the way in which the ACCS scale is viewed by assessors or therapists more broadly, but to provide in-depth insight into the opinions of a small group of experts in order to identify any aspects of the ACCS scale which required further improvement or refinement prior to more widespread use of the scale.

References

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. New York: Guildford Publications.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77 - 101. doi: 10.1191/1478088706qp063oa

Brewer, J., & Hunter, A. (2005). *Foundations of multimethod research: Synthesizing styles*. Thousand Oaks, CA: Sage Publications.

Butler, G. (1998). Clinical formulation. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology* (Vol. 6). Oxford: Elsevier Science Pergamon Press.

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23*(4), 318-327. doi: 10.3109/09540261.2011.606803

DeVellis, R. F. (2012). *Scale development: Theory and Applications*. Thousand Oask, CA: Sage.

Kennerley, H. (2007). *Socratic questioning: A guide for therapists*. Oxford: Oxford Cognitive Therapy Centre.

Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research* (4th ed.). London: Sage Publications.

Kuyken, W., Padesky, C. A., & Dudley, R. (2011). *Collaborative case conceptualisation: Working effectively with clients in cognitive-behavioral therapy*. New York: Guildford Press.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. London: Sage Publications.

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. doi: 10.1016/j.brat.2010.08.002

Overholser, J. C. (2010). Psychotherapy according to the socratic method: Integrating ancient philosophy with contemporary cognitive therapy. *Journal of Cognitive Psychotherapy, 24*(4), 354-363. doi: 10.1891/0889-8391.24.4.354

Padesky, C. A. (1993). *Socratic questioning: Changing minds of guided discovery?* Paper presented at the European Congress of Behavioural and Cognitive Therapies, London.

Reichelt, F., 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. doi: 10.1016/S0005-7916(03)00022-3. ISSN: 0005-7916.

Ritchie, J., & Spencer, L. (2002). Qualitative data analysis for applied policy research. In A. M. Huberman & M. B. Miles (Eds.), *The qualitative researcher's companion* (pp. 305 - 329). London: Sage Publications.

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognititve and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

Shaw, B. F. (1984). Specification of the training and evaluation of cognitive therapy for outcome studies. In J. B. W. Williams & R. L. Spitzer (Eds.), *Psychotherapy research: Where we are and where we should go*. New York: Guildford.

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), 837-846. doi: 10.1037/0022-006x.67.6.837

Stewart, D. W., Shamdasani, P. N., & Rook, D. W. (2007). *Focus groups: Theory and practice* (2nd ed.). London: Sage Publications.

Tee, J., & Kazantzis, N. (2011). Collaborative empiricism in cognitive therapy: A definition and theory for the relationship construct. *Clinical Psychology: Science and Practice, 18*(1), 47-61. doi: 10.1111/j.1468-2850.2010.01234.x

Vaughn, S., Schumm, J. S., & Sinagub, J. M. (1996). *Focus group interviews in education and psychology*. London: Sage Publications.

Westbrook, D., Kennerley, H., & Kirk, J. (2007). *An introduction to Cognitive Behaviour Therapy: Skills and Applications*. London: Sage Publications Ltd.

Appendices

Appendix 1. Annotated Notes from the Assessment of Core CBT Skills (ACCS) Focus Group

The following notes are taken directly from the recording of the ACCS focus group. General headings are provided simply to offer a structure to the notes and improve readability and later interpretation. They are not intended to provide a thematic structure or interpretation of the data. The order of the notes reflects the order in which the issues/topics were raised within the focus group. However, where issues were revisited at a later stage, the secondary comments from assessors were included with the original heading. These notes have been annotated, using comment boxes, according to the initial thematic framework identified.

- **General comments about ACCS:**

Assessors generally really liked the scale overall.

kate muse
General positive comment- fits into strengths??

Assessors thought it was well operationalized.

Assessors felt the scale was very relevant to CBT competence.

Assessors found it simple to use.

Assessors felt the manual was very clear.

Some assessors liked the patient complexity rating.

kate muse
Some assessors liked the patient complexity rating.

- **Generic intro / definition of performance bandings:**

Assessors liked the generic introduction and found the generic performance bandings a useful "anchor" and default rating.

kate muse
Strengths of the ACCS

- **Formulation domain:**

Raised the possibility of putting "Coherent and dynamic formulation" on the actual rating form, someone who is not using the manual might settle for coherent and not bother to judge if the formulation is being "actively" used and updated.

kate muse
Active formulation

Another issue was raised - is this actually 2 items (i.e. coherent and regularly updated/revised in light of new information). Others also thought the formulation could be broken down into multiple components (is it coherent, suitable for stage of treatment, is it shared with the patient). However, assessors noted that this is often not evident in a session, making it harder to include broader issues relating to the formulation.

kate muse
Break down formulation domain

Some assessors felt the formulation wasn't always (or often) explicitly referred to in the session, especially for later sessions, so was hard to rate. Assessors didn't always realise they should also be referring to the written formulation submitted in the supporting documentation. One possibility for overcoming this was to add guidance to the feedback form noting this.

kate muse
Can formulation be rated?

kate muse
Use of supporting materials

Some assessors felt that the formulation should be clearly evident in each session- when assessors view the formulation in combination with the recording they can get a feel about whether what is happening in the session matches what is written in the formulation. I.e. therapists need to be working on the ideas they have identified in the formulation and this should be clear in every session. This might, for example, only be a very simple maintenance cycle which is discussed in the session but this should reflect/fit in with the more detailed formulation.

kate muse
Can formulation be rated?

It was suggested that this item could be called "in session formulating" instead.

kate muse
Active formulation

- **Measuring change domain:**

It was noted that sometimes a simple VAS and frequency count is sufficient. To make this clear, it was suggested that the manual could read "standardised measure(s) and/or idiosyncratic ……".

kate muse
Measuring change doesn't have to include formal/standardised measures

It was also noted that it is not always relevant to carry out a comprehensive review of measure(s). Sometimes, later in treatment, it's sufficient for the patient to give a simple self-rating of progress and both parties know what this indicates. Then therapy mileage is in moving on to relapse management and consolidating progress rather than reviewing rating itself. Would a therapist be penalised for not being more detailed?

kate muse
Should measures be thoroughly reviewed?

Some assessors felt this was a problematic domain because sometimes measures aren't used (justifiably), which then impacts on the overall score (i.e. bring the total score down).

kate muse
Should measures be used?

Some assessors felt it was appropriate not to use measures in every session.

kate muse

Some assessors commented that a problem here might be that therapists and supervisors alike don't focus enough on measuring change in general and that this should actually be something which is evident in each session.

Some assessors felt there should be a culture shift in that measures should be viewed as being an important component of delivering CBT- some commented that this shift was already beginning across a number of settings.

Some assessors used the information about the measures handed in with the supporting documentation to help make this rating, others did not. It was suggested that, if this supporting documentation can be used, it is made more explicit in the manual and in the feedback form.

Some assessors felt that, if supporting documentation was used, then it should always be possible to rate the measuring change item. So, appropriately omitted would not be needed.

- **Scoring system:**

Some assessors, and a number of therapists (when completing self-ratings), included total scores for each domain, rather than average scores. Additionally, the final score was difficult for some to add up due to the use of average total scores. It was also felt that not having average scores would be less work for assessors.

Assessors questioned with the average scores how many decimal points you would go to.

Some assessors liked having a total score for each domain so that you could get a feel for how therapists had performed within each domain. (Helpful for therapists and assessors). Others felt this wasn't necessary as you can get a feel of performance within each domain without total scores as there aren't many items in each domain.

Very few assessors were able to get the button on the word document which provides total domain scores etc. to work.

Some assessors found it hard to decide between basic and good, or felt it was too much of a big jump from basic to good, and so were looking for something in between. Similarly, some assessors found they wanted more than four points on the scale to indicate finer differences in competence.

Some assessors felt the ACCS picked up well on people who were very skilful and those who were not skilful but perhaps missed the subtlety of those "in the middle".

It was noted that a lot of therapists were rated as "3". Some assessors worried that the ACCS wouldn't be able to monitor change in skills.

Some felt the CTSR was better able to capture the range of skills demonstrated by trainees as it provides more points on the scale.

kate muse
Use of supporting materials

kate muse Use of supporting materials

kate muse Appropriately omitted option

kate muse
Confusing scoring system

kate muse
Enough sensitivity?

Some assessors found they wanted to put half marks if it felt as though the supervisee fell between two points on the scale –some assumed this was ok to do, others thought they weren't supposed to use ½ marks. It was felt that this wasn't clear in the manual. It was also felt that ½ marks were necessary in order to accurately represent the full range of abilities.

Some assessors struggled with the interpretation of the ACCS (what do the scores really mean when combined)? Where does competence lie? It was suggested that further guidance could be provided within the manual.

- **Using the manual?**

Some assessors commented that they worried they were being a bit "slapdash" when using the ACCS- doing too quickly and just "plumping" for a number as they didn't have to support this with formative/qualitative feedback/comments.

Possibility of putting generic guidance/definition of performance bandings (in summarised form) on the actual feedback rating form – as many people said they didn't look back at the manual again.

Many assessors also said they didn't know what the generic performance bandings were (e.g. they knew "2" indicated "basic" skill but not that this referred to "Therapist's performance is somewhat appropriate with some degree of skill evident. However, major substantive problems are evident.".

Some assessors highlighted issues which were in the ACCS as missing (e.g. using formulation to guide interventions- which appears in choosing suitable interventions, keeping to the agenda- which appears in maintained focus). Upon discussion it was evident that they had only briefly read the manual and the used the feedback form without referring back to the manual. Consequently they didn't know what all of the items referred to and made their own inferences based on the item titles alone.

kate muse
Interpretation of ACCS scores

kate muse
The manual isn't used

kate muse
Confusion and misunderstandings.

kate muse
Confusions and misunderstandings

kate muse
The manual isn't used

Some assessors wondered whether they were completing detailed enough ratings. They commented that they read through the manual once, thought they understood it so didn't look again. In retrospect they realised that this was, perhaps, not always the case and it would have been better to have looked at the manual again and/or to have had training in how to use the scale.

- **Layout:**

Assessors liked the layout, style and visual appearance of the ACCS. They liked that it was colour keyed.

kate muse
Strengths of the ACCS

- **Familiarity / time**

Many commented that once they got the hang of it, and became familiar with the manual, it only took about 20 mins to use.

kate muse
More detail/specificity

- **Comparison with other scales (CTS-R)**

Assessors noted that they needed a "mind-set shift" from CTS-R in order to use the ACCS.

kate muse
Comparison with CTS-R

Assessors were confused about where an "average" therapist would lie. It was felt that this confusion was to a large degree fuelled by a comparison to the CTS-R ratings- e.g. where would a 2 on the ACCS fall on the CTS-R scale / is a 4 on the ACCS equivalent to a 6 on the CTS-R etc.

Assessors felt frustrated because they had to look at the manual, whereas because they have been using the CTS-R for a long time, they already know how it works and so don't need to look at the manual.

Assessors frequently became confused about what the items in the ACCS were due to their familiarity with the CTS-R (e.g. thinking that the ACCS includes an "interpersonal effectiveness" item which includes empathy, genuineness, warmth and the ability to develop a good relationship, as outlined in the CTS-R).

- **More detailed ratings / specificity**

Some assessors felt they were providing more detailed assessments of skill using the ACCS compared to the CTS-R.

kate muse
More detail / specificity

kate muse
Strengths of the ACCS

Some assessors liked the specificity of the ACCS.

Some assessors also liked the way the ACCS broke down / expanded upon aspects of competence. E.g. the way homework was split into reviewing, rationale, planning homework etc. E.g. the way the agenda is broken down into different aspects. E.g. feedback being split into multiple domains/expended upon. E.g. splitting up interventions into multiple components.

It was felt that specificity/detail/breaking down items helped assessors to provide more in-depth feedback and was helpful as a learning tool for students (e.g. helping them to remember to do all of the aspects of competence outlined/ understand what they were meant to be doing).

The scale was longer/more detailed than the CTS-R. People felt this length was about right.

- **Guided discovery**

Some assessors felt the ACCS was missing guided discovery / eliciting questions / Socratic enquiry. Guided discovery was referred to by assessors as therapists 'CBT stance' or the 'essence of CBT'. It was felt this was something more than collaboration, reflective summaries, and feedback. Something about the quality of the interaction.

kate muse
Missing item: guided discovery

Guided discovery was felt to be a process, where you are moving your patients mind to notice things which are out of view or putting together a maintenance cycle so that they discovery something new. Therapists facilitate the discovery through Socratic questioning (not always but often). This process of discovery is vital. The process of reaching an alternative perspective/more balanced point of view. It is something that is created within the patient, not given to the patient by the therapist but created in the process. Facilitating and enabling change.

Some assessors noted that guided discovery is hard to quantify.

Assessors commented that a lot of the qualitative feedback they provide when assessing tapes is about guided discovery because it is a very important skill that is learnt through supervision.

It was identified that this was a slightly "woolly" concept. Suggestions as to what this item would include were "what did you see that guided the person to discover new things" so it guides the rater to put something explicit / helps the student because students find it hard to understand the concept. The opposite of guided discovery was suggested to be just "telling somebody to do something" or "giving answers". It is something accurate the therapist does. Socratic enquiry was suggested as being part of guided discovery. Important is the specific questions and language being used (e.g. instead of "did you feel sad when that happened" reframing to "how did you feel when that happened"). Moving from closed to open questions.

kate muse
Missing item guided discovery

It was noted that there was something in the collaboration item which overlapped with guided discovery but it was felt that they were distinct.

It was felt it was demonstrated to some degree in collaboration but was more- you can collaborate but guided discovery was felt to be a very different skill.

Some commented that the guided discovery item in the CTS-R confounds with other items.

Some assessor felt guided discovery should in some ways be in every item- it was questioned whether it should perhaps be mentioned within each item?

Some assessors felt guided discovery was already evident in the collaboration and implementing interventions items and that, perhaps, it could be drawn out further within these items rather than individual items being added.

- **Appropriately omitted**

One option to overcome the difficulties with measuring change item (which isn't always present) and the formulation item (when it isn't explicitly mentioned in session) was an "appropriately omitted" option.

kate muse
Appropriately omitted option?

- **Formative feedback**

Some assessors felt the ACCS didn't provide much opportunity to deliver formative feedback. However, they hadn't realised they were using a "numeric" version of the feedback form. They felt that providing space for detailed qualitative feedback about each domain would have solved this problem (which is what the formative feedback form does).

kate muse
Confusion and misunderstandings.

kate muse
Useful formative tool

Assessors felt that filling out the form without the qualitative comments made it a very different experience. Without completing qualitative feedback assessors felt they weren't really "focusing" on what they were doing.

kate muse
The manual isn't used

Some assessors felt the ACCS was an excellent tool for providing formative feedback (better than CTS-R) due to the specificity of the scale (see specificity section above). Others commented that the ACCS provides them with a better 'template' for formative feedback.

kate muse
Strengths of the ACCS

kate muse
Useful formative tool

Assessors felt that completing the formative feedback would mean that the scale would take longer to use than the CTS-R because you are forced to put comments about all the different specific aspects of competence outlined in the ACCS. However, assessors felt this would be worth it to help students learn. (Again, see specificity section above).

Some felt that the specificity (again, see above) could be problematic when giving feedback to therapists' whose sessions were not so skilful. It was felt that, with less specific scales like the CTS-R you can "fluff out" your feedback and don't have to highlight everything the therapist did wrong so much. Others felt that this was not the case, that actually the formative feedback sections allowed you to provide positive comments/praise etc. On discussion it seemed that there was again a misunderstanding here that the ACCS did not allow space for formative feedback for each item (See above comment).

kate muse
Confusion and misunderstandings

kate muse
Useful formative tool

It was thought that it would be helpful for students to have a copy of the ACCS at the start of term so that they understand what they need to be achieving in CBT sessions.

kate muse
Strengths of the ACCS

It was felt that the ACCS provides a useful framework for "how to do therapy" for trainees and could be used as a good practice guide. So this made the ACCS a useful learning tool.

It was felt that it might be helpful for feedback purposes to have the description of each item on the feedback form. Although it was recognised that there is a balance to be had between the level of detail and the length of the feedback form.

- **Problems with manual**

For some items basic reads "problems in one or more of the following areas" whereas good/advanced reads "Good degree of skill in the following areas, with only minor problems or inconsistencies". This implies that you have got to check all of them- you must to all of the bullet points otherwise you can't score a 3 or 4. However, for some level 4's you might need to demonstrate skill in all of the areas but not for all of them. It was suggested that the wording should be changed to make this clearer.

Typo - page 20 in the basic section - should be spoke too rapidly.

Typo – typo "5" on Domain 8 - 8.1

kate muse
Typographical errors

- **Familiarity**

Some assessors felt that, although the ACCS was in some ways longer/more complex at first sight. Once they had understood it and practiced using it a bit, they could go through it much faster.

kate muse
More detail / specificity

- **CBT interventions**

Some assessors wondered whether there were too many items in the CBT interventions domain and whether some could be removed. Choosing suitable interventions was suggested. However, others felt this was an important component which needed to be retained. Another suggestion was to move this to the formulation domain as it refers to linking interventions to the formulation.

kate muse
More detail –specificity

- **Training**

Assessors felt that the opportunity to go through the ACCS scale and discuss what items include, how to interpret items on the scale, and to have the opportunity to practice using the scale (especially to see discrepancies / differences between assessors / hear how others assess items etc.) would have helped reduce some of the confusion they felt / difficulties they faced. They also felt this would help to improve reliability.

kate muse
Confusion and misunderstandings.

Appendix 2. Initial Thematic Framework from the Assessment of Core CBT Skills (ACCS) Focus Group

Theme	Notes	Supporting data (distilled summary from focus group recording)
Strengths of the ACCS	Things assessors particularly liked about the ACCS.	• Well operationalized. • Very relevant to CBT competence. • Simple to use. • Manual was very clear. • Assessors liked the generic introduction and found the generic performance bandings a useful "anchor" and default rating. • Assessors liked the layout and style of the ACCS. They liked that it was colour keyed. • Useful for providing formative feedback (see section below). • Increased specificity / detail (see below). • Useful learning tool (see below) • Some assessors liked the patient complexity rating.
Active formulation	The formulation should be actively used and updated and this is currently unclear in the manual/feedback form. Revising the title may help overcome this.	• Raised the possibility of putting "Coherent and dynamic formulation" on the actual rating form, someone who is not using the manual might settle for coherent and not bother to judge if the formulation is being "actively" used and updated. • It was suggested that this item could be called "in session formulating" instead.
Break down formulation domain	Could formulation be broken down into multiple items? This could be done theoretically- but it would be hard to rate these items.	• Is formulation actually 2 items (i.e. coherent and regularly updated/revised in light of new information)? • Formulation could be broken down into multiple components (is it coherent, suitable for stage of treatment, is it shared with the patient). • However, all aspects of formulation are not often evident in every session, making it harder to include broader issues relating to the formulation.
Use of supporting materials	There was uncertainty about whether supporting documentation should/could	• Assessors didn't always realise they should also be referring to the written formulation or information about the measures being used which is submitted in the supporting documentation. Consequently some used this material but others did not.

	be used and if so, for which items. Clarification in the feedback form may help.	• One possibility for overcoming this was to add guidance to the feedback form noting this • Another solution is to make this guidance more explicit in the manual.
Can formulation be rated?	Debate about whether the formulation should be evident in each session. Links to active formulation above. Also links to use of supporting materials above.	• Some assessors felt the formulation wasn't always (or often) explicitly referred to in the session, especially for later sessions, making it hard to rate. • But- other assessors felt that the formulation should be clearly evident in each session- especially when assessors view the written formulation in combination with the recording (they can get a feel about whether what is happening in the session matches what is written in the formulation). E.g. a simple maintenance cycle discussed in session which reflects/fits in with overarching formulation.
Idiosyncratic/informal measures of change are acceptable	It needs to be made clear in the title/manual that it is appropriate to include idiosyncratic measures (i.e. it's not always necessary to use standardised questionnaires).	• Some assessors didn't realise that idiosyncratic ratings were acceptable within measuring change It needs to be clarified that informal measures are appropriate (i.e. that the item doesn't only refer to the use of formal, standardised measures). • To make this clear, it was suggested that the manual could read "standardised measure(s) and/or idiosyncratic …….". • For some items basic reads "problems in one or more of the following areas" whereas good/advanced for the measuring change item reads "Good degree of skill in the following areas, with only minor problems or inconsistencies". This implies that you have got to use idiosyncratic and standardised measures to score a 3 or 4.
Should measures be used?	Should measures actually be implemented/used in each session? Some thought yes, others thought no! If they are not used (legitimately), would a therapist be unfairly penalised? (overlap with theme below)	• Some assessors felt it was appropriate not to use measures in every session. • Some assessors felt it was important to pay attention to measures and so include this in the scale. • Some assessors felt there should be a culture shift in that measures should be viewed as being an important component of delivering CBT- some commented that this shift was already beginning across a number of settings. • Some assessors felt that, if supporting documentation was used, then it should always be possible to rate the measuring change item.
Should measures be thoroughly reviewed?	Should measures actually be thoroughly reviewed in each	• It was also noted that it is not always relevant to carry out a comprehensive review of measure(s). Sometimes, later in treatment, it's sufficient for the patient to give a

	session- especially later sessions? If they are not, would a therapist be unfairly penalised? (overlap with theme above)	simple self-rating of progress and both parties know what this indicates. Then therapy mileage is in moving on to relapse management and consolidating progress rather than reviewing rating itself. Would a therapist be penalised for not being more detailed?
Appropriately omitted option?	Should an appropriately omitted option be available (especially for formulation item / measuring change item). Some thought yes other no. (links with scoring system and with should measures be used/reviewed & can formulation be rated)	• Some assessors felt that, if supporting documentation was used, then it should always be possible to rate the measuring change item. So, appropriately omitted would not be needed. • One option to overcome the difficulties with measuring change item (which isn't always present) and the formulation item (when it isn't explicitly mentioned in session) was an "appropriately omitted" option.
Confusing scoring system?	Some found the scoring system confusing (especially the mean domain score).	• Some assessors, and a number of therapists (when completing self-ratings), included total scores for each domain, rather than mean scores. Additionally, the final score was difficult for some to add up due to the use of mean total scores. • Some assessors liked having a total score for each domain so that you could get a feel for how therapists had performed within each domain. (Helpful for therapists and assessors). Others felt this wasn't necessary as you can get a feel of performance within each domain without total scores as there aren't many items in each domain. • Very few assessors were able to get the button on the word document which provides total domain scores etc. to work.
Enough sensitivity?	Some wondered whether the ACCS was sensitive to change / subtleties in skill sets. Add ½ marks? Or add another point between basic/good.	• Some felt it was too much of a jump from basic to good, and so were looking for a banding in between. • Some wanted more than four points on the scale to reflect finer differences in competence. • Some assessors felt the ACCS picked up well on people who were very skilful and those who were not skilful but perhaps missed the subtlety of those "in the middle".

		• It was noted that a lot of therapists were rated as "3". Some assessors worried that the ACCS wouldn't be able to monitor change in skills. • Some felt the CTS-R was better able to capture the range of skills demonstrated by trainees as it provides more points on the scale. • Some assessors found they wanted to put half marks if it felt as though the supervisee fell between two points on the scale –some assumed this was ok to do, others thought they weren't supposed to use ½ marks. It was felt that this wasn't clear in the manual. It was also felt that ½ marks were necessary in order to accurately represent the full range of abilities.
Interpretation of ACCS scores	What do the scores on the ACCS mean?	• Some assessors struggled with the interpretation of the ACCS (what do the scores really mean when combined)? Where does competence lie? It was suggested that further guidance could be provided within the manual.
Not using the manual.	Many didn't use the manual at all or only looked at it briefly. Perhaps exacerbated by providing numerical ratings only (don't have to justify ratings / provide detail). Overcome it be putting more detail in the feedback form? Could be overcome but putting more detail/guidance in the feedback form? Also be providing assessor training? (Why? Because not using the manual and because thinking about the CTS-R.).	• Many assessors didn't know what the generic performance bandings were. • Some assessors highlighted issues which were in the ACCS as missing (e.g. using formulation to guide interventions- which appears in choosing suitable interventions, keeping to the agenda- which appears in maintained focus). Upon discussion it was evident that they had only briefly read the manual and the used the feedback form without referring back to the manual. Consequently they didn't know what all of the items referred to and made their own inferences based on the item titles alone. • Some assessors commented that they worried they were being a bit "slapdash" when using the ACCS- just "plumping" for a number as they didn't have to support this with formative/qualitative feedback/comments. • Some assessors wondered whether they were completing detailed enough ratings. They commented that they read through the manual once, thought they understood it so didn't look again. In retrospect they realised that this was, perhaps, not always the case and it would have been better to have looked at the manual again and/or to have had training in how to use the scale. • Assessors felt that filling out the form without the qualitative comments made it a very different experience. Without completing qualitative feedback assessors felt they weren't really "focusing" on what they were doing.

		• Possibility of putting generic guidance/definition of performance bandings (in summarised form) on the actual feedback rating form – as many people said they didn't look back at the manual again. • Assessors felt that the opportunity to go through the ACCS scale and discuss what items include, how to interpret items on the scale, and to have the opportunity to practice using the scale (especially to see discrepancies / differences between assessors / hear how others assess items etc.) would have helped reduce some of the confusion they felt / difficulties they faced. They also felt this would help to improve reliability.
Comparison with CTS-R	Comparisons with the CTS-R lead to frustration and confusion.	• Assessors noted that they needed a "mind-set shift" from CTS-R in order to use the ACCS. • Assessors felt frustrated because they had to look at the manual, whereas because they have been using the CTS-R for a long time, they already know how it works and so don't need to look at the manual. • Assessors frequently became confused about what the items in the ACCS were due to their familiarity with the CTS-R.
More detail / specificity	ACCS has more detail / specificity- good for formative feedback, not overly long. (links with formative tool).	• Assessors liked the specificity of the ACCS: the way the ACCS broke down / expanded upon aspects of competence. E.g. the way homework was split into reviewing, rationale, planning homework etc. E.g. the way the agenda is broken down into different aspects. E.g. feedback being split into multiple domains/expended upon. E.g. splitting up interventions into multiple components. • The scale was longer/more detailed than the CTS-R. • Specificity allows more detailed assessments of skill / more in-depth feedback (compared to the CTS-R). • Once assessors' understood ACCS & became familiar with it / practiced using it, they felt it was fairly quick to use. • People felt this length was about right. • Too many items in the CBT interventions domain? Remove some? Or move to formulation domain?

Missing item: Socratic enquiry	It was felt this item was missing from the ACCS and should be included, but it was hard to quantify/operationalise what this was.	• Some assessors felt the ACCS was missing eliciting questions / Socratic enquiry. • It was identified that this was a slightly "woolly" concept. Some assessors noted that Socratic enquiry is hard to quantify • It was noted that there was something in the collaboration item which overlapped with Socratic questioning / guided discovery but it was felt that they were distinct. • It was felt it was demonstrated to some degree in collaboration and implementing interventions items but participants felt this wasn't enough. • Some assessor felt Socratic enquiry should in some ways be in every item- it was questioned whether it should perhaps be mentioned within each item? • Some assessors felt Socratic enquiry was already evident in the collaboration and implementing interventions items and that, perhaps, it could be drawn out further within these items rather than individual items being added.
Useful formative tool	The ACCS is useful for providing formative feedback. (links with more detail/specificity).	• ACCS is an excellent formative feedback / learning tool. • Better than CTS-R. • Due to the specificity of the scale. • Provides a helpful 'template' for formative feedback. • Provides a useful "how to" framework or good practice guide = useful learning tool for trainees. Helps them to remember / understand what they were meant to be doing. • Downside = scale takes longer to use than the CTS-R because you are forced to put formative comments about all the different specific aspects of competence outlined in the ACCS. But- worth it to help students learn. • Improvements = include description of each item on the feedback form (as long as length of the feedback form not too greatly increase).
Typographical errors		• Typo - page 20 in the basic section - should be spoke too rapidly. • Typo – typo "5" on Domain 8 - 8.1

Appendix 3. Revised Assessment of Core CBT Skills (ACCS) Feedback Form, Manual and Cover Sheet (version 3).

Assessment of Core CBT Skills (ACCS)

www.accs-scale.co.uk

USER MANUAL

Final Version (3) - July 2014

Scale contributors: Kate Muse, Freda McManus, Sarah Rakovshik & Helen Kennerley

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus and was conducted in affiliation with the University of Oxford's Department of Psychiatry and the Oxford Cognitive Therapy Centre (OCTC).

With special thanks to those who participated in the development and evaluation of the ACCS, particularly staff at OCTC, First Step, University of Derby CBT course, and South of Scotland CBT course.

Contents Page

Contents Page 437
General Guidelines 438
Domain 1 - Agenda Setting 442
1.1. Suitable Items 442
1.2. Feasible Agenda 443
Domain 2 - Formulation 444
2.1. Coherent and Dynamic Formulation 444
Domain 3 - CBT Interventions 445
3.1. Appropriate Intervention Targets 445
3.2. Choosing Suitable Interventions 446
3.3. Rationale for Interventions 447
3.4. Implementing Interventions 448
3.5. Reviewing Interventions 449
Domain 4 - Homework 450
4.1. Reviewing Homework 450
4.2. Choosing Suitable Homework 451
4.3. Rationale for Homework 452
4.4. Planning Homework 453
Domain 5 - Assessing Change 454
5.1. Choosing Suitable Measures 454
5.2. Implementing Measures 455
Domain 6 - Effective Use of Time 456
6.1. Pace 456
6.2. Time Management 457
6.3. Maintained Focus 458
Domain 7 - Fostering Therapeutic Relationship 459
7.1. Interpersonal style 459
7.2. Empathic Understanding 460
7.3. Collaboration 461
Domain 8 - Effective Two-way Communication 462
8.1. Patient Feedback 462
8.2. Reflective Summaries 463
Patient Complexity Rating 464
References 465

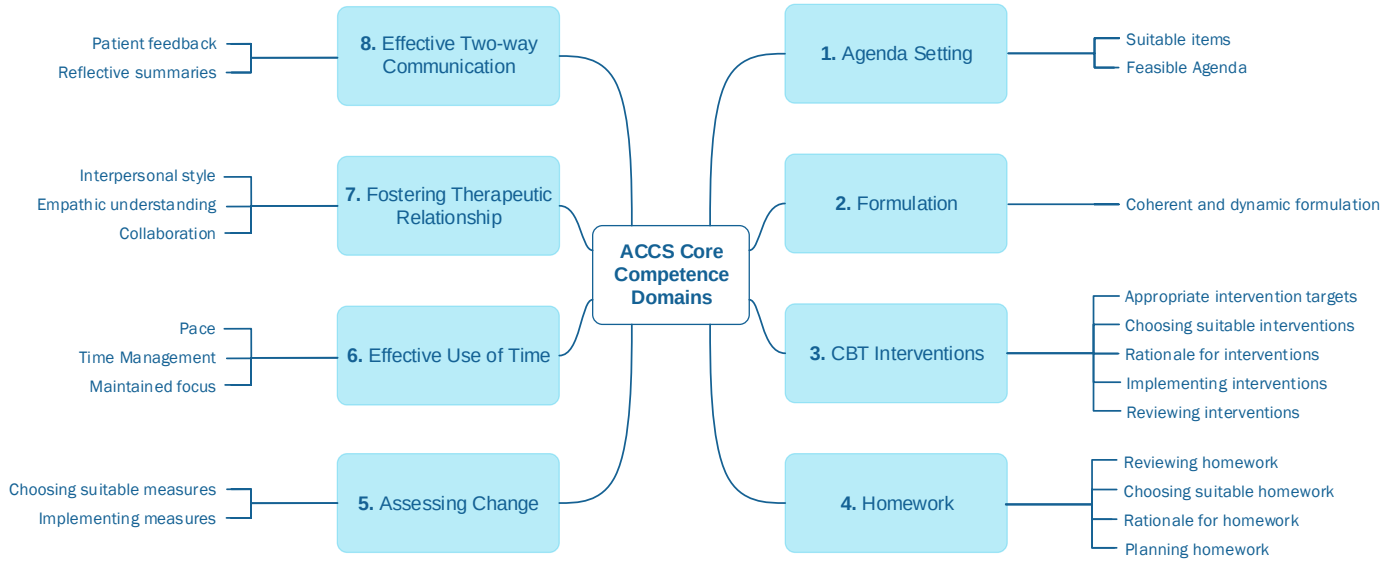
General Guidelines

Scale Scope

The Assessment of Core CBT Skills (ACCS) provides a framework for assessors to deliver formative and summative feedback about a therapist's performance within an observed treatment session. The scale assesses core general therapeutic and CBT-specific skills required to appropriately deliver individual cognitive-behavioural therapy to adults experiencing mental health problems. The scale focuses on competences which are transdiagnostic, rather than skills which are specific to a particular disorder or treatment protocol. The scale is also intended to assess a mid-treatment session and so it does not include skills which, although important, do not occur in most active treatment sessions (e.g. goal setting, relapse prevention etc.).

Core competences were identified for inclusion in the scale on the basis of a review of relevant literature. In particular, the authors drew upon the Cognitive Therapy Scale (CTS, or Cognitive Therapy Rating Scale: CTRS, www.beckinstitute.org), the Cognitive Therapy Scale-Revised (CTS-R: Blackburn et al., 2001) and Roth and Pilling's (2007) competence framework.

As demonstrated in the figure below, the scale features eight competence domains, each of which includes one to five items:



The scale provides a numerical total score and an average item score, which are designed to identify the therapist's competence across the session overall. Numerical scores are also provided for each item in order to provide more detailed feedback as to the therapist's specific areas of strength and weakness. In addition, the feedback form includes space for assessors to include further qualitative feedback and to indicate what steps could be taken to further develop competence. Thus the scale can be used both to establish whether therapists have reached an adequate standard of competence (useful for accreditation, evaluating the impact of training, selecting and monitoring research trial therapists) and to provide detailed feedback to aid a therapist's development and progression (useful within supervision and training settings, for continuing personal development, and examining skill acquisition).

How to use the ACCS

This manual is intended to provide guidance for assessors as to how to make judgements about the quality or skilfulness of therapists' performance using the ACCS rating scale. It is recommended that assessors spend some time reading and familiarising themselves with the manual before conducting an assessment. Initially, the manual should be consulted on an item by item basis alongside completing the feedback form. Once the assessor becomes familiar with the scale, the manual can then be used as a frame of reference which can be consulted whenever questions or uncertainties arise. It is also recommended that assessors periodically re-review the manual to prevent assessor-drift.

Step 1: Provide a numerical rating for each item

Each of the eight domains is broken down into one to five items, each of which is scored on a four-point scale. If the therapist is assessed as falling between two descriptors, the assessor can select a ½ mark (i.e. 1.5, 2.5, or 3.5). The overarching definitions of the scale are as follows:

Performance Band	Generic Definition of Performance Band
1. Limited	• Therapist fails to include feature outlined. • ***Or*** therapist demonstrates a significant absence of skill or an inappropriate performance which is likely to have negative therapeutic consequences.
2. Basic	• Therapist's performance is somewhat appropriate with some degree of skill evident. • However, major substantive problems are evident.
3. Good	• Therapist demonstrates a good degree of skill with no major problems. • However, minor problems or inconsistencies are evident in the therapist's performance.
4. Advanced	• Therapist consistently demonstrates a high level of skill with only very few and very minor problems.

Concrete examples of behaviours consistent with each scale point are provided for each item individually. These 'exemplar therapist behaviours' are intended to provide guidance as to the type of performance which is consistent with each point on the rating scale for that particular item. However, there may be occasions where a therapist demonstrates behaviours from more than one performance banding. Under these circumstances, the assessor is asked to use the overarching definitions outlined above to inform their decision as to which banding to assign. For example, within the 'choosing suitable homework' item, the therapist may have selected homework which related to key maintenance cycles discussed in-session (band 4) but the homework task was too challenging for the patient at this time (band 2). As the homework being too challenging for the patient constitutes a 'major substantive problem', the assessor would assign a rating of 2- Basic. When providing qualitative feedback the assessor has the opportunity to explain why the rating of 2-Basic was given as well as highlighting aspects of the therapist's performance which were consistent with the 4-advanced banding.

The feedback form should be used to assign a rating for each item on the scale. An overall session score can then be provided by adding the score assigned for each of the items. An average item score can also be provided by dividing the total score by 22 (the number of items).

Issues to consider-

- The exemplar behaviours are not able to provide an exhaustive list of possible performance indicators. Furthermore, ratings of competence will inevitably include a degree of subjectivity. However, in order to promote consensus between different assessors, it is important that assessors base their ratings on the item-specific exemplar therapist behaviours wherever possible.
- Assessors may find it helpful to make notes of about the therapist's performance that correspond with each item as they view the session. However, assessors should only score items after they have listened to the entire session.
- The 'halo effect' indicates that, when an assessor rates individual aspects of a therapist's competence, they are often unduly influenced by their overall impression of the therapist's performance (Cooper, 1981). As a result the therapist is provided with a uniformly high or low rating, rather than specific feedback as to their strengths and weaknesses. It is important for assessors to bear in mind that therapist's performance may well vary across domains- indeed a therapist can perform poorly in one competence domain and then do very well in another. It is therefore important to score each item independently of other items, and to avoid relying on an overall global impression.

Step 2: Provide qualitative feedback for each domain:

After completing numerical ratings for each domain, assessors are given space to highlight areas of strength and learning needs. Assessor can draw on the exemplar therapist behaviours provided and the session material being assessed to provide examples of what made some aspects of the therapist's performance more successful than others. Where areas of weakness are identified, assessors can suggest alternative, more skilful ways of working within this domain and highlight ways in which the therapist could further develop their skills in this area. Such in-depth feedback is invaluable in aiding further development and progression.

Step 3: Provide a global performance rating

The scale also includes a global competence item which is rated using the generic four-point scale outlined above. This global rating should reflect the assessor's overall impression of the therapist's performance within the session as a whole. This global rating will normally correspond to domain ratings. However, this will not always be the case. The assessor may, for example, wish to provide a global rating which reflects the presence or absence of specific aspects of competence which, although important, have not been included within the core competence domains of this scale because they are not relevant across patients and/or treatment sessions.

Step 4: Rate patient complexity:

Patient complexity has not been incorporated within the scoring system of individual items as this may unduly confound the ratings. Furthermore, reserving the highest ratings for performance demonstrated in the face of difficulties or added complexity may unfairly prevent those working with more straightforward cases from being awarded a high performance rating. However, assessors are asked to provide a separate patient complexity rating at the end of the scale. This is so that the context within which the therapist's performance has been evaluated can be taken into account when viewing the ACCS ratings (i.e. it identifies whether the therapist has competently delivered CBT with a patient who is straightforward to work with vs. a patient who has more complex difficulties).

Minimum Competence Score

A minimum total score above which a therapist can be considered competent has not been provided for the ACCS. This is because further research is required to establish an empirically proven 'tipping point' for competence using the ACCS. Furthermore, the level of competence and thus the score on the ACCS that is required may vary according to the context and purpose of the assessment. For example, when determining whether candidates are suitable for inclusion on a CBT training course, a lower level of performance may be acceptable, whereas those seeking accreditation may be required to demonstrate a higher level of performance. However, the total score which a therapist would achieve if their performance was consistently rated as 1, 2, 3, or 4 is provided to aid in interpretation of the total score: One-limited = total score of 22, Two-basic = total score of 44, Three-good = total score of 66, Four-advanced = total score of 88. The average item score also provides an indication of the therapist's average performance.

Assessment Materials

The ACCS is designed to assess therapists' performance within an 'active' CBT treatment session. Assessments should therefore be made on the basis of a mid-treatment session which involves CBT intervention strategies, rather than an initial assessment session or end of treatment/review session.

Ratings should be based upon source material which retains the most information possible, as this is more likely to provide the most accurate representation of the treatment session and thus the most accurate assessment of the therapists' competence (Waltz, Addis, Koerner, & Jacobson, 1993). As it is typically not possible to physically sit in on treatment sessions, rating are usually made by viewing recordings of a treatment session. Visual recordings provide the richest source of information and can therefore be viewed as the optimal source material, followed by audio recordings. Ratings should not be made on the basis of session transcripts or therapists' verbal or written reports of the session.

It is recommended that assessors base their ratings on a full treatment session, rather than segments of a session. This is because the reliability of individual item ratings is poorer when assessors view session segments rather than full sessions, with a number of aspects of competence not being assessed at all using this method (Weck, Bohn, Ginzburg, & Ulrich, 2011).

In addition to submitting a session recording, it is recommended that therapists complete the ACCS submission cover sheet provided with this scale. This briefly outlines key contextual information such as the stage of therapy, the patient's presenting problem, the

session agenda, an outline of homework assignments, the main focus of the session, a brief summary of interventions carried out so far, a rational for chosen treatment interventions, a treatment plan based on the formulation and treatment targets/patient's goals, and any relevant patient outcome data. Any materials completed in the session or for homework (e.g. questionnaires, diaries, thought records etc.) should also be submitted alongside the recording. Viewing this additional contextual information will help assessors make an informed rating of the therapist's performance.

Assessor Selection

It is unlikely that assessors who are themselves novice therapists would be able to provide detailed formative feedback about a therapist's specific areas of strengths and weaknesses. Indeed research suggests that therapist competence cannot be reliably rated by trained novices (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011). However, at present it is unclear what expertise is necessary to assess competence. As a minimum, it is recommended that assessments are carried out by therapists who have received formal training or accreditation as a CBT therapist.

There is some evidence that assessors rate competence differently depending on their level of independence (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012a), although it is not known whose ratings are more accurate. There are a number of pros and cons involved with using either supervisors who have a relationship with the therapist or independent assessors to rate a therapist's performance using the ACCS. Independent assessors are less likely to be influenced by social demands, relationship dynamics or information other than that obtained through viewing treatment sessions (e.g. prior competence, ability in other domains). On the other hand, supervisors may have access to a greater wealth of contextual information (e.g. treatment context, therapist's broader work) and be better placed to provide appropriate formative feedback. Those selecting assessors will therefore need to weigh up these pros and cons and make a decision based on pragmatics and the purpose of the assessment. If it is decided to use supervisors, one possibility for overcoming potential biases is to implement reliability checks by independent assessors.

Therapist Self-ratings

The scale can also be used by therapists to rate their own competence. As the ability to accurately rate one's own competence has been called into question, it is not recommended that self-ratings be used as a formal measure of therapist competence (Muse & McManus, 2013). However, self-ratings can be a useful learning tool. This is particularly the case when they are used in combination with assessor-ratings as this process of comparison can help to foster self-reflection and highlight on-going learning needs.

Cautions

The ACCS is designed to assess whether a therapist has demonstrated the core generic and CBT specific skills required to deliver effective CBT within a given treatment session. Thus a number of aspects of competence are not assessed by the ACCS. First, the ACCS does not assess therapists' knowledge or understanding of CBT, aspects of competence which can instead be assessed using multiple choice questionnaires, essays, case reports or clinical vignettes (Muse & McManus, 2013). Second, as the ACCS focuses on 'intervention competence', broader professional skills (e.g. ethical practice, effective use of supervision) are not covered. Third, the scale focuses on competences which are transdiagnostic, rather than skills which are specific to a particular disorder or treatment protocol. Fourth, as the ACCS assesses core CBT skills evident during active, mid-treatment therapy sessions, it does not assess therapists' assessment or relapse-prevention skills. Hence it is recommended that, where the scale is used for summative assessment purposes, the ACCS should not be used as a stand-alone measure of competence. Instead the ACCS should form part of a multi-method competence assessment programme.

Although competence is often inferred from rating one or two isolated treatment sessions per therapist, recent studies suggest that a much higher sample of clinical work needs to be assessed in order to reliably assess therapist competence (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012b; Keen & Freeston, 2008). Furthermore, collecting multiple assessors' ratings of the same treatment session can increase reliability (Vallis, Shaw, & Dobson, 1986) and reduce the impact of the halo effect (Streiner & Norman, 2003). Thus in order to be confident of a reliable judgement as to a therapist's competence, multiple assessors would be used to assess a number of therapy sessions drawn across different patients (between 15 and 24 sessions rated by a single assessor, or 19 sessions rated by two assessors: Keen & Freeston, 2008). Although financial and time constraints often make this unfeasible, it is important to recognise the limitations involved in drawing conclusions about a therapist's competence and thus fitness to practice on the basis of a limited number of session ratings.

Domain 1 - Agenda Setting

NB This item refers to **planning and setting** an appropriate agenda **not adhering to it** which should be rated under Item 6.3 'Maintained Focus'.

1.1. Suitable Items

Ability to help the patient identify and prioritise specific, relevant and appropriate agenda items.

1. Limited

Absence of skill or an ***inappropriate performance:***

- Therapist made little or no attempt to outline any specific topics to be covered in the session (e.g. asked "How would you like to spend our time today?" with no further follow-up).

2. Basic

Therapist helped the patient identify and prioritise topics to be covered but a ***number of items*** had ***major substantive problems*** in one or more of the following areas:

- Vague, unclear or overly general (e.g. my anxiety).
- Not suited to stage of therapy or patient's presentation.
- Not relevant to patient's presenting problem or treatment goals [1].
- Unlikely to make progress in solving the problem [1].
- Inappropriate given potential risk factors [2].

- ***And/or*** a number of relevant standard items inappropriately omitted [3].

3. Good

Therapist helped the patient identify and prioritise items which met ***most*** of the following criteria, with only ***minor problems or inconsistencies***:

- Well-defined, specific, and measurable (e.g. improving difficulty getting to sleep, addressing financial concerns).
- Suited to stage of therapy and patient's presentation.
- Relevant to patient's presenting problem / goals.
- Likely to make progress in solving the problem.
- Appropriate given potential risk factors.

- ***And*** included most relevant standard items.

4. Advanced

Therapist ***consistently*** helped the patient identify and prioritise items which met ***all*** of the following criteria:

- Well-defined, specific, and measurable.
- Suited to stage of therapy and patient's presentation.
- Relevant to patient's presenting problem / goals.
- Likely to make progress in solving the problem.
- Appropriate given potential risk factors.

- ***And*** included all relevant standard items.

[1] NB: Attention may sometimes need to be given to items the patient deems important, even if these do not appear relevant to their presenting problem / treatment goals or it is unlikely that progress can be made in solving the problem (e.g. allowing the patient time to 'vent' when they have nobody else to talk to about the issue).

[2] Potential risk factors = risk of harm to self or others (e.g. expression of suicidal wishes), degree of distress associated with problem.

[3] Standard items = items which should usually be included on the agenda (e.g. mood check, week review, key learning from previous session, homework review, setting homework, session feedback etc.). NB: There may be occasions when it is appropriate *not* to include all standard items on the agenda of a session (e.g. if a routine for carrying out a mood check at the beginning of every session has been firmly established, it may not be useful to restate this agenda item).

1.2. Feasible Agenda

Ability to set an agenda which is realistic and feasible given the time available.

1. Limited

- Therapist made little or no attempt to outline any specific topics to be covered in the session

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas***:***

- Items far too numerous, large or overwhelming to be feasible / enable any progress in solving the problem.
- ***Or*** too few or insignificant items which could easily be covered in a very small time frame (e.g. reviewing a small piece of homework).

2. Basic

Major substantive problems:

- Mismatch between available time and number of items ***or*** scope and significance of items- making it unlikely that adequate time could be spent on all items ***or*** items likely to be covered well before the end of the session.

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- Appropriate number of items with suitable scope and significance given the time available, making it likely all items could be adequately covered in- session.

4. Advanced

Consistently high level of skill:

- Optimal number of items with ideal scope and significance given the time available, making it highly likely that maximum use could be made of the session and that all items could be thoroughly covered.

Domain 2 - Formulation

An initial formulation is typically developed as part of the assessment process. However, a formulation should not simply precede treatment but should evolve throughout therapy as further clinical information arises and recurrent patterns and themes emerge. Thus a therapist's ability to formulate a patient's problems should be evident throughout the course of treatment.

NB

In addition to viewing the session recording, it is necessary to review any written or diagrammatic formulation materials in order to make an informed rating on this item.

Whether the therapist drew upon the formulation to highlight potential treatment targets and intervention strategies should be rated under domain 3 'CBT Interventions'.

2.1. Coherent and Dynamic Formulation

Ability to develop a clear formulation which draws upon appropriate evidence-based theory to offer a concise, comprehensive and personalised explanation of relevant history, triggers and maintaining features of the patient's problems.

1. Limited

- Therapist made very little or no attempt to formulate the patient's problems.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist's formulation of the patient's problems had major flaws (e.g. did not use evidence-based CBT theory [1], attempted to fit the patient into an inappropriate model, inaccurately reproduced a relevant CBT model, included largely irrelevant information, failed to incorporate key features of the patient's presentation [2], failed to identify links between different features of the patient's problems, formulation overly complex or far too simplistic given stage of treatment or patient's ability to understand, formulation incoherent or disjointed).
- If new clinical material emerged, therapist seemed unwilling or unable to review or revise the formulation.

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist utilised appropriate evidence-based theory to provide an account of the patient's problems. However, this account was limited (e.g. vague or abstract, unclear or confusing, too complex or simplistic given stage of treatment or patient's ability to understand, included information of limited relevance, failed to show how features of the patient's presentation related to each other or the patient's problems in a meaningful way).
- If new clinical material emerged, therapist found it difficult to or seemed reluctant to revise the formulation.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist utilised appropriate evidence-based theory to provide an explanation of the history, triggers and maintaining features of the patient's problems. This explanation was clear and logical but had some problems (e.g. lack of personalisation, lack of concrete and specific examples, missed opportunities to include relevant strength and resilience factors, a more straightforward or detailed formulation would have better suited the stage of treatment or patient's ability to understand).
- If new clinical material emerged, therapist made some attempt to integrate this into the formulation.

4. Advanced

Consistently high level of skill in the following areas:

- Therapist flexibly utilised appropriate evidence-based theory to provide a clear, concise, comprehensive and personalised account of relevant history, triggers and maintaining features of the patient's problems, as well as any relevant strength and resilience factors. This formulation was well suited to the stage of treatment and patient's ability to understand.
- If new clinical material emerged, therapist responded flexibly and fully incorporated this within the formulation.

[1] Relevant evidence-based theory should be used to inform the formulation. If a relevant evidence-based model is available, this should be used to develop the formulation (e.g. disorder-specific models such as Clark's model of panic). If not, the formulation should be based on an appropriate generic model (e.g. Beck's cognitive model of emotional disorder, personalised vicious cycles).

[2] Presentation features = behaviours, cognitions, emotions, physiological experiences, environmental factors, interpersonal issues.

Domain 3 - CBT Interventions

This item refers to the selection and application of cognitive and/or behavioural interventions (also referred to as CBT 'techniques') designed to promote therapeutic change by targeting cognitions, behaviours or emotions likely to be maintaining the patient's problems. A range of CBT interventions can be used to promote change. These include, but are not limited to, cognitive restructuring, exposure/response prevention, addressing safety behaviours, imagery re-scripting, examining pros and cons, thought records, activity monitoring/scheduling, behavioural experiments, role play, graded task assignment, problem solving, assertiveness or other skills training, behaviour modification.

NB A number of factors other than therapist competence can influence whether an intervention is successful in bringing about change. Please, therefore, focus on the quality of the cognitive-behavioural interventions and not whether change actually occurred.

3.1. Appropriate Intervention Targets

Ability to skilfully define, clarify and specify intervention targets which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were highly likely to be maintaining problems.

1. Limited

- Therapist did not identify any intervention targets (e.g. cognitions, behaviours, emotions).

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist lacked a clear rationale for selecting intervention targets (e.g. focused on cognitions, behaviours or emotions which neither evidence-based theory or the patient's formulation suggested were likely to be maintaining the patient's problems).
- Intervention targets were not appropriate to the patient's therapeutic context [1] (e.g. therapist tried to re-evaluate or develop new core beliefs too early in therapy).

2. Basic

Major substantive problems in one or more of the following areas:

- Interventions targeted cognitions, behaviours or emotions which were of limited therapeutic benefit (e.g. cognitions or behaviours were not linked to distressing emotions, relevant evidence-based theory suggested the targets could be maintaining the patient's problems but the idiosyncratic formulation suggested that the targets were irrelevant or peripheral to the patient's problems).
- Intervention targets were largely appropriate to the patient's therapeutic context but required fine tuning (e.g. therapist targeted safety behaviours which the patient was not yet ready to consider changing).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Interventions targeted relevant cognitions, behaviours or emotions which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were maintaining problems. However, these targets were not clearly specified (e.g. were generic, vague or ambiguous).
- Interventions targets were suited to the patient's therapeutic context (e.g. therapist targeted thoughts that were more amenable to change before progressing to more strongly held beliefs).

4. Advanced

Consistently high level of skill in the following areas:

- Interventions targeted relevant cognitions, behaviours or emotions which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were highly likely to be maintaining problems. These targets were skilfully defined, clarified and specified.
- Interventions targets were suited to the patient's therapeutic context.

[1] Patient's therapeutic context = patient's presenting problem, current capacity of the patient, stage of treatment etc.

3.2. Choosing Suitable Interventions

Ability to select cognitive-behavioural interventions which form part of a logical, coherent and unified treatment strategy which is likely to bring about therapeutic change in the treatment target(s) and is suited to the patient's therapeutic context. This selection was accurately guided by appropriate theory-based practice or practice based on evidence when possible.

1. Limited

- Therapist did not select any CBT interventions

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Interventions formed part of an incoherent treatment strategy (e.g. therapist appeared to select interventions in a random, hit-and-miss fashion). ***And/or*** this strategy was highly unlikely to bring about therapeutic change in the treatment targets (e.g. interventions were not at all matched to treatment targets, such as selecting imagery re-scripting to increase activity levels).
- Interventions were poorly suited to the patient's therapeutic context [2] (e.g. intervention far too complex for current capacity of the patient).
- Selection of interventions was not guided by theory or evidence-based CBT practice.

2. Basic

Major substantive problems in one or more of the following areas:

- Interventions formed part of a vague, unclear or disjointed treatment strategy. ***And/or*** this strategy showed little promise in bringing about therapeutic change in the treatment targets (e.g. interventions were poorly matched to treatment targets, such as selecting a core belief worksheet to work with automatic thoughts).
- Interventions were largely appropriate given the patient's therapeutic context but required fine tuning (e.g. too easy or overly ambitious, challenging or complicated for the current capacity of the patient).
- Selection of interventions was guided by inappropriate theory or evidence-based CBT practice (e.g. a treatment protocol which was only vaguely related to the patient's presentation) or therapist significantly misinterpreted or misapplied appropriate theory or evidence-based CBT practice.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Interventions formed part of a generally coherent therapeutic strategy which showed some promise in bringing about therapeutic change in the treatment targets (e.g. interventions were appropriate to the treatment targets, such as selecting activity scheduling to improve mood by increasing activity levels)
- Interventions were appropriate given the patient's therapeutic context (e.g. complexity and difficulty matched with patient's motivation and cognitive functioning).
- Selection of interventions was accurately guided by appropriate theory-based practice or practice based on evidence when possible.

4. Advanced

Consistently high level of skill in the following areas:

- Interventions formed part of a logical, coherent and unified treatment strategy which showed a great deal of promise in bringing about therapeutic change in the treatment targets.
- Appropriate timing and grading meant that interventions were optimally suited to the individual patient's therapeutic context.
- Selection of interventions was accurately guided by appropriate theory-based practice or practice based on evidence when possible.

[2] Patient's therapeutic context = patient's presenting problem, current capacity of the patient, stage of treatment etc.

3.3. Rationale for Interventions

Ability to facilitate the patient's understanding of the importance and potential benefits of interventions.

1. Limited

- Therapist did not implement any CBT interventions.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to facilitate the patient's understanding of the rationale underpinning interventions (e.g. did not explain the importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, these were ignored or disregarded or therapist responded in a negative manner (e.g. defensive, hostile).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist facilitated the patient's understanding of the rationale underpinning interventions with limited skill (e.g. developed a limited or confusing rationale for importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, therapist gave these limited consideration (e.g. acknowledging a question but failing to answer it).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist facilitated the patient's understanding of the rationale underpinning interventions with some degree of skill (e.g. developed a rationale for importance / potential benefits) but some minor problems were evident (e.g. did not relate rationale to the patient's presenting problems / formulation / treatment goals, conveyed the rationale with limited clarity).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, these were acknowledged and therapist made some attempt to discuss and resolve the issues.

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully facilitated the patient's understanding of the rationale underpinning interventions (e.g. clearly and succinctly conveyed a rationale for the importance and potential benefits which aligned with the patient's presenting problems / formulation / treatment goals).
- If the patient did not understand, was sceptical about the rationale for interventions or asked questions, therapist responded openly, fully discussed the issues and (where possible) resolved the issues (e.g. clarified any ambiguities).

3.4. Implementing Interventions

Ability to systematically implement intervention(s) in a fluent and articulate manner. To be sensitive and responsive to the therapeutic context and provide optimal levels of support, encouragement and praise.

1. Limited

- Therapist did not implement any CBT interventions.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Interventions were misused or were implemented incorrectly in a way that was not justified by therapeutic context (e.g. omitting 'evidence for the belief' when doing a thought record, terminating exposure before any anxiety habituation occurred).
- Therapist lacked skill in implementing interventions (e.g. therapist was inarticulate or incoherent, appeared unfamiliar with techniques).
- Therapist showed very little or no awareness of the therapeutic context ***or*** responded in an unhelpful or negative manner (e.g. therapist was too forceful, appeared discouraging, ignored or dismissed very high levels of arousal or patient distress).

2. Basic

Major substantive problems in one or more of the following areas:

- Interventions were not implemented in the manner in which they were intended and there was no clinical justification for these deviations (e.g. interventions were incomplete, were abandoned prematurely or were missing key components [e.g. therapist neglected to get ratings of pleasure/achievement in activity scheduling]).
- Therapist implemented interventions with limited skill (e.g. at times the therapist was unclear, difficult to understand or appeared confused or muddled).
- Therapist showed some awareness of the therapeutic context but responded with limited sensitivity (e.g. provided limited encouragement or was overly pushy) ***and/or*** appropriate flexibility (e.g. failed to adapt intervention despite indication that it was too distressing or complex for the patient to grasp).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Interventions were mainly carefully and methodically implemented in the manner in which they were intended, unless there was clinical justification for deviating from this. However, some minor skill deficits or inconsistencies were evident (e.g. thought record could have been more detailed).
- Therapist implemented interventions with adequate skill (e.g. therapist was articulate and comprehensible).
- Therapist was largely sensitive to the therapeutic context (e.g. provided appropriate support and encouragement) and made some attempt to adapt interventions to suit the patient's needs where necessary.

4. Advanced

Consistently high level of skill in the following areas:

- Clear and consistent evidence that interventions were carefully, thoroughly and systematically implemented from start to finish in the manner in which they were intended, unless there was clinical justification for deviating from this.
- Therapist consistently implemented interventions with a high level of skill (e.g. therapist was fluent, articulate, clear and comprehensible).
- Therapist was consistently sensitive to the therapeutic context (e.g. provided optimal levels of support, encouragement and praise) and was able to custom tailor interventions to suit the patient's needs where necessary (e.g. broke down a thought record into sections to make the task more manageable/less overwhelming, recognised and addressing subtle safety behaviours, modelled the intervention for a patient who felt it was too threatening).

3.5. Reviewing Interventions

Ability to conduct a comprehensive review of the results of interventions (whether positive or negative) in order to help the patient identify what they learned from the experience.

1. Limited

- Therapist did not implement any CBT interventions.

Or absence of skill or an ***inappropriate performance***:

- Therapist failed to review ***or*** conducted a brief or cursory review of the results of interventions which did not help the patient identify learning (e.g. asked "how do you think that went?" with no further follow up).

2. Basic

Major substantive problems:

- Therapist reviewed the results of interventions and made some attempt to highlight learning implications, but with limited skill (e.g. struggled to relate the results back to the relevant cognition, behaviours or emotions that they were designed to target, did not address negative results, simply told the patient what they had learned from the experience).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- Therapist adequately reviewed the results of interventions (whether positive or negative). This review helped the patient independently draw conclusions about the learning implications (e.g. therapist used curious and Socratic questions to examine outcomes in relation to prior predictions).

4. Advanced

Consistently high level of skill:

- Therapist skilfully engaged the patient in a comprehensive review of the results of interventions (whether positive or negative) which helped the patient to independently draw conclusions about useful and relevant learning implications. Therapist helped the patient link this learning to the relevant cognition, behaviours or emotions they were designed to target ***and*** to their formulation / treatment goals (e.g. used curious and Socratic questions to help the patient re-evaluate previous conclusions or construct new ideas).

Domain 4 - Homework

4.1. Reviewing Homework

Ability to conduct a comprehensive review of previous homework (whether completed or not) in order to help the patient identify what they learned from the experience.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist failed to review ***or*** conducted a brief or cursory homework review which did not help the patient identify learning (e.g. asked "how did you get on with the homework?" with no further follow up).
- Therapist responded to completed or uncompleted homework in an unhelpful or negative manner (e.g. scolding, judgemental, dismissed homework as unimportant).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist reviewed homework and made some attempt to highlight learning implications from completing (or not completing) homework, but with limited skill (e.g. learning had limited relevance to the patient's problems, therapist simply told the patient what they had learned from the experience).
- If relevant, therapist made some attempt to reflect on why non-completion occurred but failed to identify relevant emotional, cognitive or practical difficulties ***and/or*** did not identify ways to overcome future 'blocks' to completion.
- Completed homework was given limited recognition ***and/or*** therapist responded to non-completion with limited sensitivity and understanding.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist adequately reviewed homework. This review helped the patient to independently draw conclusions about the learning implications from completing (or not completing) homework (e.g. therapist used curious and Socratic questions to examine outcomes in relation to prior predictions).
- If relevant, therapist helped the patient identify relevant emotional, cognitive or practical reasons for non-completion ***and*** discussed ways to overcome future 'blocks' to completion (e.g. asked "how could you overcome xx next time?").
- Therapist responded to completed homework with appropriate encouragement ***and*** to non-completion in a

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully engaged the patient in a comprehensive homework review. This review helped the patient to independently draw conclusions about useful and relevant learning implications from completing (or not completing) homework and to link this learning to their formulation / treatment goals (e.g. therapist helped them to explore what it tells them about how best to understand or make progress with their difficulties).
- If relevant, therapist worked with the patient to identify relevant emotional, cognitive or practical reasons for non-completion ***and*** helped them identify specific ways to overcome future 'blocks' to completion (e.g. using a calendar for a patient who has difficulty remembering to complete homework, using problem solving to break homework down into less threatening components).
- Therapist responded to completed homework with optimal levels of encouragement and praise ***and*** to non-completion in a highly supportive, sensitive and open manner.

4.2. Choosing Suitable Homework

Ability to plan homework which is tailored to the therapeutic context and builds upon session material or previous homework.

1. Limited

- Therapist failed to plan further homework arising from the session or building on previous homework.

Or absence of skill or an ***inappropriate performance***:

- Homework was poorly suited to the patient's therapeutic context [1] (e.g. expecting the patient to carry out a highly feared exposure after first session or planned a difficult or large assignment for a patient with low motivation).

2. Basic

Major substantive problems in one or more of the following areas:

- Homework was largely appropriate given the patient's therapeutic context, but required fine tuning (e.g. overly ambitious, easy, challenging or complicated for the patient, 'standard' homework used without necessary adaptation).
- Homework was not related to issues (e.g. problems, cognitions, behaviours) dealt with in-session ***or*** previous homework, therefore restricting opportunities for relevant information gathering or generalising of concepts and skills learnt into daily life.
- Significant risk the patient would fail to benefit from homework or would have predictions confirmed in a way that was counter therapeutic.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Homework was appropriate given the patient's therapeutic context (e.g. amount and difficulty was achievable given the patient's motivation and cognitive functioning).
- Homework related to issues (e.g. problems, cognitions, behaviours) dealt with in-session ***or*** previous homework, therefore enabling some opportunities for information gathering and/or generalising of concepts and skills learnt into daily life. However, opportunities for more central homework tasks or further learning were missed.

4. Advanced

Consistently high level of skill in the following areas:

- Custom tailoring of homework meant it was optimally suited to the individual patient's therapeutic context (e.g. represented a moderate and realistically achievable challenge for the patient).
- Homework built upon important issues (e.g. problems, cognitions, behaviours) dealt with in-session ***or*** previous homework, therefore enabling opportunities for relevant information gathering and/or generalising of concepts and skills learnt into daily life.
- Therapist thoroughly considered the impact of likely outcomes of homework (e.g. set homework up as 'no lose' or worked with patient to make a plan for dealing with potential negative outcomes).

[1] Therapeutic context = patient's presenting problems, goals, stage of treatment, formulation, presentation etc.

4.3. Rationale for Homework

Ability to facilitate the patient's understanding of the importance and potential benefits of homework.

1. Limited

- Therapist failed to plan further homework arising from the session or building on previous homework.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to facilitate the patient's understanding of the purpose of homework (e.g. failed to develop a rationale for importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, these were ignored or disregarded ***or*** therapist responded in a negative manner (e.g. defensive, hostile).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist facilitated the patient's understanding of the purpose of homework with limited skill (e.g. developed a limited or confusing rationale for importance / potential benefits).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, therapist gave these limited consideration (e.g. acknowledging a question but failing to answer it).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist facilitated the patient's understanding of the purpose of homework with some degree of skill (e.g. developed a rationale for importance / potential benefits) but some minor problems were evident (e.g. did not relate rationale to the patient's presenting problems / treatment goals / formulation, conveyed the rationale with limited clarity).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, these were acknowledged and therapist made some attempt to discuss and resolve the issues.

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully facilitated the patient's understanding of the purpose of homework (e.g. worked with the patient to develop a clear and succinct rationale for the importance and potential benefits which aligned with the patient's presenting problems / treatment goals / formulation).
- If the patient did not understand, was sceptical about the rationale for homework or asked questions, therapist responded openly, fully discussed the issues and (where possible) resolved the issues (e.g. clarified any ambiguities).

4.4. Planning Homework

Ability to work with the patient to ensure they have a clear and detailed understanding of the homework task(s).

1. Limited

- Therapist failed to plan further homework arising from the session or building on previous homework.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Content or practicalities for homework were vague or unclear (e.g. experiment with going out).
- Therapist made little or no attempt to identify potential obstacles. ***Or*** if the patient raised potential obstacles, these were ignored or disregarded.

2. Basic

Major substantive problems in one or more of the following areas:

- Content and practicalities of homework were poorly specified (e.g. go into a specific situation and see if I get anxious).
- Therapist attempted to identify potential obstacles, with limited skill (e.g. asked complicated or vague questions such as "do you think you could complete the task?"). ***Or*** if the patient raised potential obstacles, therapist gave these limited consideration (e.g. acknowledging an obstacle but failing to address it).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Content and practicalities of homework were outlined, though some specific details were lacking (e.g. exactly what, where, when, how, outcome evaluation).
- Therapist made some attempt to prepare the patient for how to complete homework (e.g. talked through a homework example, made sure the task was written down).
- Therapist identified potential obstacles (e.g. asked relevant questions such as "What might get in the way of completing this task?"). ***And*** made some attempt to discuss ways to overcome these obstacles (e.g. asked "how might we overcome that?").

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully worked with the patient to agree a clear and detailed plan of exactly what homework involved (e.g. what, how, where, when, frequency, duration, outcome measurement).
- Therapist thoroughly prepared the patient for how to complete homework (e.g. worked through a homework example together in-session, used guided imagery techniques).
- Therapist worked with the patient to identify any potential obstacles, fully discussed these obstacles ***and*** (where possible) identified ways to overcome them (e.g. using problem solving to break homework into smaller tasks if the patient felt overwhelmed, discussing advantages and disadvantages of homework).

Domain 5 - Assessing Change

A range of methods can usefully be employed to assess change, these include standardised questionnaires **and/or** informal idiosyncratic measures (e.g. diaries, frequency counts, ratings of duration of event or experience, self-ratings of emotions and cognitions etc.). These methods should be employed ***across treatment*** to measure change in symptoms and movement towards goals over the course of treatment as a whole. However, these methods should also be employed ***within treatment sessions*** to measure the impact of cognitive-behavioural interventions completed within session and for homework (e.g. belief ratings pre/post intervention).

NB In addition to viewing session recording, it is necessary to review any relevant written materials in order to make an informed rating on this item (e.g. completed questionnaires, diaries, belief ratings, treatment goals, etc.).

5.1. Choosing Suitable Measures

Ability to select appropriate, clinically relevant standardised **and/or** idiosyncratic methods for measuring change in symptoms and associated features (beliefs, behaviours, feelings) and movement towards goals.

1. Limited

- Therapist neglected to measure change in symptoms, associated features or movement towards goals.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist selected standardised measures with inadequate reliability, validity or usability, despite the availability of better measures[1] ***and/or*** created inadequate idiosyncratic measures with poor usability (e.g. too complicated for the patient to complete), completely unsuitable response scales (e.g. yes / no to measure symptom change), or undefined targets (e.g. rate your 'feelings').
- Measures were irrelevant and of no clinical use (e.g. depression questionnaire for patient with anxiety, belief ratings of irrelevant cognitions, frequency of irrelevant behaviours).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist selected standardised measures with poor validity, reliability or usability, despite availability of better measures ***and/or*** created idiosyncratic measures which were difficult to use, had inappropriate response scales, or had poorly defined targets.
- Measures were of limited relevance and clinical use (e.g. general anxiety questionnaire for patient with panic disorder, belief ratings of peripheral cognitions).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist selected standardised measures with adequate reliability, validity and usability, though better measures were available ***and/or*** created usable idiosyncratic measures with relevant response scales and outlined targets (e.g. 0 to100 emotional intensity ratings, 0 to 8 or 0 to 10 problem severity ratings).
- Measures were relevant and served a clinical purpose (e.g. panic cognitions questionnaire for patient with panic disorder, belief ratings for related cognitions, rating severity of relevant problem).

4. Advanced

Consistently high level of skill in the following areas:

- Therapist selected the most reliable, valid and usable standardised measures which were available ***and/or*** created idiosyncratic measures with clearly defined targets and highly appropriate response scales (e.g. 0 to 100 rating of clearly specified behaviour).
- Measures were highly relevant and served an important clinical function (e.g. belief ratings for key cognitions, diary provided useful insight into relationship between different aspects of the patient's problems).

[1] Therapists should select the most reliable, valid and usable standard measures which were *available* to them. Thus therapists should *not* be penalised if they were unable to access 'gold standard' questionnaire measures (e.g. due to service constraints, cost).

5.2. Implementing Measures

Ability to administer measures at suitable time points across and within session and to skilfully review, interpret and respond to the information gleaned.

1. Limited

- Therapist neglected to measure change in symptoms, associated features or movement towards goals.

Or absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Monitoring schedule was random, haphazard, inappropriate or not clinically useful (e.g. no baseline function to measure change against [e.g. thought record without pre-intervention belief rating], patient overly burdened by completing measures too frequently, progress not regularly checked).
- Therapist failed to review or conducted an extremely brief or cursory review of completed measures which did not provide useful clinical information (e.g. didn't compare post-intervention scores with pre-intervention scores, lack of understanding of completed measures).
- If relevant, therapist made little or no attempt to identify or explore barriers (e.g. difficulty understanding or completing measures, doubts about the purpose of monitoring) or responded to barriers in an unhelpful or negative manner (e.g. scolded patient for incorrectly completing measure).

2. Basic

Major substantive problems in one or more of the following areas:

- Monitoring schedule was unsuitable or of limited clinical use (e.g. measures not used regularly enough or used more frequently than necessary, measures implemented at inappropriate time points).
- Therapist conducted a rushed or superficial review of completed measures which provided little useful clinical information (e.g. therapist showed a limited ability to interpret completed measures).
- If relevant, therapist made some attempt to identify barriers to monitoring and responded to these in an understanding manner. However, therapist made little or no attempt to explore or overcome barriers.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Monitoring schedule was appropriate (e.g. at the beginning, middle and end of treatment, pre/post interventions) and served a useful clinical purpose (e.g. provided insight into changes in patient's symptoms or impact of interventions).
- Therapist adequately reviewed completed measures and attempted to use this information for clinical purposes but some minor problems were evident (e.g. struggled to relate the results to the individual patient).
- If relevant, therapist identified barriers to monitoring, responded in a supportive manner and made some attempt to explore or overcome them.

4. Advanced

Consistently high level of skill in the following areas:

- Monitoring schedule was systematic, appropriate and served an important clinical function (e.g. provided clear and relevant information as to effectiveness of interventions or treatment progress).
- Therapist comprehensively reviewed completed measures and used information gleaned for relevant clinical purposes (e.g. exploring improvements or deteriorations to better understand maintaining factors, highlighting progress to encourage motivation, making decisions about where best to focus therapeutic efforts).
- If relevant, therapist thoroughly identified and explored barriers to monitoring. Where possible therapist worked with the patient to resolve issues (e.g. working through measure with the patient, modifying or simplifying measures).

Domain 6 - Effective Use of Time

6.1. Pace

Ability to pace the session in a manner which is well suited to the therapeutic context and patient's capacity for learning.

1. Limited

Absence of skill or an ***inappropriate performance:***

- Pace consistently much too fast or slow for therapeutic context [1] or patient's optimal learning (e.g. therapist repeatedly belaboured a point long after the patient had grasped the message, important issues were rushed- making it difficult for the patient to grasp or retain key concepts, therapist switched from topic to topic far too rapidly).

2. Basic

Major substantive problems:

- There were a number of times when the pace was too quick or slow for the therapeutic context or patient's optimal learning (e.g. therapist slightly overstressed a point, moved on from emotionally difficult topics too quickly, switched topics before conclusions had been drawn, spoke too rapidly or too slowly).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- For most of the session therapist set a pace which was appropriate given the therapeutic context and patient's learning capacity. However, there were one or two occasions when the pace was a little too quick or slow to be optimally effective (e.g. therapist failed to slow the pace suitably where cognitive difficulties arose, moved the patient along before they had fully understood concepts).

4. Advanced

Consistently high level of skill:

- Therapist was responsive to and consistently matched the pace of the session to the therapeutic context and patient's learning capacity (e.g. gave more time and attention to issues where cognitive difficulties arose, allowed the patient to experience emotions rather than rushing onto an intervention / another topic).

[1] Therapeutic context = patient's presenting problems, goals, stage of treatment, formulation, presentation etc.

6.2. Time Management

Ability to manage time within the session in a balanced and efficient manner.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Inefficient, unbalanced allocation of time meant a number of important issues were not covered at all (e.g. not enough time left to plan homework, summarise session content, exchange feedback etc.).
- Omitted issues were given limited or no acknowledgement.
- Session ran very significantly over ***or*** under scheduled time without clinical justification [1] (e.g. 50 minute session lasted 70 minutes or ended after 30 minutes with no clinical justification).
- Overly rigid adherence to a schedule was unhelpful to the patient (e.g. comments such as "We said we'd move on after 10 minutes so we need to move on now").

2. Basic

Major substantive problems in one or more of the following areas:

- Less than optimal use of time meant some important issues were not covered ***or*** were inadequately covered (e.g. insufficient time to review learning from an intervention or to plan how to take the work forward).
- Omitted issues were acknowledged but were not rescheduled.
- Session ran a little over ***or*** under scheduled time without sufficient clinical justification (e.g. 50 minute session lasted for 60 minutes or ended after 40 minutes).

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist's use of time was generally well balanced and efficient, meaning most important issues were adequately covered. However, to make optimal use of time, it would have been necessary to spend a little less or a little more time on some issues (e.g. not quite enough time left at end of session to plan homework thoroughly).
- If issues were not covered, this was somewhat appropriate given the therapeutic context (e.g. focusing the entire session on one agenda item which was more distressing / significant than anticipated).
- Omitted issues were rescheduled for a different time or format (e.g. next session, homework).
- Length of session was largely as scheduled ***or*** there was some clinical justification for deviating from planned schedule.

4. Advanced

Consistently high level of skill in the following areas:

- Consistently balanced and efficient allocation of time meant all important issues were thoroughly covered. ***Or*** if issues were not covered, this was necessary given the therapeutic context and this decision was made openly and explicitly with the patient.
- Omitted issues were rescheduled for a different time or format (e.g. next session, homework).
- Length of session was as scheduled ***or*** there was clear clinical justification for deviating from planned schedule.

[1] Examples of clinical justification for running over / under time = significant patient distress, expression of risk etc.

6.3. Maintained Focus

Ability to maintain focus on important issues, whilst demonstrating appropriate flexibility in response to unanticipated issues.

1. Limited

Absence of skill or an ***inappropriate performance***:

- Session lacked any clear focus and appeared aimless (e.g. patient dominated session completely and therapist simply followed the patient's lead- making little or no attempt to focus session back to appropriate topics, general unstructured 'chat' which did not appear to have any specified aims or reach any particular conclusions).
- ***Or*** overly rigid focus was unhelpful to the patient (e.g. therapist dismissed potentially important issues raised by the patient or disregarded patient's distress because it did not relate to the agenda / fit with the session plan).

2. Basic

Major substantive problems:

- Therapist showed limited skill in maintaining focus on appropriate topics, resulting in frequent repetitions, unproductive digressions or unjustified discussions of peripheral issues (e.g. therapist frequently switched topics without clear justification, was slow to respond to digressions, was abrupt or tactless when re-directing discussion to appropriate topics).
- ***Or*** therapist focused on established priorities in a rigid, inflexible or mechanistic manner (e.g. failed to acknowledge or respond to unanticipated issues).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies*** in the following areas:

- Therapist largely maintained focus on session priorities (i.e. agenda items). However, if repetitions, unproductive digressions or unnecessary discussions of peripheral issues arose, these were re-directed back to an appropriate topic in a timely manner and with some degree of sensitivity.
- ***And/or*** if unanticipated issues arose, therapist acknowledged these and showed appropriate flexibility (e.g. allowing the session to diverge from planned topics to make maximum therapeutic progress or maintain client engagement / therapeutic relationship).

4. Advanced

Consistently high level of skill in the following areas:

- Therapist consistently maintained focus on session priorities (i.e. agenda items). However, if repetitions, unproductive digressions or unnecessary discussions of peripheral issues arose, these were quickly acknowledged, sensitively evaluated and skilfully steered back to an appropriate topic (e.g. "We seem to be straying from our plan for the session- is this something that it is important that we focus on? Or would it be more useful to return to [planned focus]?").
- ***And*** if unanticipated issues arose, therapist acknowledged these and showed appropriate flexibility. Any decisions to change focus were negotiated openly and explicitly with the patient.

Domain 7 - Fostering Therapeutic Relationship

7.1. Interpersonal style

Ability to embody a positive interpersonal style which is congruent with the therapeutic context.

1. Limited

Absence of skill or an ***inappropriate performance***:

- Therapist consistently demonstrated a negative interpersonal style (e.g. impatient, aloof, insincere, patronising, distant, distracted, cold, uncaring, intimidating, reprimanding, hostile, demeaning etc.).
- ***Or*** therapist's interpersonal style was poorly suited to the patient's therapeutic context [1] (e.g. therapist praised patient excessively, demonstrated high level of warmth to a highly avoidant / socially anxious patient thus making them feel uncomfortable).

2. Basic

Major substantive problems:

- Therapist generally demonstrated a reasonable interpersonal style (e.g. was friendly, positive, constructive etc.) which was appropriate given the patient's therapeutic context. However, at times therapist showed limited flexibility in adapting their interpersonal style in response to therapeutic context (e.g. insensitive use of humour which invalidated the patient's feelings, inappropriate self-disclosure, inappropriate disclosure of ideas about the patient's problems [e.g. told patient with eating disorder idea for final weight too early in treatment]).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- Therapist generally demonstrated a positive interpersonal style (e.g. open, warm, respectful, genuine etc.) which was suited to individual patient's therapeutic context (e.g. friendly but not overly familiar, optimistic without being excessively positive).

4. Advanced

Consistently high level of skill:

- Therapist consistently demonstrated a positive interpersonal style (e.g. embodied and effectively communicated sincerity, open-minded curiosity, genuineness, honesty, sensitive concern, optimism, professionalism, encouragement, unconditional positive regard etc.) ***and*** seamlessly adapted their interpersonal style so that it was congruent with the therapeutic context (e.g. optimal levels of warmth, concern etc. given the individual patient's current needs).

[1] Therapeutic context = patient's interpersonal style, presenting problems, history, stage of treatment, formulation & presentation.

7.2. Empathic Understanding

Ability to accurately grasp the content and emotional tone of the patient's viewpoint (i.e. their understanding of themselves and the world around them) and to sensitively and appropriately conveying this understanding.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist frequently projected their own attitudes, conventional attitudes, or attitudes derived from particular theoretical systems onto the patient.
- Therapist repeatedly ignored or completely misinterpreted obvious elements of the patient's viewpoint.
- Therapist negated, undermined or was critical of the patient's viewpoint (e.g. appeared disapproving, disparaging or judgemental) or over-identified with patient's perspective (i.e. was not objective).
- Therapist seemed unwilling or unable to revise their understanding of the patient's viewpoint in response to new clinical information or in-session changes.

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist accurately reflected more obvious, verbal cues (i.e. what patient explicitly said), but repeatedly ignored or completely misinterpreted more subtle emotional or nonverbal cues (e.g. tone of voice, body language etc.) ***and/or*** responded inappropriately (e.g. excessive empathy).
- Therapist found it difficult or seemed reluctant to revise their understanding of the patient's viewpoint in response to new clinical information or in-session changes.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist picked up on most verbal and more subtle emotional or nonverbal cues to gain an understanding of the patient's viewpoint but at times therapist slightly misinterpreted patient or focused on peripheral aspects of what patient said.
- Therapist attempted to convey their understanding of the patient's viewpoint but with limited skill (e.g. restating patient's words unnecessarily, overly long summaries, unsuitable tone of voice, intellectualising issues).

4. Advanced

Consistently high level of skill in the following areas:

- Therapist consistently picked up on verbal, emotional and nonverbal cues to gain an accurate and intimate understanding of the patient's viewpoint.
- Therapist consistently used appropriate verbal (e.g. paraphrasing, summarising, reflecting) and non-verbal responses (e.g. tone of voice, facial expression) to convey their understanding of the patient's viewpoint. This helped the patient feel listened to and understood on an emotional and cognitive level throughout the session (e.g. therapist conveyed the importance of working on issues which were important to the patient when prioritising agenda items, acknowledged participants bravery in undertaking anxiety provoking tasks etc.)

7.3. Collaboration

Ability to encourage the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation.

1. Limited

Absence of skill or an ***inappropriate performance:***

- Therapist repeatedly made very little or no attempt to work with or involve the patient (e.g. therapist appeared to come to the session with pre-planned ideas that were not open to discussion, focused on issues that were not a priority for the patient, told the patient what to do without requesting any input, ignored or disregarded issues raised by the patient, did not elicit or value the patient's ideas, did not share information with the patient, tried to control or dominate the patient, seemed to be 'cross-examining' or arguing with the patient).
- ***Or*** therapist was frequently too passive and offered the patient too much responsibility (e.g. provided no input where the patient was not able to do so independently, inappropriately followed the patient's lead).

2. Basic

Major substantive problems:

- Therapist made some attempt to involve and work with the patient. However, on a number of occasions, the therapist offered the patient too much or too little responsibility given the stage of therapy or patient's presentation [1] (e.g. therapist expected the patient to take the lead in very early sessions, did not give the patient enough choice, paid minimal attention to the patient's ideas, did not give the patient sufficient space and time to think, used overly complicated or vague questions such as "What shall we talk about today?" or "What next?", lectured the patient, debated with or tried to persuade the patient).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- For most of the session, the therapist worked effectively with the patient and helped them to share responsibility in a manner appropriate to the stage of therapy and patient's presentation (e.g. therapist offered the patient choices, paid attention to the patient's ideas, gave the patient some time and space to think, asked relevant questions such as "What do you think the most important things are for us to cover today?", "How might we test that out?"). However, there were one or two occasions when the therapist and patient did not appear to be working as a team (e.g. therapist was too instructive or passive).

4. Advanced

Consistently high level of skill:

- Therapist consistently encouraged the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation (e.g. therapist offered the patient appropriate choices, used skilful questioning to elicit and incorporated the patient's ideas, gave the patient plenty of time and space to think, asked skilful questions such as "How might you take this forward next week?", "What's the most important thing you need to do to keep the progress going?", "You have mentioned X, Y and Z. What do you think is most important to cover today, given that we are nearing the end of therapy?").

[1] NB it may be appropriate for therapists to take more responsibility in initial sessions, but responsibility should be increasingly transferred to the patient as therapy progresses, depending on patient's presentation.

Domain 8 - Effective Two-way Communication

8.1. Patient Feedback

Ability to elicit, explore and respond to feedback about the patient's understanding of and reaction to all aspects of session.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to check the patient's understanding of and reaction to the session ***or*** discouraged the patient from giving feedback.
- Therapist repeatedly ignored or disregarded patient feedback ***or*** responded in a negative manner (e.g. became defensive or argumentative).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist inconsistently checked the patient's understanding of and reaction to the session, meaning that feedback was not elicited for a number of important aspects of the session (e.g. formulation, homework, interventions etc.).
- Therapist used overly complicated or vague questions to elicit patient feedback (e.g. asked "Is that OK?", "Does everything seem to be going well?").
- Therapist failed to ask for patient feedback about any negative reactions to session, despite clear need to do so (e.g. patient's comments indicated potential concerns).
- Therapist acknowledged patient feedback but made little or no attempt to discuss or respond to it.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist regularly checked the patient's understanding of and reaction to the session. However, feedback was not elicited for one or two important aspects of the session (e.g. formulation, homework, interventions etc.).
- Therapist used some appropriate questions to elicit patient feedback (e.g. asked "How do you think the session is going so far", "Does this homework seem manageable?", "Is the homework clear?", "How does the idea of doing X strike you?", "How do you feel about attempting X?").
- Therapist made some attempt to elicit potentially negative reactions to session (e.g. doubts, scepticism or concerns) as well as positive reactions (e.g. asked "Do you have any concerns about X", "You seem a little uncertain, is there anything you feel unsure about?").
- Therapist discussed patient feedback and made some attempt to respond to patient's feedback (e.g. answered questions, revised focus in response to patient's concerns).

4. Advanced

Consistently high level of skill in the following areas:

- Therapist consistently checked the patient's understanding of and reaction to all aspects of session (e.g. formulation, homework, interventions, agenda, measuring change, patient's response to therapist etc.).
- Therapist consistently used skilful questions to elicit patient feedback (e.g. asked "From your point of view, why is it important to do X?", "What might you learn from doing X?").
- Therapist actively encouraged the patient to express positive and negative reactions to all aspects of session in a warm and friendly manner which invited exploration (e.g. asked "What has been least helpful today?", "Is there anything I've said that didn't make sense?").
- Therapist thoroughly and openly discussed patient feedback in a non-judgemental, supportive and sensitive manner. ***And*** therapist responded appropriately to patient's feedback (e.g. adapted interventions, adjusted their own behaviour, clearly and fully answered questions).
- If relevant, therapist picked up on subtle cues (e.g. tone of voice, hesitation etc.) to identify difficulties giving authentic feedback (e.g. patient responded in accordance with what they thought therapist wished to hear) and openly discussed these with the patient.

8.2. Reflective Summaries

Ability to work with the patient to reflect upon and summarise salient session content in order to facilitate joint understanding of crucial therapeutic material and consolidate key learning.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist made little or no attempt to encourage the patient to review and reflect on session content.
- Therapist made little or no attempt to summarise important information or provided very minimal, highly inappropriate or inaccurate summaries.

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist prompted the patient to review and reflect on session content ***and/or*** provided summaries either too rarely (e.g. only once or twice) ***or*** too frequently (e.g. continually providing summaries or asking the patient to reflect on session content).
- Reflective summaries were of limited therapeutic benefit as they were superficial ***and/or*** focused on irrelevant information or peripheral learning points.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- At the beginning and end of session, and a number of times within session, therapist prompted the patient to review and reflect on session content (e.g. asked relevant questions such as “what’s the main thing you’d like to remember about this session") ***and*** provided accurate capsule summaries (e.g. “so to recap, at the beginning of the session we discussed X and the impact it had on your mood. Then we talked about Y etc.”).
- Reflective summaries focused on relevant information and learning points and thus were of some use in helping to reinforce understanding of important therapeutic material and consolidate key learning.

4. Advanced

Consistently high level of skill in the following areas:

- At the beginning and end of session, and at appropriately regular intervals within session, therapist used skilful questioning to elicit in-depth patient reflections on and summaries of session content ***and*** provided clear, succinct and meaningful capsule summaries.
- Reflective summaries skilfully drew together salient information and key learning points in a way which helped to reinforce and facilitate joint understanding of crucial therapeutic material, consolidate key learning and enabled the patient to gain new insights and make therapeutic shifts.

Global Performance Rating

How would you rate the therapist's overall performance within this session?

1. Limited

- Therapist demonstrates a significant absence of skill or an inappropriate performance which is likely to have negative therapeutic consequences.

2. Basic

- Therapist's performance is somewhat appropriate with some degree of skill evident.
- However, major substantive problems are evident.

3. Good

- Therapist demonstrates a good degree of skill with no major problems.
- However, minor problems or inconsistencies are evident in the therapist's performance.

4. Advanced

- Therapist consistently demonstrates a high level of skill with only very few and very minor problems.

Patient Complexity Rating

How complex do you feel the patient was to work with? Issues that may influence complexity include level of patient motivation, level of engagement, the presence of social or environmental problems, the number of problems, the severity of problems, and suitability for CBT.

1. Very straightforward

2. Somewhat Straightforward

3. Somewhat Complex

4. Very Complex

References

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. doi: 10.1017/S1352465801004040

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*(2), 218-244. doi: 10.1037/0033-2909.90.2.218

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012a). Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, 1-11. doi: 10.1080/10503307.2012.716528

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

Keen, A., & Freeston, M. (2008). Assessing competence in cognitive-behavioural therapy. *British Journal of Psychiatry, 193*(1), 60-64. doi: 10.1192/bjp.bp.107.038588

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

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognititve and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use (3rd ed.)*. Oxford: Oxford University Press.

Vallis, T., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54*(3), 381-385. doi: 10.1037/0022-006X.54.3.381

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), 620-630. doi: 10.1037/0022-006X.61.4.620

Weck, F., Bohn, C., Ginzburg, D. M., & Ulrich, S. (2011). Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire sessions. *Psychotherapy Research, 21*(6), 658-669. doi: 10.1080/10503307.2011.602751

Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *Journal of Nervous and Mental Disease, 199*(4), 276-279. doi: 10.1097/NMD.0b013e3182124617

Young, J., & Beck, A. T. (1980). Cognitive Therapy Scale Rating Manual. Retrieved from www.beckinstitute.org 16.10.2013

Assessment of Core CBT Skills (ACCS)

www.accs-scale.co.uk

FEEDBACK FORM

Final Version (3) - July 2014

Scale contributors: Kate Muse, Freda McManus, Sarah Rakovshik & Helen Kennerley

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus and was conducted in affiliation with the University of Oxford's Department of Psychiatry and the Oxford Cognitive Therapy Centre (OCTC).

With special thanks to those who participated in the development and evaluation of the ACCS, particularly staff at OCTC, First Step, University of Derby CBT course, and South of Scotland CBT course.

Brief Rating Guidelines

Using the ACCS manual

It is recommended that assessors spend some time reading and familiarising themselves with the ACCS manual before conducting an assessment. Initially, the manual should be consulted on an item by item basis alongside completing the feedback form. Once the assessor becomes familiar with the scale, the manual can then be used as a frame of reference which can be consulted whenever questions or uncertainties arise. It is also recommended that assessors periodically re-review the manual to prevent assessor-drift.

Numeric rating

For each item assign a rating between 1 and 4. If the therapist is assessed as falling between two descriptors, the assessor can select a ½ mark (i.e. 1.5, 2.5, or 3.5). The overarching definitions of the scale are as follows:

Performance Band	Generic Definition of Performance Band
1. Limited	• Therapist fails to include feature outlined. • ***Or*** therapist demonstrates a significant absence of skill or an inappropriate performance which is likely to have negative therapeutic consequences.
2. Basic	• Therapist's performance is somewhat appropriate with some degree of skill evident. • However, major substantive problems are evident.
3. Good	• Therapist demonstrates a good degree of skill with no major problems. • However, minor problems or inconsistencies are evident in the therapist's performance.
4. Advanced	• Therapist consistently demonstrates a high level of skill with only very few and very minor problems.

Please refer to the manual for guidance as to the type of performance which is consistent with each point on the rating scale for each individual item.

Qualitative feedback

For each domain, draw on the exemplar therapist behaviours provided and the session material being assessed to provide examples of what made some aspects of the therapist's performance particularly successful what made some aspects of the therapist's performance less successful. Suggest alternative, more skilful way of working within any weaker domains and highlight ways in which the therapist could further develop their skills in this area.

Assessment Materials

As well as a recording of a mid-treatment CBT session, **please also use the ACCS submission cover sheet** which provides additional contextual information which will help you make an informed rating of the therapist's performance.

Therapist's Name: _____ **Assessor's Name:** _____

Date of Submission: _____

Domain 1 - Agenda Setting

See user manual page 7 – 8

1.1. Suitable Items: Ability to help the patient identify and prioritise specific, relevant and appropriate agenda items.

1.2. Feasible Agenda: Ability to set an agenda which is realistic and feasible given the time available.

Key strengths and learning needs:

Domain 2 - Formulation

See user manual page 9

2.1. Coherent and dynamic formulation: Ability to develop a clear formulation which draws upon appropriate evidence-based theory to offer a concise, comprehensive and personalised explanation of relevant history, triggers and maintaining features of the patient's problems.

Note: This item should be based on any written or diagrammatic formulation materials in addition to the session recording.

Key strengths and learning needs:

Domain 3 – CBT Interventions

See user manual page 10 - 14

3.1. Appropriate Intervention Targets: Ability to skilfully define, clarify and specify intervention targets which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were highly likely to be maintaining problems.	
3.2. Choosing Suitable Interventions: Ability to select cognitive-behavioural interventions which form part of a logical, coherent and unified treatment strategy which is likely to bring about therapeutic change in the treatment target(s) and is suited to the patient's therapeutic context. This selection was accurately guided by appropriate theory-based practice or practice based on evidence when possible.	
3.3. Rationale for Interventions: Ability to facilitate the patient's understanding of the importance and potential benefits of interventions.	
3.4. Implementing Interventions: Ability to systematically implement intervention(s) in a fluent and articulate manner. To be sensitive and responsive to the therapeutic context and provide optimal levels of support, encouragement and praise.	
3.5. Reviewing Interventions: Ability to conduct a comprehensive review of the results of interventions (whether positive or negative) in order to help the patient identify what they learned from the experience.	

Key strengths and learning needs:

Domain 4 – Homework

See user manual page 15 - 18

4.1. Reviewing Homework: Ability to conduct a comprehensive review of previous homework (whether completed or not) in order to help the patient identify what they learned from the experience.	
4.2. Choosing Suitable Homework: Ability to plan homework which is tailored to the therapeutic context and builds upon session material or previous homework.	
4.3. Rationale for Homework: Ability to facilitate the patient's understanding of the importance and potential benefits of homework.	
4.4. Planning Homework: Ability to work with the patient to ensure they have a clear and detailed understanding of the homework task(s).	

Key strengths and learning needs:

Domain 5 – Appropriate Tracking of Progress

See user manual page 19 - 20

5.1. Choosing Suitable Measures: Ability to select appropriate, clinically relevant standardised **and/or** idiosyncratic methods for measuring change in symptoms and associated features (beliefs, behaviours, feelings) and movement towards goals.

5.2. Implementing Measures: Ability to administer measures at suitable time points across and within session and to skilfully review, interpret and respond to the information gleaned.

Note: This item should be based on any relevant written materials (e.g. completed questionnaires, diaries, belief ratings, treatment goals, etc.) in addition to the session recording.

Key strengths and learning needs:

Domain 6 – Effective Use of Time

See user manual page 21 - 23

6.1. Pace: Ability to pace the session in a manner which is well suited to the therapeutic context and patient's capacity for learning.

6.2. Time Management: Ability to manage time within the session in a balanced and efficient manner.

6.3. Maintained Focus: Ability to maintain focus on important issues, whilst demonstrating appropriate flexibility in response to unanticipated issues.

Key strengths and learning needs:

Domain 7 – Fostering Therapeutic Relationship

See user manual page 24 - 26

7.1. Interpersonal style: Ability to embody a positive interpersonal style which is congruent with the therapeutic context.

7.2. Empathic Understanding: Ability to accurately grasp the content and emotional tone of the patient's viewpoint (i.e. their understanding of themselves and the world around them) and to sensitively and appropriately conveying this understanding.

7.3. Collaboration: Ability to encourage the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation.

Key strengths and learning needs:

Domain 8 – Effective Two-way Communication

See user manual page 27 - 28

8.1. Patient Feedback: Ability to elicit, explore and respond to feedback about the patient's understanding of and reaction to all aspects of session.

8.2. Reflective Summaries: Ability to work with the patient to reflect upon and summarise salient session content in order to facilitate joint understanding of crucial therapeutic material and consolidate key learning.

Key strengths and learning needs:

Overall Session Score

Total Score (*add the scores assigned for each item*):

	Range = 22 to 88

To aid in interpretation, the total score a therapist would achieve if their performance was consistently rated as 1, 2, 3, or 4 is: One-limited = total score of 22, Two-basic = total score of 44, Three-good = total score of 66, Four-advanced = total score of 88.

Average item score *(total score divided by 22):*

Global Performance Rating

Therapist's overall performance within the session was:

1. Limited **2. Basic** **3. Good** **4. Advanced**

If you believe the therapist falls between two of the descriptors, select a ½ mark (1.5, 2.5, or 3.5).

Patient Complexity Rating

How complex do you feel the patient was to work with?

1. **Very Straightforward** **2. Somewhat Straightforward** **3. Somewhat Complex** **4. Very Complex**

Skills Profile

Therapist's Key Strengths

What were the therapist's key strengths within the session recording? How can the therapist continue to build upon these strengths in the future?

Therapist's Learning Needs

Based on the session recording, what skills does the therapist need to concentrate on further improving and developing? How could the therapist further develop their skills in these areas?

ACCS Submission Cover Sheet

Therapist's Name: _____

Date of Submission: _____

Assessor's Name: _____

Supporting Material. The recording should be accompanied by any materials completed in the session or for homework (e.g. diaries, thought records etc.) and any completed questionnaires.

Formulation. Please attach a copy of the formulation (no more than one A4 sheet). The formulation should demonstrate the therapist's understanding - at that point in time - of how the patient's problems developed (background experiences, beliefs and assumptions, critical incidents) and what is maintaining them (vicious circles).

Session Summary:

- Stage of therapy / session number

- Patient's presenting problem(s)

- Patient's treatment goals

- Agenda

- Homework review (What were the assignments? Outcome?)

- Main topic(s) – What were the main things you worked on? What did you do (e.g. worked through problem on thought record, carried out behavioural experiment, looked for alternative explanations)? What is the rationale for this work?

- New homework (What are the assignments? What is the rationale for them?)

- Patient's feedback on the session (reactions, take home messages)

- Did you use any standardised questionnaires and / or idiosyncratic measures (e.g. frequency counts, diaries etc.) to assess change within the treatment session (e.g. to measure the impact of cognitive-behavioural interventions) or across treatment (i.e. to measure change in symptoms and movement towards goals over the course of treatment as a whole)?

Appendix 4. Text relating to Socratic Enquiry.

3.5. Reviewing Interventions

Ability to conduct a comprehensive review of the results of interventions (whether positive or negative) in order to help the patient identify what they learned from the experience.

1. Limited

- Therapist did not implement any CBT interventions.

Or absence of skill or an ***inappropriate performance***:

- Therapist failed to review ***or*** conducted a brief or cursory review of the results of interventions which did not help the patient identify learning (e.g. asked "how do you think that went?" with no further follow up).

2. Basic

Major substantive problems:

- Therapist reviewed the results of interventions and made some attempt to highlight learning implications, but with limited skill (e.g. struggled to relate the results back to the relevant cognition, behaviours or emotions that they were designed to target, did not address negative results, simply told the patient what they had learned from the experience).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- Therapist adequately reviewed the results of interventions (whether positive or negative). This review helped the patient independently draw conclusions about the learning implications (e.g. therapist used curious and Socratic questions to examine outcomes in relation to prior predictions).

4. Advanced

Consistently high level of skill:

- Therapist skilfully engaged the patient in a comprehensive review of the results of interventions (whether positive or negative) which helped the patient to independently draw conclusions about useful and relevant learning implications. Therapist helped the patient link this learning to the relevant cognition, behaviours or emotions they were designed to target ***and*** to their formulation / treatment goals (e.g. used curious and Socratic questions to help the patient re-evaluate previous conclusions or construct new ideas).

ACCS Manual

Domain 4 - Homework

4.1. Reviewing Homework

Ability to conduct a comprehensive review of previous homework (whether completed or not) in order to help the patient identify what they learned from the experience.

1. Limited

Absence of skill or an ***inappropriate performance*** in one or more of the following areas:

- Therapist failed to review ***or*** conducted a brief or cursory homework review which did not help the patient identify learning (e.g. asked "how did you get on with the homework?" with no further follow up).
- Therapist responded to completed or uncompleted homework in an unhelpful or negative manner (e.g. scolding, judgemental, dismissed homework as unimportant).

2. Basic

Major substantive problems in one or more of the following areas:

- Therapist reviewed homework and made some attempt to highlight learning implications from completing (or not completing) homework, but with limited skill (e.g. learning had limited relevance to the patient's problems, therapist simply told the patient what they had learned from the experience).
- If relevant, therapist made some attempt to reflect on why non-completion occurred but failed to identify relevant emotional, cognitive or practical difficulties ***and/or*** did not identify ways to overcome future 'blocks' to completion.
- Completed homework was given limited recognition ***and/or*** therapist responded to non-completion with limited sensitivity and understanding.

3. Good

Good degree of skill in the following areas, with only ***minor problems or inconsistencies***:

- Therapist adequately reviewed homework. This review helped the patient to independently draw conclusions about the learning implications from completing (or not completing) homework (e.g. therapist used curious and Socratic questions to examine outcomes in relation to prior predictions).
- If relevant, therapist helped the patient identify relevant emotional, cognitive or practical reasons for non-completion ***and*** discussed ways to overcome future 'blocks' to completion (e.g. asked "how could you overcome xx next time?").
- Therapist responded to completed homework with appropriate encouragement ***and*** to non-completion in a

4. Advanced

Consistently high level of skill in the following areas:

- Therapist skilfully engaged the patient in a comprehensive homework review. This review helped the patient to independently draw conclusions about useful and relevant learning implications from completing (or not completing) homework and to link this learning to their formulation / treatment goals (e.g. therapist helped them to explore what it tells them about how best to understand or make progress with their difficulties).
- If relevant, therapist worked with the patient to identify relevant emotional, cognitive or practical reasons for non-completion ***and*** helped them identify specific ways to overcome future 'blocks' to completion (e.g. using a calendar for a patient who has difficulty remembering to complete homework, using problem solving to break homework down into less threatening components).
- Therapist responded to completed homework with optimal levels of encouragement and praise ***and*** to non-completion in a highly supportive, sensitive and open manner.

ACCS Manual

7.3. Collaboration

Ability to encourage the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation.

1. Limited

Absence of skill or an ***inappropriate performance:***

- Therapist repeatedly made very little or no attempt to work with or involve the patient (e.g. therapist appeared to come to the session with pre-planned ideas that were not open to discussion, focused on issues that were not a priority for the patient, told the patient what to do without requesting any input, ignored or disregarded issues raised by the patient, did not elicit or value the patient's ideas, did not share information with the patient, tried to control or dominate the patient, seemed to be 'cross-examining' or arguing with the patient).
- ***Or*** therapist was frequently too passive and offered the patient too much responsibility (e.g. provided no input where the patient was not able to do so independently, inappropriately followed the patient's lead).

2. Basic

Major substantive problems:

- Therapist made some attempt to involve and work with the patient. However, on a number of occasions, the therapist offered the patient too much or too little responsibility given the stage of therapy or patient's presentation [1] (e.g. therapist expected the patient to take the lead in very early sessions, did not give the patient enough choice, paid minimal attention to the patient's ideas, did not give the patient sufficient space and time to think, used overly complicated or vague questions such as "What shall we talk about today?" or "What next?", lectured the patient, debated with or tried to persuade the patient).

3. Good

Good degree of skill, with only ***minor problems or inconsistencies***:

- For most of the session, the therapist worked effectively with the patient and helped them to share responsibility in a manner appropriate to the stage of therapy and patient's presentation (e.g. therapist offered the patient choices, paid attention to the patient's ideas, gave the patient some time and space to think, asked relevant questions such as "What do you think the most important things are for us to cover today?", "How might we test that out?"). However, there were one or two occasions when the therapist and patient did not appear to be working as a team (e.g. therapist was too instructive or passive).

4. Advanced

Consistently high level of skill:

- Therapist consistently encouraged the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation (e.g. therapist offered the patient appropriate choices, used skilful questioning to elicit and incorporated the patient's ideas, gave the patient plenty of time and space to think, asked skilful questions such as "How might you take this forward next week?", "What's the most important thing you need to do to keep the progress going?", "You have mentioned X, Y and Z. What do you think is most important to cover today, given that we are nearing the end of therapy?").

[1] NB it may be appropriate for therapists to take more responsibility in initial sessions, but responsibility should be increasingly transferred to the patient as therapy progresses, depending on patient's presentation.

CHAPTER SEVEN:

Is assessor training necessary? Examining the impact of training on usability and inter-rater reliability of the Assessment of Core CBT Skills (ACCS) rating scale

Chapter Overview

The Assessment of Core CBT Skills (ACCS) is a novel rating scale for assessing the competence with which therapists deliver Cognitive Behavioural Therapy (CBT). This study examines the impact of a one-day assessor training workshop in how to use the ACCS in order to determine whether training is required for assessors to reliably use the ACCS scale, or whether use of the ACCS manual alone is sufficient. Therapists (*N* = 40) from a routine practice CBT service rated two role-play treatment sessions: one before training (using the ACCS manual alone) and one after training. The agreement between therapists' ratings on the ACCS (i.e. inter-rater reliability) and therapist's usability ratings for the ACCS were assessed both before and after training. Results show an improvement in both inter-rater agreement and perceived usability of the ACCS following training. However, both inter-rater agreement and usability ratings were already acceptable pre-training (i.e. with use of the ACCS manual alone). Thus, whilst training does seem to be beneficial, simply providing assessors with the ACCS manual may be sufficient to achieve acceptable levels of reliability and usability.

Acknowledgements:

With many thanks to practitioners within the First Step service in Cumbria for participating in the study and to Richard Thwaites, Laura Lockhart, David Sandford, and Ryan Askey-Jones for supporting this research within their service.

Contents

Introduction 483
Method 485
Procedure 485
Participants 486
Materials 486
Training Package 486
ACCS Scale 487
Role-plays 488
ACCS Usability Questionnaire 488
Satisfaction with Workshop 489
Results 489
Descriptive Statistics 489
Inter-Rater Reliability 489
Variance in scores. 490
Score discrepancies. 492
Absolute agreement. 493
Usability 495
Satisfaction with Workshop 496
Discussion 497
Summary of Results 497
Implications for the ACCS Scale 499
Quality of Assessor Training 500
Strengths and Limitations 501
Future Research 502
References 504
Appendicies 507

Appendix 1 – Hand-out of PowerPoint slides for ACCS training 507

Appendix 2: Numerical ACCS Feedback form used to rate the therapist's performance within the role-play treatment sessions. 516

Appendix 3: Submission cover sheets provided for participants alongside the role-plays 520

Appendix 4: Feedback questionnaire to determine the usability of the ACCS 526

Appendix 5: Workshop feedback questionnaire 527

Appendix 6: Qualitative feedback about the ACCS scale 528

Appendix 7: Qualitative feedback on the workshop 534

Introduction

Inter-rater reliability (IRR) is the level of agreement between different assessors' ratings on an observational measure (Barker, Pistrang, & Elliot, 2002). Achieving adequate IRR is important for rating scales used to assess therapists' competence in delivering cognitive behavioural therapy (CBT) because these assessments are used to determine whether therapists have reached levels of competence that warrant receipt of formal qualifications and to make decisions about therapists' fitness to practice. However, it has proved difficult to achieve adequate IRR on the most widely used assessor-rated CBT competence scales, with a number of studies reporting only poor or moderate agreement for both total scores and a number of individual items on the cognitive therapy scale (CTS: www.beckinstitute.org) and revised cognitive therapy scale (CTS-R: Blackburn et al., 2001).

A number of reasons have been suggested as contributing to difficulties achieving adequate IRR for CBT competence rating scales. These include difficulties defining and operationalising CBT competence, differences in assessors' understandings of CBT competence, differences in how much attention assessors pay to different aspects of competence, high degrees of assessor inference required to rate scale items, and inadequate definition of scale points (Barber, Sharpless, Klostermann, & McCarthy, 2007; Blackburn et al., 2001; Whisman, 1993). Findings from the qualitative study examining experts' experiences of assessing CBT competence (see Chapter two) also highlight a number of these issues. In particular, participants commented that one of the main difficulties faced when assessing CBT competence was that defining and operationalising the specific skills involved in the competent delivery of CBT is far from straightforward and that there is often a lack of agreement within the field about what CBT is, the way it should be applied and therefore what assessors consider competent.

The Assessment of Core CBT Skills (ACCS) was designed to improve upon currently available CBT competence rating scales (in particular the CTS and CTS-R) in terms of inter-rater reliability. Specifically, the ACCS sought to give more detailed rating guidance in the scale manual, provide clearer operational definitions of each scale item, offer more detailed item-specific guidance about the specific behavioural indicators associated with each point on the rating scale, and ensure that items related to discrete aspects of competence (rather than assessing multiple aspects of CBT competence). It was hoped that these revisions would increase standardisation and objectivity of measurement of CBT competence and thus overcome some of the barriers to achieving adequate IRR faced by previous scales.

For these efforts to improve IRR to be effective, it is vital that assessors use the manual accompanying the ACCS feedback form. The ACCS manual provides guidance for assessors as to how to make judgements about the quality or skilfulness of therapists' performance using the rating scale. As well as general guidelines, the manual provides clarification about the constructs measured by each item and details item-specific examples of behaviours consistent with each scale point. It is recommended that assessors spend some time reading and familiarising themselves with the manual before conducting an assessment. Initially, the manual should be consulted on an item by item basis alongside completing the feedback form. Only once the assessor becomes familiar with the scale can the manual then be used as a frame of reference that can be consulted whenever questions or uncertainties arise. In addition to this, it is also recommended that assessors periodically re-review the manual to prevent assessor-drift. Despite this guidance, the focus group in chapter six identified that, due to time pressures, many assessors from the Oxford Cognitive Therapy Centre (OCTC) were unable to find the time to read the manual in any depth and often completed ratings without consulting the manual at all. This is problematic as it means that assessors are effectively providing ratings on the basis of the overarching anchor definitions alone (i.e. 1- limited, 2- basic, 3- good, 4- advanced), rather than the item-specific behavioural definitions. Given the ever increasing workload of therapists working in routine practice and training settings, it is unlikely that this problem would be unique to OCTC. In addition, even if the manual is thoroughly read and referred to, this does not provide assessors with the opportunity to practice using the ACCS rating scale or to ask questions or clarify their understanding of the scale. Thus, simply providing a copy of the ACCS training manual alone may not be sufficient to ensure adequate IRR.

Another method of improving IRR is to provide training packages which help to standardise assessors' interpretations of scale items, teach assessors the specific performance indicators associated with each point on a rating scale, and provide assessors with practice and thus improve familiarity with the rating scale (Barber et al., 2007). Assessor training packages have been found to improve IRR for a range of clinician-rated scales (Guar, Kaviani, Bansal, & Lee, 2010; Kobak, Alan, & lipstiz, 2005; Kobak, Engelhardt, & Lipstiz, 2006; Kobak, Opler, & Engelhardt, 2007; Muller et al., 1998; Rosen et al., 2008; Tracy, Adler, Rotrosen, Edson, & Lavori, 1997) and clinical performance assessments (Bachrach, Mitntz, & Lurborsky, 1971; Lin et al., 2013), as well as CBT competence rating scales (Gordon, 2006; Reichelt et al., 2003). Although attending training is time-consuming, there are perhaps fewer barriers to attending training than there are to finding time in a busy schedule to read a rating manual (this is a task which is likely to remain on the bottom of one's to-do list, being replaced by more urgent tasks until it is ultimately

forgotten). Within many CBT training courses it is also a mandatory requirement for therapists working in a supervisory capacity to attend training in assessing therapist competence. Furthermore, both the expert CBT therapists who participated in the qualitative study and the focus group (see chapters two and six respectively) felt that assessing CBT competence is a difficult and complex skill that needs to be fostered through experiential assessor training. However, the provision of assessor training is time consuming, costly, and involves practical difficulties (e.g. finding appropriate venues, availability of experienced trainers with knowledge of the rating scale, arranging for multiple assessors to have the same day off work etc.). The need for using trained assessors, therefore, has implications for the feasibility and scalability of the ACCS scale. Thus, it is important to establish whether the provision of assessor training in how to use the ACCS scale is necessary in order to achieve adequate IRR.

This study aimed to develop and examine the impact of an assessor training package in order to determine whether training is required for assessors to reliably use the ACCS scale, or whether use of the ACCS manual alone is sufficient. It was predicted that participating in the training session would result in improved agreement between different assessors' ratings on the ACCS (i.e. improved IRR) compared to use of the manual ACCS alone. It was also predicted that the training would improve assessors' understanding of how to use the ACCS scale compared to use of the manual ACCS alone, therefore leading to an increase in usability ratings (i.e. the clarity of the items in the ACCS, how easy the ACCS was to use, and overall satisfaction with the ACCS). A secondary aim of the study was to examine assessors' reactions to the training, as this is not only a useful first step in evaluating the impact of training, but can also provide useful information about how best to improve future training efforts (Decker, Jameson, & Naugle, 2011; Kirkpatrick & Kirkpatrick, 2006).

Method

Procedure

This study investigated the impact of a one-day training workshop in how to use the Assessment of Core CBT Skills (ACCS) scale on the usability of the ACCS and the inter-rater reliability of scores on the ACCS. Participants were therapists from an NHS CBT treatment service. All participants were provided with a copy of the ACCS manual prior to training and were asked to spend approximately one hour reading and familiarising themselves with the manual prior to the training. None had previously used the ACCS to rate their own or other therapists' in-session performance. Participants completed ACCS ratings for two different role-played therapy sessions;

one before and one after taking part in the training. Participants were asked to work alone to complete the ACCS ratings. A short usability questionnaire was completed immediately following the pre- and post-training ratings. At the end of the workshop, participants also completed a brief questionnaire assessing their satisfaction with the training.

Participants

The workshop was attended by 40 CBT therapists from the First Step service in Cumbria, an NHS Increasing Access to Psychological Therapies treatment service for people with mild to moderate depression and anxiety (see http://www.cumbriapartnership.nhs.uk/first-step.htm). Participants were predominantly female (80%) and had a mean age of 46.21 years (*SD* = 9.32). All participants were CBT therapists (accredited by the British Association of Behavioural and Cognitive Psychotherapies) and all had completed a higher education qualification in CBT (36 completed a PGDip in CBT, one completed an MSc in CBT, and three completed a BSc in CBT). Participants had been practicing CBT for between one and 15 years (*M* = 6.02, *SD* = 3.34). The majority of participants (*n* = 32, 80%) had treated over 200 CBT cases, with seven (17.5%) having treated 50 – 200 cases and one (2.5%) having treated under 50 cases. Participants were all familiar with and had previously used the scales upon which the ACCS were based: the Cognitive Therapy Scale (CTS: www.beckinstitute.org) and/or the Cognitive Therapy Scale-Revised (CTSR: Blackburn et al., 2001) to rate their own/others' performance.

Materials

Training Package

Training in how to use the ACCS scale comprised of a one-day (six hour) workshop. The training was designed by Kate Muse and was co-delivered by Kate Muse and Dr Freda McManus. The training package aimed to (1) familiarise assessors with the ACCS scale and manual, (2) standardise 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 item, and (3) provide assessors with an opportunity to practice conducting ratings and compare these to standardised ratings.

The training comprised of a number of brief didactic presentations outlining the background to the development of the ACCS scale, the aims of the ACCS, what the ACCS assesses, who should use the ACCS, and finally a four-step guide to using the ACCS. Participants also had the opportunity to practice using the ACCS. This involved first watching a role-play of a treatment session and then rating the skilfulness of the therapists' performance within the session using the

ACCS. Self-reflection exercises and facilitated group discussions were then used to (1) encourage participants to reflect on *why* they had assigned the ACCS scores they had chosen for the therapist in the role-play, (2) identify whether any trainees in the group had provided different scores and, if so, *why* this had occurred, and (3) to highlight any problems or uncertainties in relation to using the ACCS. Larger group discussions were then used to address any questions or uncertainties identified by participants. The degree to which participants agreed on the rating assigned for each ACCS item was also discussed, with particular attention being paid to any items where large discrepancies occurred. A PowerPoint presentation was used throughout the workshop and participants were given a hand-out of the slides (see Appendix one).

ACCS Scale

The Assessment of Core CBT Skills (ACCS: See chapters three and four[1]) is a 22-item rating scale, covering eight domains: agenda setting, formulation, CBT interventions, homework, effective two-way communication, managing therapeutic relationship, effective use of time, and measuring change. All items are rated on a four-point scale (1- limited, 2 - basic, 3 - good, 4 - advanced), with item-specific behavioural descriptions provided for each anchor. A total domain score is provided by calculating the average score of items within the domain and each total domain score is then added to provide an overall session score (ranging from eight to 32), with higher scores representing a higher level of skill.

ACCS Manual

Participants were emailed a copy of the ACCS manual (see Appendix seven Chapter four: *page 265*) and were asked to read this over prior to attending the training. The manual was also used throughout the day to help clarify the description and function of each item. The manual provides guidance for assessors as to how to make judgements about the quality or skilfulness of therapists' performance using the rating scale. As well as general guidelines, the manual provides clarification about the constructs measured by each scale item by detailing item-specific examples of behaviours consistent with each scale point.

[1] The study was completed prior to the evaluation of the ACCS outlined in chapters five and six. Thus, revisions made to the ACCS scale in light of these studies were not included in the materials used in this study. Instead, version 2 of the ACCS manual and feedback form were used.

ACCS Feedback Form

A numerical version of the ACCS feedback form[1] (i.e. a feedback form with space for in-depth formative comments removed) was used by participants to rate the therapist's performance within the role-play sessions (see Appendix two).

Role-plays

Due to patient confidentiality issues it was not possible to use recordings of real treatment sessions, therefore role-played treatment sessions (created by staff at the Oxford Cognitive Therapy Centre [OCTC] for training purposes) were used. Permission was obtained from OCTC and the individuals role-playing the therapists and patients to use these recordings for the purposes of the current study. A supporting information cover sheet was created for each role-play to provide workshop participants with some basic information about the therapeutic context of the role-play session (see Appendix three). Participants read these cover sheets before they watched the role-plays. The first role-play was a mid-treatment, active CBT session for a patient with bulimia and lasted for 35 minutes and 17 seconds. The second role-play was a mid-treatment, active CBT session for a patient with depression and lasted for 55 minutes and 10 seconds. Unfortunately, a technical problem meant that second the role-play would not play past 54 minutes and 34 seconds, meaning that the very end of the session was missing. Participants rated the complexity of the patients in both role-plays on a 4-point scale (1 - very straightforward, 2 - somewhat straightforward, 3 - somewhat complex, 4 - very complex). As the data was not normally distributed, non-parametric Wilcoxon signed-rank tests were performed to examine the difference between the complexity scores assigned for the pre- and post-training role-plays (N = 37 as complexity ratings were missing for three participants). The pre-training role-play (M = 2.41, SD = .65) was rated as significantly more complex than the post-training role-play (M = 1.61, SD = .52), T = 0, p <.001.

ACCS Usability Questionnaire

A seven-item feedback questionnaire was used to determine the usability of the ACCS scale (see Appendix four). Participants used a four-point scale to rate the degree to which the ACCS was relevant to the assessment of CBT competence (1- *not relevant*, 2- *somewhat relevant*, 3- *quite relevant*, 4- *very relevant*), the clarity of the items on the ACCS (1- *not clear*, 2- *somewhat clear*, 3- *quite clear*, 4- *very clear*), the overall style, appearance and layout of the ACCS (1- *poor*, 2- *fair*, 3- *good*, 4- *very good*), how easy they found the ACCS to use (1- *not easy*, 2- *somewhat easy*, 3- *quite easy*, 4- *very easy*), the overall style, appearance and layout of the ACCS (1- *poor*, 2- *fair*, 3- *good*, 4- *very good*), how appropriate they found the scoring system (1- *not appropriate*, 2-

somewhat appropriate, 3- *quite appropriate*, 4- *very appropriate*), and their overall satisfaction with the ACCS (1- *not satisfied*, 2- *somewhat satisfied*, 3- *quite satisfied*, 4- *very satisfied*). Two open-ended questions were also included to identify what participants most and least liked about the ACCS.

Satisfaction with Workshop

To provide insight into participants' reactions to and satisfaction with training, a one-page feedback form was completed anonymously by all participants at the end of the workshop (see Appendix five). Participants used a 5-point scale (1- *not at all*, 3- *moderately*, 5- *extremely*) to rate five aspects of the training day: interesting content, useful content, structure of training, quality of presentation, usefulness and relevance of hand-outs, opportunity to participate, and the degree to which the workshop helped individuals understand how to use the ACCS scale. Three open-ended questions also asked participants to highlight the best things about the workshop, the least satisfactory things about the workshop, and advice about how the workshop could be better run another time.

Results

Descriptive Statistics

The internal consistency of the ACCS was tested by computing Cronbach's alpha coefficient. The results indicated that the ACCS was highly internally consistent, achieving alpha values of .91 pre-training and .89 post-training. The total ACCS scores[2] assigned by the 40 participants ranged from 20.37 to 30.50 (M = 24.92, SD = 3.10) for the pre-training role-play and from 21.00 to 32.00 (M = 26.67, SD = 2.65) for the post-training role-play. Table one outlines the individual item scores assigned by the 40 participants for the pre- and post-training role-play.

Inter-Rater Reliability

The 'gold standard' approach to examining inter-rater reliability (IRR) is to calculate intraclass correlation coefficients (ICCs: Shrout & Fleiss, 1979). However, it was not possible to carry out this analytic strategy as, due to time and resource constraints, only one CBT session was rated pre- and post-training. As there are no clear guidelines regarding the most appropriate statistical approach for examining IRR when only one rating is available, the decision was taken to

[2] Total ACCS scores were calculated by summing the total domain scores (the average scores of items within the domain).

employ three analytic strategies in order to 'triangulate' the results[3]. The results from the three strategies were broadly consistent (demonstrating an increase in IRR from pre- to post-training) and are outlined separately below.

Variance in scores.

Examining the range of scores and the standard deviation (i.e. the spread of scores around the mean) provides an indication of the degree of variation in the scores assigned for each item and domain (see Table one). The standard deviation was low for five out of eight domains (62.50 % < .58) both pre- and post-training. Three domains were highlighted as having high variation in scores. First, there was high degree of variation in the scores participants assigned to domain two (formulation: including one item 'coherent formulation') both pre- and post-training, with the degree of variation being higher post-training (*SD* = .77) compared to pre-training (*SD* = .66). Second, there was high degree of variation in domain five (measuring change), with a large amount of variance in the scores assigned to both of the items in the domain (5.1: 'choosing suitable measures' and 5.2: 'implementing measures') at both pre- and post-training (*SD* for total domain score = .87 both pre- and post-training). Third, there was high degree of variation in the ratings assigned to both items in domain one (agenda setting: 1.1. 'suitable items', 1.2 'feasible agenda') pre-training. However, the spread of scores had reduced post-training (*SD* for total domain score = .64 pre- and .45 post-training).

The difference in the standard deviation for the ACCS scores assigned pre- and post-training was calculated for each item and total domain score (see Table one). For 21 of the 22 items (95.45 %), the standard deviation for the ACCS scores was lower post-training compared to pre-training. A sign test indicated that this number of reductions in variance was highly unlikely to be due to chance (p = <.001). The item which did not reduce in variance pre- to post-training was item two (coherent formulation). A Wilcoxon Signed Rank test revealed that the standard

[3] This analytic approach was developed through consultation with statistical experts with experience of scale development and evaluation. It would have been possible to randomly assign the assessors as 'rater 1' and 'rater 2' and then examine the correlation between raters or calculate ICC values. This analytic approach was taken by Reichelt et al. (2003). However, this study had much higher number of assessors rating each session than this previous study, meaning that there are many more possible combinations of paired raters. Given the resulting higher degree of error, this approach was not taken here. Pbar and total score difference scores were felt to be more appropriate because they allow comparison across all possible pairs of raters and as such provide a much more accurate representation of the data. Pbar is also a more stringent method of calculating inter-rater agreement as it relates to absolute agreement (i.e. it looks for raters who assigned exactly the same score).

deviations for ACCS scores assigned for individual items were significantly smaller post-training compared to pre-training (T = 14.50, p <.001). For six of the eight domains (75.00%), the standard deviation for the ACCS item scores was lower post-training compared to pre-training. A sign test indicated that this number of reductions in variance could have been due to chance (p = .15). The domains which did not reduce in variance pre- to post-training were domain two (formulation) and domain five (measuring change). A Wilcoxon Signed Rank test revealed that the standard deviations for ACCS total domain scores were not significantly smaller post-training compared to pre-training (T = 6.00, p = .09).

Table 1: Individual ACCS item scores for the pre- and post-training role-play.

	Pre-training role-play		**Post-training role-play**		**Pre - Post**
Item / domain[1]	**Mean *(SD)***	**Range**	**Mean *(SD)***	**Range**	***SD* Difference**
1.1	3.13 *(0.69)*	2.00 – 4.00	3.46 *(.49)*	3.00 – 4.00	0.20
1.2	3.12 *(0.68)*	2.00 – 4.00	3.53 *(.49)*	3.00 – 4.00	0.19
Domain 1	3.12 *(0.64)*	2.00 – 4.00	3.49 *(.45)*	3.00 – 4.00	0.19
2.1	3.33 *(0.60)*	2.00 – 4.00	3.08 *(.77)*	2.00 – 4.00	-0.17
Domain 2	3.33 *(0.60)*	2.00 – 4.00	3.08 *(.77)*	2.00 – 4.00	-0.17
3.1	3.22 *(0.59)*	2.00 – 4.00	3.63 *(.53)*	2.00 – 4.00	.06
3.2	3.19 *(0.60)*	2.00 – 4.00	3.82 *(.37)*	3.00 – 4.00	.23
3.3	3.15 *(0.54)*	2.00 – 4.00	3.55 *(.51)*	2.50 – 4.00	.03
3.4	3.18 *(0.51)*	2.00 – 4.00	3.69 *(.45)*	3.00 – 4.00	.06
3.5	3.28 *(0.56)*	2.00 – 4.00	3.54 *(.49)*	3.00 – 4.00	.07
Domain 3	3.21 *(0.37)*	2.30 – 4.00	3.65 *(.33)*	3.00 – 4.00	.04
4.1	3.45 *(0.50)*	3.00 – 4.00	3.88 *(.31)*	3.00 – 4.00	.19
4.2	3.31 *(0.61)*	2.00 – 4.00	3.78 *(.39)*	3.00 – 4.00	.22
4.3	3.24 *(0.56)*	2.00 – 4.00	3.72 *(.46)*	2.50 – 4.00	.10
4.4	2.99 *(0.61)*	2.00 – 4.00	3.63 *(.50)*	2.50 – 4.00	.11
Domain 4	3.25 *(0.43)*	2.50 – 4.00	3.75 *(.34)*	2.88 – 4.00	.09

5.1	2.05 *(0.91)*	1.00 – 4.00	1.74 *(.90)*	1.00 – 4.00	.01
5.2	2.13 *(0.92)*	1.00 – 4.00	1.78 *(.86)*	1.00 – 4.00	.06
Domain 5	2.09 *(0.87)*	1.00 – 3.50	1.76 *(.87)*	1.00 – 4.00	.00
6.1	3.32 *(0.62)*	2.00 – 4.00	3.68 *(.61)*	2.00 – 4.00	.01
6.2	3.21 *(0.69)*	2.00 – 4.00	3.59 *(.52)*	2.00 – 4.00	.17
6.3	3.51 *(0.57)*	2.30 – 4.00	3.76 *(.43)*	3.00 – 4.00	.14
Domain 6	3.34 *(0.55)*	2.10 – 4.00	3.68 *(.44)*	2.33 – 4.00	.11
7.1	3.37 *(0.58)*	2.00 – 4.00	3.72 *(.48)*	2.00 – 4.00	.01
7.2	3.31 *(0.69)*	2.00 – 4.00	3.65 *(.48)*	3.00 – 4.00	.21
7.3	3.38 *(0.54)*	2.00 – 4.00	3.85 *(.31)*	3.00 – 4.00	.23
Domain 7	3.35 *(0.49)*	2.33 – 4.00	3.78 *(.37)*	2.67 – 4.00	.12
8.1	3.22 *(0.62)*	2.00 – 4.00	3.50 *(.58)*	2.00 – 4.00	.04
8.2	3.23 *(0.71)*	2.00 – 4.00	3.54 *(.58)*	2.00 – 4.00	.13
Domain 8	3.22 *(0.58)*	2.00 – 4.00	3.52 *(.54)*	2.00 – 4.00	.04
Global Score	3.22 *(.41)*	2.50 – 4.00	3.64 *(.38)*	3.00 – 4.00	.03

[1] Total domain scores are the average scores of items within the domain. It is worth noting that domain two only includes one item- thus the total domain score for domain two is the same as the item score.

Score discrepancies.

Another way to examine the variability of scores across participants is to look at the discrepancy in the total score assigned by each of the possible pairwise comparisons (Gordon, 2006). Within this study there are 40 participants and thus 780 different pairwise comparisons of the total score. Total ACCS scores were calculated by summing the total domain scores (the average scores of items within the domain) and range from eight to 32 (see Table two). Examining the discrepancy in total scores reveals low inter-rater disagreements pre-training (mean discrepancy = 3.58, *SD* = 2.50 range = 0.00 – 10.13), with only 220 pairings (28.21 %) demonstrating a discrepancy of more than five points. A Wilcoxon Signed Rank test revealed that the level of inter-rater disagreements then fell significantly further post-training (T = 2289.50, p <.001, mean discrepancy = 3.00, *SD* = 2.17, range = 0.00 – 11.00), with only 141 pairings (18.08 %) demonstrating a discrepancy of more than five points.

Table two: Discrepancies in Assessment of Core CBT Skills (ACCS) total scores pre- and post-training.

	Pre-training	**Post-training**
Discrepancy in ACCS total score pre – post training	**Number of pairwise comparisons (%)** [1]	
10.01 - 11 points	3 (.38 %)	4 (0.51 %)
9.01 - 10 points	12 (1.59 %)	3 (0.38 %)
8.01 - 9 points	30 (3.85 %)	15 (1.92 %)
7.01 - 8 points	50 (6.41 %)	19 (2.44 %)
6.01 - 7 points	60 (7.69 %)	39 (5.00 %)
5.01 - 6 points	69 (8.85 %)	61 (7.82 %)
4.01 - 5 points	83 (10.64 %)	93 (11.92 %)
3.01 - 4 points	91 (11.67 %)	107 (13.72 %)
2.01 - 3 points	115 (14.74 %)	130 (16.67 %)
.01 - 2 points	144 (18.46 %)	151 (19.36 %)
0 - 1 points	123 (15.77 %)	158 (20.26 %)

[1] There are 40 participants and thus 780 different pairwise comparisons.

Absolute agreement.

P-bar analysis was conducted to examine the proportion of absolute agreement between the ACCS ratings assigned by participants pre- and post-training (Kalton & Stowell, 1979). This measure of reliability is based on the various possible comparisons of ratings for each item. Within the current study, there are 40 participants and 780 different pairwise comparisons of the ratings assigned for each item. To conduct this analysis, the 4-point response scale was converted into seven discrete categories as follows: 1, 1.5, 2, 2.5, 3, 3.5, and 4. Using discrete categories allowed an index of reliability to be calculated by computing the proportion of all the 780 paired comparisons in which the two participants assigned the same coding (referred to as P-bar[4]).

Nearly all the p-bar values (81.82 %: see Table three) lie between .30 and .63 at both pre- and post-training. The exceptions are as follows. Items 5.1 (choosing suitable measures) and 5.2 (implementing measures) demonstrated low absolute agreement pre-training (.25 and .28

[4] An aid to the interpretation of the P-bar coefficients is provided by noting that if 39 of the 40 participants agree on the rating of an item, then P-bar = .95 for that item; if the participants split into twenty in agreement and nineteen in agreement, then P-bar = .46.

respectively), although this increased following training (.34 and .30 respectively). Finally, item 2.1 (coherent formulation) fell to .26 post-training and item 4.1 (reviewing homework) rose to .73 post-training. McNemar chi-square revealed that the proportion of absolute agreements increased significantly from pre- to post-training for 17 out of the 22 items (77.27 %), with the largest improvements being seen in the following items: 3.2. choosing suitable interventions, 4.1. reviewing homework, 6.1. pace, and 6.3. maintained focus (see Table three). The proportion of absolute agreements decreased significantly from pre- to post-training for five out of the 22 items (22.73 %): 2.1. coherent formulation, 3.3. rationale for interventions, 3.4. implementing interventions, 3.5. reviewing interventions, and 8.1. patient feedback (see Table three). For most of these items, the proportion of absolute agreements remained relatively high (> .39) post-training. However, for the formulation item, the proportion of absolute agreement was relatively low both pre- and post- training (.34 pre vs. .26 post).

Table three: Proportion of absolute agreement between ACCS ratings assigned pre- and post-training.

Item	P-bar [1]		McNemar chi-square	Pre- post training change?[2]
	Pre-training	Post-training		
1.1	.34	.44	p = <.001	*Sig. improvement*
1.2	.35	.41	p = <.001	*Sig. improvement*
2.1	.34	.26	p = <.001	*Sig. decrease*
3.1	.44	.48	p = <.001	*Sig. improvement*
3.2	.43	.62	p = <.001	*Sig. improvement*
3.3	.53	.42	p = <.001	*Sig. decrease*
3.4	.56	.50	p = <.001	*Sig. decrease*
3.5	.48	.44	p = <.001	*Sig. decrease*
4.1	.47	.73	p = <.001	*Sig. improvement*
4.2	.43	.58	p = <.001	*Sig. improvement*
4.3	.47	.50	p = <.001	*Sig. improvement*
4.4	.43	.45	p = <.05	*Sig. improvement*
5.1	.25	.34	p = <.001	*Sig. improvement*

5.2	.28	.30	p = <.001	*Sig. improvement*
6.1	.35	.58	p = <.001	*Sig. improvement*
6.2	.33	.41	p = <.001	*Sig. improvement*
6.3	.41	.60	p = <.001	*Sig. improvement*
7.1	.42	.51	p = <.001	*Sig. improvement*
7.2	.42	.51	p = <.001	*Sig. improvement*
7.3	.47	.61	p = <.001	*Sig. improvement*
8.1	.40	.39	p = <.05	*Sig. decrease*
8.2	.37	.38	p = <.05	*Sig. improvement*

[1] P-bar = the proportion of all the 780 different pairwise comparisons in which the two participants assigned the same coding.

[2] 'Significant increase' indicates an *increase* in the proportion of agreements from pre- to post-training. 'Significant decrease' indicates a *decrease* in the proportion of agreements from pre- to post-training.

Usability

Results of the usability questionnaire are summarised in Table four. As the data was not normally distributed, non-parametric Wilcoxon signed-rank tests were performed to examine the difference between usability scores assigned by participants pre- and post-training. There was a significant increase in participants' ratings of the clarity of the items in the ACCS, how easy the ACCS was to use, and overall satisfaction with the ACCS from pre- to post-training. There was no significant difference between participants' ratings of the relevance of the ACCS or the overall style, appearance, and layout of the ACCS from pre- to post-training, with participants rating the ACCS as being 'very relevant' and as having a 'very good' style, appearance and layout at both time points.

Table Four: ACCS usability scores assigned by participants pre- and post-training.

	Pre- training Mean (*SD*)	**Post-training Mean (SD)**	**Mean diff *(SD)***	***T***
Relevance	3.58 *(0.55)*	3.69 *(0.51)*	.11 *(.59)*	42.50
Clarity	3.13 *(0.61)*	3.49 *(0.50)*	.36 *(.64)*	9.00 *
Style / Appearance	3.44 *(0.50)*	3.51 *(0.50)*	.08 *(.24)*	0.00
Ease of Use	2.70 *(0.65)*	3.18 *(0.50)*	.48 *(.60)*	0.00 *
Overall Satisfaction	2.93 *(0.69)*	3.33 *(0.53)*	.40 *(.71)*	28.50 *

* $p < .001$

Responses to the open-ended questions about the ACCS scale are outlined in Appendix six. Twenty six participants responded to the open ended question about what they 'most liked' about the ACCS pre-training and thirty one responded post-training. Ease of use was highlighted by a number of participants both pre- (*n* = 12) and post- (*n* = 14) training. Participants also liked the specificity of the scale (*n* = four participants pre- and post- training), the inclusion of basic therapeutic skills (*n* = three participants pre- and one participant post- training), the comprehensive nature of the scale (*n* = two participants pre- and three participants post-training), the formative function of the scale (*n* = three participants post- training), and the relevance of the scale (*n* = two participants pre- and one participant post-training). Twenty six participants responded to the open ended question about what they 'least liked' about the ACCS pre-training and twenty responded post-training. Many of the comments were idiosyncratic (see Appendix six). However, a recurring theme was the amount of detail in the ACCS and, as a consequence of this, the need for practice and familiarity with the scale before use and the amount of time taken to complete a rating (*n* = ten participants pre- and six participants post-training).

Satisfaction with Workshop

The workshop was rated highly by participants, with mean scores on the 1-5 scale as follows: the content was interesting (*M* = 4.23, *SD* = 0.62), the content was useful (*M* = 4.40, *SD* = 0.67), the training was well structured (*M* = 4.53, *SD* = 0.55), the presentation was good (*M* = 4.50, *SD* = 0.55), the hand-outs were useful and relevant (*M* = 4.55, *SD* = 0.55), I was offered the opportunity to participate as much as I wished (*M* = 4.58, *SD* = 0.50), and the workshop helped me to understand how to use the ACCS scale (*M* = 4.55, *SD* = 0.60).

Qualitative feedback from participants' questionnaires is summarised in Appendix seven. Thirty-four participants commented on the 'best thing' about the workshop. The most widely endorsed comment was that the workshop included opportunity to practice using the ACCS scale and subsequently to engage in group discussions about why people assigned the ratings they did and the reasoning for any differing viewpoints (highlighted by 27 participants). Twelve participants highlighted one or more 'least satisfactory thing' about the workshop, the most common of which were difficulty concentrating when watching/rating a long treatment session at the end of the day (highlighted by five participants) and issues with the training environment (highlighted by four participants). Fourteen participants offered one or more piece of 'advice about how the workshop could be run better another time', the most common of which were more group discussions, less paperwork, and shorter videos (all highlighted by two participants).

Discussion

Summary of Results

It was hypothesised that participating in a training session about how to use the ACCS scale would result in improved inter-rater reliability (IRR) and the results of the study do indicate some improvements in the level of agreement between ACCS scores assigned by participants pre- and post-training. Standard deviation for ACCS scores was lower post-training compared to pre-training for the majority of scale items (95.45 %) and domains (75.00%) and the standard deviations for ACCS scores assigned for individual items were significantly smaller post-training compared to pre-training. These results indicate that the degree of variation in the ACCS scores assigned fell from pre- to post-training. Additionally, the proportion of absolute agreements between participants increased significantly from pre- to post-training for most of the scale items (77.27 %) and the level of disagreements between total scores assigned by participants also fell from pre- to post-training.

Not all of the results, however, indicated an improvement in IRR following training. The standard deviations for ACCS total domain scores were not significantly smaller post-training, although the trend in the data indicated a decrease from pre- to post-training which narrowly missed significance (p = .09). The proportion of absolute agreements also decreased significantly from pre- to post-training for five (22.73 %) of the scale items. However, the proportion of agreement for four of these items remained high (> .39) post training. Additionally, absolute agreement is a very stringent test of IRR because it requires participants to provide exactly the same rating in order to be considered in agreement. Thus, despite these inconsistencies, the

results from this study overall do seem to indicate that a one-day training session in how to use the ACCS scale leads to improved IRR. These results are in line with previous research which shows that assessor training leads to an improvement in IRR for clinician-rated scales (Guar et al., 2010; Kobak et al., 2005; Kobak et al., 2006; Kobak et al., 2007; Muller et al., 1998; Rosen et al., 2008; Tracy et al., 1997) and clinical performance assessments (Bachrach et al., 1971; Gordon, 2006; Lin et al., 2013; Reichelt et al., 2003).

It was also hypothesised that participating in the training would result in improved usability ratings. Results from the study demonstrate that the training did have a positive impact on the perceived usability of the ACCS, with participants finding the ACCS items clearer, the ACCS easier to use, and overall being more satisfied with the ACCS after training. There was no significant difference in the overall style, appearance, and layout or the relevance of the ACCS from pre- to post-training. It is, however, not surprising that training did not impact on these variables, as they relate to stable aspects of the ACCS scale, the perception of which would not necessarily change as the result of training. Indeed, appearance and relevance scores were high at both time points, indicating that participants were satisfied with the appearance and relevance of the ACCS before training and remained so after training.

Given the improvements in IRR and usability of the ACCS, it could be concluded that participant training is necessary in order to use the ACCS scale. However, the degree of variation in the ACCS scores assigned (as measured by the standard deviation) was low for most of the ACCS domains both pre- and post-training (62.50 % < .58). The level of disagreement between total ACCS scores assigned by participants were also low both pre- and post-training, with the number of pairings with a discrepancy of more than five points only 28.21 % pre- and 18.08% post-training. This level of disagreement is lower than the percentage of pairings with a discrepancy of more than five points reported by Gordon (2006) for the CTS-R (50%) and CTS-Psy (55%) (Cognitive Therapy Scale- Psychosis: Haddock et al., 2001), despite both scales having a larger range in total scores than the ACCS. Similarly, the proportion of complete agreement between participants was high both pre- and post-training for most items (81.82% of items had a p-bar value > .30). A similar pattern can also be seen for usability, with the participants at both pre- and post- training rating the ACCS as being at least quite clear (M > 3 pre and post), somewhat easy to use (M > 2.5 pre and post), and being at least somewhat satisfied with the ACCS (M > 2.50 pre and post). Taken together, these results indicate that, whilst training had a positive impact on usability and IRR, both were already acceptable pre-training (i.e. with use of the ACCS manual alone). Simply providing assessors with the ACCS manual may, therefore, be sufficient to achieve acceptable levels of IRR and usability. However, this strategy is dependent

upon assessors being both willing and able to spend time reading the manual and then referring to it when they begin using the ACCS to rate therapists' competence. Furthermore, whilst participants in this study did highlight the specificity and comprehensively of the ACCS as a key strength of the scale, they also felt that the amount of detail in the ACCS manual was challenging. Consequently participants felt that the opportunity to practice and become familiar with the scale before using it was important. Thus, assessors may feel more confident and able to use the ACCS if they are able to participate in training.

Implications for the ACCS Scale

Findings from this study highlight two domains as 'problematic' in terms of IRR. These were domain two 'formulation' (including one item: 'coherent formulation') and domain five 'measuring change' (including two items: 'choosing suitable measures' and 'implementing measures'). There was high degree of variation in the scores (as measured by standard deviation) participants assigned to items in both domains at pre- and post-training, with neither of the domains showing a reduction in variance from pre- to post-training (although there was a reduction in variance for the individual measuring change items). The items in the formulation and measuring change domains also demonstrated low absolute agreement pre-training. This did not increase post-training for the formulation item (in fact the level of agreement fell further), but did increase following training for the measuring change items. Furthermore, there was a reduction in the degree of absolute agreement for the formulation item from pre- to post-training. It is notoriously difficult to achieve agreement between clinicians regarding CBT formulations, with research consistently demonstrating only modest agreements (Dudley, Park, James, & Dodgson, 2010; Kuyken, Fothergill, Musa, & Chadwick, 2005; Mumma & Smith, 2001; Persons & Bertagnolli, 1999; Persons, Mooney, & Padesky, 1995). This difficulty in reaching consensus about what constitutes a relevant and coherent formulation could have influenced the ratings provided within the formulation item. Measuring change is an aspect of CBT competence which has not been previously measured within CBT competence rating scales. The novel nature of the domain could to some degree explain the low IRR pre-training, followed by an increase in absolute agreement post-training. Participants in the focus group (see chapter six) also commented that both the formulation and measuring change domains are sometimes difficult to observe within treatment sessions, meaning that the ratings provided are likely to rely more heavily on assessor inference than other domains in the ACCS. This higher degree of assessor inference may also go some way to explaining the lower IRR found for these domains.

There was also a reduction in the degree of absolute agreement for four other items (i.e. in addition to the formulation item as outlined above). The reduction in agreement for the patient feedback item (8.1) may be explained by the fact that the end of the second recording failed to play. Although therapists are expected to elicit, explore, and respond to feedback about the patient's understanding of and reaction to all aspects of session, a significant amount of this skill set is typically evident when the therapist elicits patient feedback at the end of the session. Thus, missing this aspect of the session may have skewed the ratings provided for this item. There was also a reduction in absolute agreement for three items within domain three (CBT interventions): 3.3 rationale for interventions, 3.4 implementing interventions, and 3.5 reviewing interventions. One possible explanation for this reduction in agreement may be that the CBT interventions implemented in the first role play were simpler and clearer than those implemented in the second role play, thus resulting in less uncertainty and disagreement between participants.

Quality of Assessor Training

The training workshop was rated highly by participants: they found the content interesting and useful and thought the training was well structured, the presentation was good, the hand-outs were useful and relevant, and that they were offered ample opportunity to participate. Perhaps most importantly, they also felt that the workshop helped them to understand how to use the ACCS. Qualitative feedback indicated that the opportunity to practice using the ACCS scale and engage in group discussions about why people assigned the ratings they did and the reasoning for any differing viewpoints appeared to be particularly helpful in providing insight into how to use the ACCS. The main weakness of the workshop highlighted by participants related to the role-plays that were used. Although they were suitable in that they provided a representation of a typical CBT treatment session, they were originally created for another purpose. Thus, it would be useful to record new role plays for the specific purpose of this workshop. This would enable the therapist in the role-play to demonstrate a wider range of skill levels, including some common therapist errors or mistakes, and for the 'patient' to role play a more challenging, complex presentation. Some trainees also felt that it was difficult to concentrate on the second role play as it was too long (one hour). Although this is a realistic length for a therapy session, it is suggested that individuals often struggle to concentrate for more than 20 minutes on single task (Fry & Ketteridge, 2009). Thus, in the future, it may be more effective for training purposes to use shorter role-plays.

Strengths and Limitations

The study provided a naturalistic evaluation of the impact of a one-day training workshop in how to use the ACCS scale. This naturalistic approach is beneficial in that it demonstrates high external validity. For example, the therapists who participated in the training and completed ACCS ratings were representative of those who will utilise the scale in practice (i.e. CBT therapists and supervisors working in general practice who are unfamiliar with the ACCS scale). This is in stark contrast to the majority of studies evaluating the psychometric properties of competence rating scales that use assessors who are experts in the treatment they are assessing (e.g. many have been prominent in the development of CBT treatment packages) and who were involved in the development of the rating scale being evaluated (Barber et al., 2007; Muse & McManus, 2013). The sample size included in this study is also relatively large (n = 40), especially given that the study involved viewing and rating recordings of CBT treatment sessions and thus was both time consuming and required trained CBT therapists to participate.

The naturalistic approach employed also brings with it a number of limitations which are inherent in this type of 'real world' research. In particular, the lack of a controlled experimental setting means that a number of confounding variables could have influenced the results of the study. For example, different role-plays were used pre-and post-training. This was necessary because training involved discussion about the discrepancies between different participants' ratings of the first role-play session, which would inevitably result in greater consistency between raters if the same session was re-rated. However, the two role-plays that were used differed in a number of important ways. These included length (with the second being nearly 20 minutes longer), type of presenting problem (bulimia vs. depression), complexity of patient's problems (the first being rated as more complex), and the therapist delivering CBT. Perhaps more problematic is the technical error with the second role-play which meant that the end of the recording failed to play. A number of participants also noted that they found their ability to concentrate had diminished by the second (and longer) role-play rated at the end of the day, suggesting that fatigue may also have impacted on the usability ratings and ACCS scores assigned for the second role play. Another limitation in this study is the lack of a control group who did not receive any training in how to use the ACCS scale. Without this, it is not possible to rule out practice effects as the source of improvement. Indeed, it is possible that merely providing assessors with the opportunity to practice using, and thus to become familiar with the ACCS, is all that is needed to achieve reliability ratings. It is also possible that demand characteristics played a role in the usability ratings and workshop feedback provided by attendees, especially as senior service managers were present at the workshop. However, both of these feedback forms were

completed anonymously in order to minimise the impact of any social desirability biases. All of these factors could have impacted on the usability ratings and ACCS scores assigned, making it difficult to be confident that any changes in the usability and inter-rater agreement of ACCS ratings from pre- to post-training were due to the impact of training.

Providing ratings on role-plays, rather than real treatment sessions, may also have been problematic in that it may have provided a simplified or unauthentic representation of a 'real' CBT session typically seen within routine clinical practice (Sharpless & Barber, 2009). Indeed, some therapists who attended the workshop commented that the patients in the role-plays presented with less 'severe' problems than those they typically see and that the individuals delivering CBT had a particularly high skill level. Although research has shown that recordings of professional actors playing patients with depression could not be distinguished between those of actual patients (Rosen, Mulsant, Bruce, Mittal, & Fox, 2004), the individuals role-playing the patients in this study had no experience of acting. That being said, they were experienced CBT therapists with a good understanding of mental health problems. It is also worth noting that, due to time and resource constraints, the supporting information was created by the study authors, rather than the individuals who created the role-play.

Future Research

A number of questions that warrant further exploration remain unanswered by this pilot study. It is, for example, unclear how much training is necessary and sufficient to achieve acceptable IRR and for assessors to feel confident enough to use the scale. The lack of a follow-up assessment also means that the question of whether the improvements in usability and IRR are maintained over time remains unanswered, as does the question of whether further 'top up' training is required to prevent rater drift. The sample of therapists in the current study varied in terms of CBT experience and experience of using rating scales. However, the range and size of the sample was too small to enable stratification into subgroups. Thus, it would also be interesting for future research to examine what impact clinical experience, prior use of rating scales, or education level might have on the effectiveness of training. It may, for example, be the case that training is most needed, and therefore most effective, for assessors with limited experience of delivering CBT and using rating scales. Alternatively, it may be that training reduces confusing comparisons being drawn between the ACCS and previous scales, and thus is of more benefit to those with more experience of using other rating scales.

Another avenue that may be worth further exploration is the development of an online training package in how to use the ACCS. Although initial development costs for online training

packages may be higher compared to face-to-face workshops, this method of training is likely to be more cost-effective in the long term. Online training also overcomes many of the practical difficulties involved in providing live training sessions (Fairburn & Cooper, 2011). For example, online training can be completed remotely and individually and thus does not require a large number of busy practitioners, often from geographically diverse sites, to meet in the same place at the same time. There is also no need for 'experts' with a good understanding of the ACCS scale to be physically present for an online training session. Providing online training would also improve accessibility, particularly for those in remote areas. Furthermore, given that inter-rater differences may still arise between assessors who receive different training (Barber et al., 2007), the use of a standardised training package may be beneficial. However, it will be important to develop a package which does more than providing recordings of didactic talks which attempt to transmit knowledge and information to the trainees, as form of teaching that typically induces passivity (Brookfield, 2006). A much more effective approach is one which allows trainees to play an active role in their learning, thereby improving engagement, understanding and retention of subject matter (Bonwell & Eison, 1991). Indeed, the opportunity to practice using the ACCS scale, and to discuss the ratings made with peers and trainers, was a core component of the current training and one that participants reported to be one of the most useful elements of the workshop. Interactive online tutorials have been developed and used successfully to train raters in how to use clinician-rated scales for depression, anxiety and schizophrenia (Kobak et al., 2005; Kobak et al., 2006; Kobak et al., 2007; Rosen et al., 2008). These training packages used a mixture of didactic talks (viewed via CD-ROM) and remote training sessions where participants interviewed a standardized patient-actor while being remotely observed in real time and given feedback. Future research could usefully develop and compare the impact of interactive online and face-to-face training packages in how to use the ACCS rating scale.

References

Bachrach, H., Mitntz, J., & Lurborsky, L. (1971). On rating empathy and other psychoetherapy variables. *Journal of Consulting and Clinical Psychology, 36*(3), 445 - 445. doi: 10.1037/h0031109

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology-Research and Practice, 38*(5), 493-500. doi: 10.1037/0735-7028.38.5.493

Barker, C., Pistrang, N., & Elliot, R. (2002). *Foundations of quantitative measurement*. Chichester: John Wiley & Sons.

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. doi: 10.1017/S1352465801004040

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom*. Washington DC: ASHE-ERIC Higher Education Report No 1.

Brookfield, S. D. (2006). *The skillful teacher: On technique, trust, and responsiveness in the classroom* (2nd ed.). San Fransisco: Jossey Bass.

Decker, S., Jameson, M., & Naugle, A. (2011). Therapist training in empirically supported treatments: A review of evaluation methods for short- and long-term outcomes. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(4), 254-286. doi: 10.1007/s10488-011-0360-1

Dudley, R., Park, I., James, I., & Dodgson, G. (2010). Rate of agreement between clinicians on the content of a cognitive formulation of delusional beliefs: The effect of qualifications and experience. *Behavioural and Cognitive Psychotherapy, 38*(2), 185-200. doi: 10.1017/S1352465809990658

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49*(6-7), 373-378. doi: 10.1016/j.brat.2011.03.005

Fry, H. S., & Ketteridge, S. (2009). *A handbook for teaching and learning in higher education: Enhancing academic practice* (3rd ed.). London: Routledge.

Gordon, P. K. (2006). A comparison of two versions of the cognitive therapy scale. *Behavioural and Cognitive Psychotherapy, 35*(3), 343-353. doi: 10.1017/S1352465806003390

Guar, R., Kaviani, H., Bansal, S., & Lee, N. (2010). Rater training on PANSS and SANS. *Schizophrenia Research, 117*(2 - 3), 265 - 266. doi: dx.doi.org/10.1016/j.schres.2010.02.422

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. doi: 10.1017/S1352465801002089

Kalton, G., & Stowell, R. (1979). A study of coder variability. *Journal of the Royal Statistical Society Series C (Applied Statistics), 28*(3), 276 - 289.

Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs: The four levels* (3rd ed.). San Fransisco: Berrett-Koehler.

Kobak, K. A., Alan, D. F., & lipstiz, J. D. (2005). Interview quality and signal detection in clinical trials. *The American Journal of Psychiatry, 162*, 628 - 628. doi: 10.1176/appi.ajp.162.3.628

Kobak, K. A., Engelhardt, N., & Lipstiz, 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. doi: 10.1016/j.jpsychires.2005.07.012

Kobak, K. A., Opler, M. G. A., & Engelhardt, N. (2007). PANSS rater training using Internet and videoconference: Results from a pilot study. *Schizophrenia Research, 92*(1 - 3), 63 - 67. doi: 10.1016/j.schres.2007.01.011

Kuyken, W., Fothergill, C. D., Musa, M., & Chadwick, P. (2005). The reliability and quality of cognitive case formulation. *Behaviour Research and Therapy, 43*(9), 1187-1201. doi: 10.1016/j.brat.2004.08.007

Lin, C.-J., Change, J. Z.-C., Hsu, T.-C., Lui, Y. J., Yu, S.-H., Tsai, 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. doi: 10.1016/j.jds.2013.01.002

Muller, M. J., Rossbach, W., Dannigkeit, P., Muller-Siechender, F., Szegedi, A., & Wetzel, H. (1998). Evaluation of standardized rater training for the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Research, 32*(3), 151 - 160. doi: 10.1016/S0920-9964(98)00051-6

Mumma, G. H., & Smith, J. L. (2001). Cognitive-Behavioral-Interpersonal scenarios: Interformulator reliability and convergent validity. *Journal of Psychopathology and Behavioral Assessment, 23*(4), 203-221. doi: 10.1023/A:1012738802126

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

Persons, J., & Bertagnolli, A. (1999). Inter-rater reliability of cognitive-behavioral case formulations of depression: A replication. *Cognitive Therapy and Research, 23*(3), 271-283. doi: 10.1023/a:1018791531158

Persons, J., Mooney, K., & Padesky, C. (1995). Interrater relability of cognitive-behavioral case formulations. *Cognitive Therapy and Research, 19*(1), 21-34. doi: 10.1007/bf02229674

Reichelt, F., 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. doi: 10.1016/S0005-7916(03)00022-3

Rosen, J., Mulsant, B. H., Bruce, M. L., Mittal, V., & Fox, D. (2004). Actors' portrayals of depression to test interrater reliability in clinical trials. *American Journal of Psychiatry, 161*(10), 1909 - 1911. doi: 10.1176/appi.ajp.161.10.1909

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. doi: 10.1016/j.psychres.2008.03.001

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. doi: 10.1016/j.cpr.2008.09.008

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

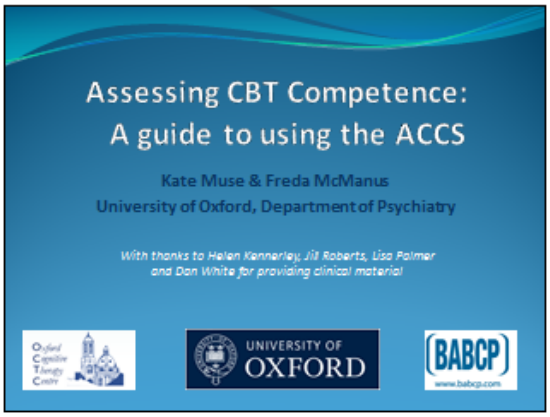
Tracy, K., Adler, L. A., Rotrosen, J., Edson, R., & Lavori, P. (1997). Interrater reliability issues in multicenter trials, Part I: Theoretical concepts and operational procedures used in Department of Veterans Affairs Cooperative Study #394. *Psychopharmacol Bulletin, 33*(1), 53 - 57.

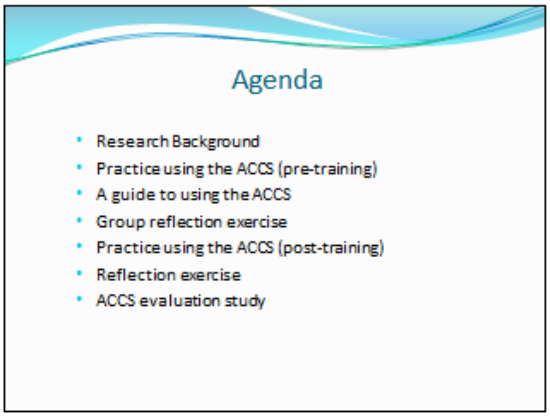
Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114*(2), 248-265. doi: 10.1037/0033-2909.114.2.248

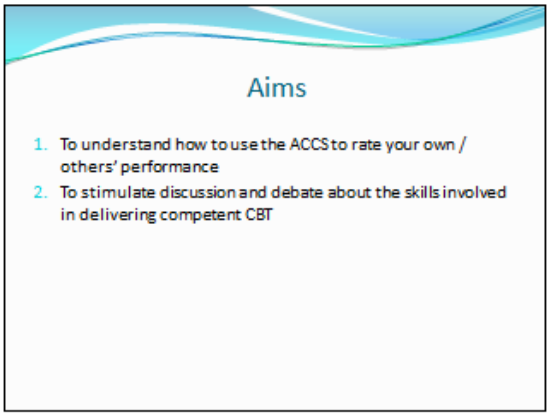
Williams, R. M., Moorey, S., & Cobb, J. (1991). Training in cognitive-behaviour therapy: Pilot evaluation of a training course using the cognitive therapy scale. *Behavioural and Cognitive Psychotherapy, 19*(04), 373-376. doi: 10.1017/S0141347300014075

Appendicies

Appendix 1 – Hand-out of PowerPoint slides for ACCS training

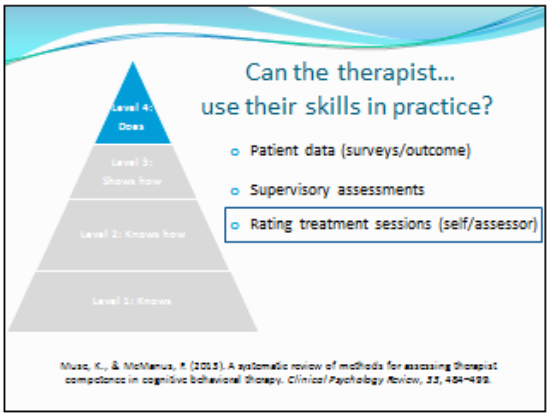












Why is a new scale needed?

- Standardised scales already available- *most valid / reliable scales:*
  - CTS (Beck Institute)
  - CTS-R (Blackburn et al., 2001)
  - CTS-Psy (Haddock et al., 2001)
  - CTCS-SP (Van Emmerik, Clark & Kamphuis, 2011)

→ *But* these are protocol / disorder-specific (not feasible in RCP)

Why is a new scale needed?

CTS (Beck Institute) + CTS-R (Blackburn et al., 2001)

But problems with these:

Validity -
- Requires updating in light of recent advancements in CBT
- Cognitive bias
- Focus on depression
- Under emphasis of basic therapeutic skills
- Item overlap
- Multiple concepts addressed by single items

Why is a new scale needed?

Reliability -
- Inadequate definition of scale points / lack of behavioural specificity
- Poor inter-rater reliability (even with training)
- Limited sensitivity

Usability -
- Unclear instruction manual
- High assessor inference
- Lack of formative feedback

So - there is a need for generic competence measures with improved validity, reliability & usability

Aims of the ACCS

To provide a generic CBT competence measure with improved validity, reliability & usability.

Development of the ACCS

Three key stages:

1. Scale development –
   Create a first draft (lit review, CTS, CTS-R, Roth & Pilling's (07) framework)
2. Expert review –
   Find out what experts in the field think & make adaptations based on their feedback
3. Scale evaluation -
   How well does the scale work in practice? Any further refinements necessary?

Questions, comments, thoughts?



Lets practice using the ACCS!

"But you haven't told us how to use the scale yet!?"

This is because:

1. We are interested in how easy the scale is to use when assessors only have the manual (i.e. no training)
2. We will review your ratings of the role-play CBT session when we discuss the core ACCS domains later in the day

Lets practice using the ACCS!

1. Read over the supporting information
2. Watch the video and make notes
3. Rate the the skilfulness of the therapists' performance (numerical only)
4. Complete the 'usability' questionnaire + demographics form

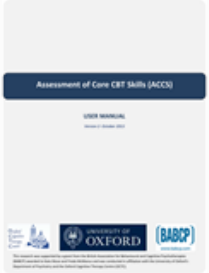
Notes:

- Use the **ACCS manual** as your guide
- Work alone, rather than conferring with colleagues (there will be time for this later!)

Group reflection exercise

- Get into small groups (4/5 people)
- In your groups discuss the following:
  1. What score did you assign for each item? Why did you give this score?
  2. Did any of your colleagues provide different scores to you? Why did these differences arise?
  3. Is the meaning of all of the items clear to everybody?
  4. Did you struggle to rate any items? Why do you think this happened?
  5. Does your group have any questions or uncertainties about how to use the scale?

A guide to using the ACCS



What is the aim of the ACCS?

To provide a framework for:

- Assessors to deliver formative & summative feedback about a therapist's performance within an observed CBT treatment session
- Therapists to rate and reflect on their own performance within a CBT treatment session

What does the ACCS assess?

- Core general therapeutic + CBT specific skills
- Individual CBT to adults experiencing mental health problems
- Range of cognitive/behavioural therapies
- Transdiagnostic
- 'Active' CBT session

What does the ACCS *not* assess?

- CBT knowledge (MCQs, essays, case reports, vignettes)
- Broader professional skills (ethics, supervision etc.)
- Assessment skills
- Relapse-prevention skills
- Skills specific to a disorder / protocol (e.g. reliving in PTSD, video feedback in SP)
- CBT with children / adolescents

Core ACCS domains



Domain 1: Agenda setting

- **Planning and setting an agenda <u>not</u> adhering to it (see domain 6 'effective use of time')**

2 items

Suitable Items
- Help patient identify & prioritise specific, relevant and appropriate items.

Feasible Agenda
- Agenda realistic & feasible given time available.

Domain 2: Formulation

- **Should be evident throughout the course of treatment** (evolves as further clinical information arises / recurrent patterns & themes emerge)
- Also view written / diagrammatic formulations
- ***Not*** whether therapist drew upon formulation to highlight treatment targets / intervention strategies (see domain 3 'CBT interventions')

1 item

Coherent Formulation
- Develop a clear formulation which draws upon appropriate evidence-based theory to offer a concise, comprehensive & personalised explanation of relevant history, triggers & maintaining features of patient's problems.

Domain 3: CBT Interventions

- Cognitive and/or behavioural interventions (CBT 'techniques')
- Designed to promote therapeutic change by targeting maintaining factors (cognitions, behaviours, emotions)



Domain 3: CBT Interventions

5 items

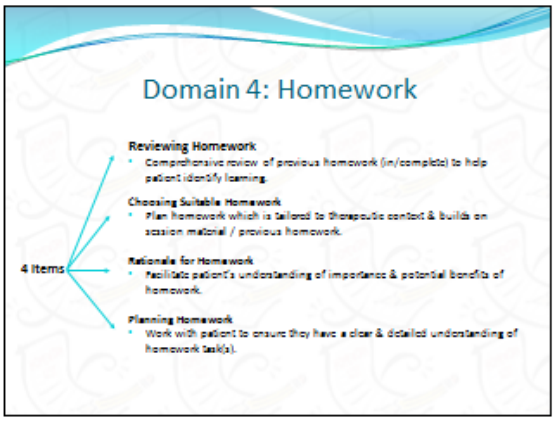
Appropriate Intervention Targets
- Define, clarify & specify targets which relevant evidence-based theory + formulation suggests are maintaining problems.

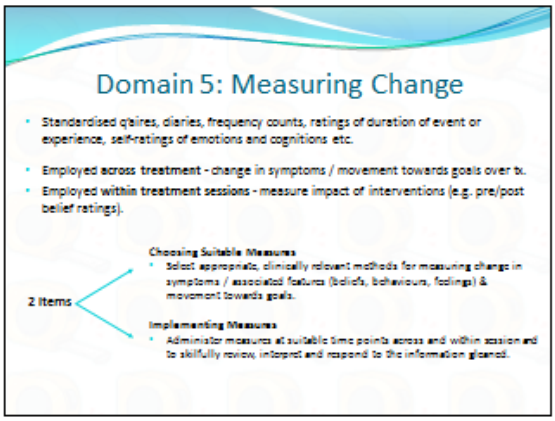
Choosing Suitable Interventions
- Interventions form part of a logical, coherent & unified treatment strategy.
- Strategy likely to bring about therapeutic change + is suited to therapeutic context
- Guided by theory / evidence (when possible).

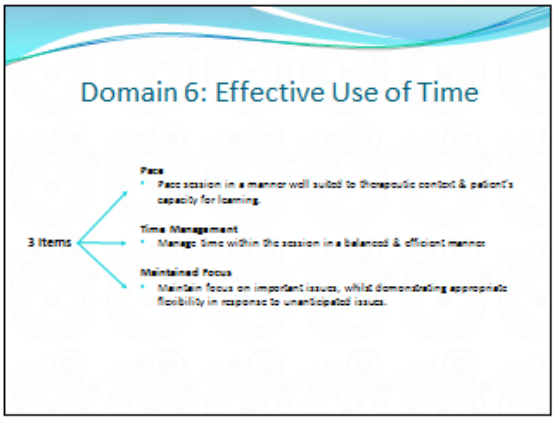
Rationale for Interventions
- Facilitate patient's understanding of importance & potential benefits.

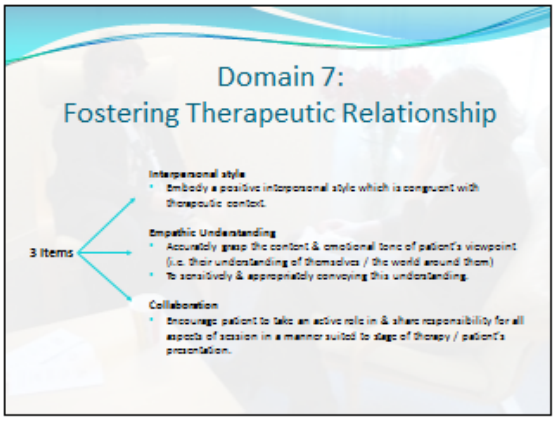
Implementing Interventions
- Systematically implement intervention(s) in fluent & articulate manner
- Sensitive & responsive to therapeutic context
- Provide optimal support, encouragement & praise.

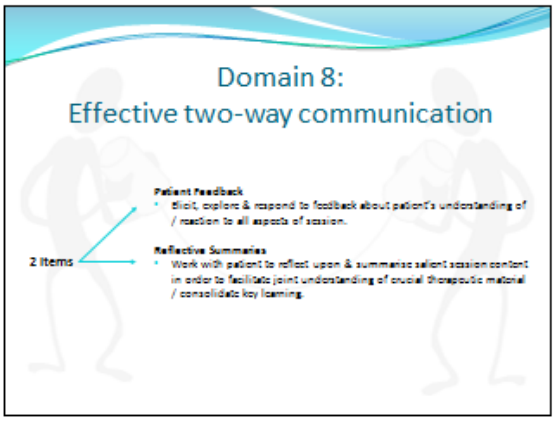
Reviewing Interventions
- Comprehensive review of results (+ / -) to help patient identify learning

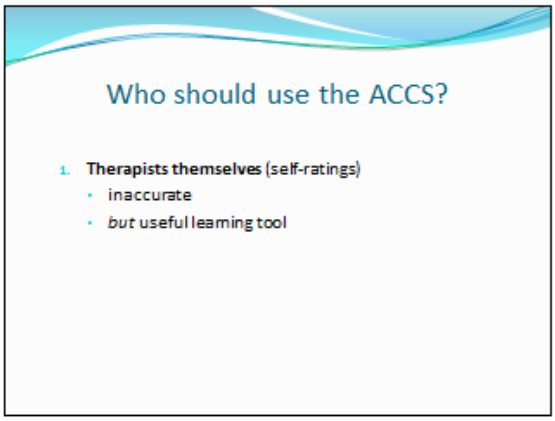

Domain 4: Homework
4 items
Reviewing Homework
Comprehensive review of previous homework (in/complete) to help patient identify learning.
Choosing Suitable Homework
Plan homework which is tailored to therapeutic context & builds on session material / previous homework.
Rationale for Homework
Facilitate patient's understanding of importance & potential benefits of homework.
Planning Homework
Work with patient to ensure they have a clear & detailed understanding of homework task(s).


Domain 5: Measuring Change
Standardised q'aires, diaries, frequency counts, ratings of duration of event or experience, self-ratings of emotions and cognitions etc.
Employed across treatment - change in symptoms / movement towards goals over tx.
Employed within treatment sessions - measure impact of interventions (e.g. pre/post belief ratings).
2 items
Choosing Suitable Measures
Select appropriate, clinically relevant methods for measuring change in symptoms / associated features (beliefs, behaviours, feelings) & movement towards goals.
Implementing Measures
Administer measures at suitable time points across and within session and to skilfully review, interpret and respond to the information gleaned.


Domain 6: Effective Use of Time
3 items
Pace
Pace session in a manner well suited to therapeutic context & patient's capacity for learning.
Time Management
Manage time within the session in a balanced & efficient manner.
Maintained Focus
Maintain focus on important issues, whilst demonstrating appropriate flexibility in response to unanticipated issues.


Domain 7:
Fostering Therapeutic Relationship
3 items
Interpersonal style
Embody a positive interpersonal style which is congruent with therapeutic context.
Empathic Understanding
Accurately grasp the content & emotional tone of patient's viewpoint (i.e. their understanding of themselves / the world around them)
To sensitively & appropriately convey this understanding.
Collaboration
Encourage patient to take an active role in & share responsibility for all aspects of session in a manner suited to stage of therapy / patient's presentation.


Domain 8:
Effective two-way communication
2 items
Patient Feedback
Elicit, explore & respond to feedback about patient's understanding of / reaction to all aspects of session.
Reflective Summaries
Work with patient to reflect upon & summarise salient session content in order to facilitate joint understanding of crucial therapeutic material / consolidate key learning.


Who should use the ACCS?
1. Therapists themselves (self-ratings)
inaccurate
but useful learning tool

Who should use the ACCS?

2. **Assessors**

- **Independent assessors**
  - more difficult to provide formative feedback
  - *but* less 'biased'
- **Supervisors**
  - more susceptible to 'biases'
  - *but* more contextual information (context, broader work etc.)
  - *but* better placed to provide formative feedback

Assessment Materials

- Video or audio
- 'Active' CBT treatment session
- Full treatment session
- **Short** written case summary (formulation, material completed in session, stage of therapy, presenting problem, goals, main topics etc.)

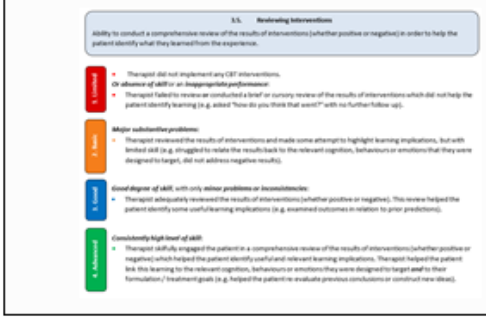
4-step guide to using the ACCS

Step 1: Provide a numerical rating for each item

A- Overarching scale

1. Limited	Therapist fails to include feature outlined. *Or* therapist demonstrates a significant absence of skill or an inappropriate performance which is likely to have negative therapeutic consequences.
2. Basic	Therapist's performance is somewhat appropriate with some degree of skill evident. However, major substantive problems are evident.
3. Good	Therapist demonstrates a good degree of skill with no major problems. However, minor problems or inconsistencies are evident in the therapist's performance.
4. Advanced	Therapist consistently demonstrates a high level of skill with only very few and very minor problems.

B- Item-specific 'exemplar therapist behaviours'



Issues to consider

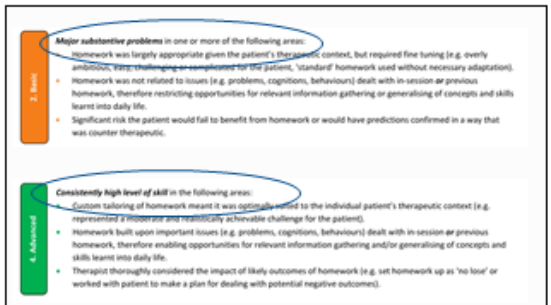
- Use item-specific exemplar therapist behaviours wherever possible (improves consistency)
- Watch entire session *then* rate performance
- Score each item independently (avoid global impressions)

What if therapist demonstrates behaviours from more than one performance banding?

= look at overarching definitions

E.g. 'choosing suitable homework' item

- therapist selected homework related to key maintenance cycles discussed in-session (band 4)
- ***but*** homework task was too challenging for patient (band 2)

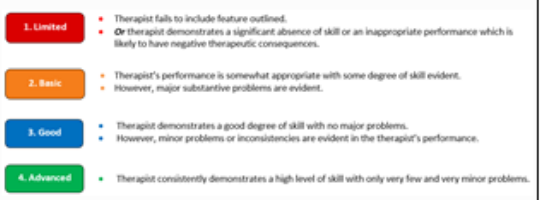


So a rating of 2- Basic

But- use qual feedback to highlight aspects of level 4

What if therapist demonstrates behaviours which are not in the item-specific exemplar therapist behaviours?

= use overarching definitions



Step 2: Provide qualitative feedback

When rating others' sessions:

- For each item & overall skills profile
- Identify key strengths and learning needs
- Give specific examples (exemplar behaviours + session material)
- For weaker domains - suggest more skilful ways of working + ways the therapist could further develop skills

= aids further development & progression

When rating your own sessions:

- Reflect on whole session
- Recognise key strengths (what went particularly well?)
- Identify areas you need to further develop (what didn't go so well?)
- Highlight anything specific you would like feedback on

= promotes self-reflection + provides info to assessor

Step 3: Provide a global performance rating

- Overall impression of therapist's performance within session as a whole
- Use generic 4-point scale
- Will normally correspond to domain ratings (but not always)

Step 4: Rate patient complexity

- Not incorporated within scoring system (confounds ratings)
- ***But*** provides context within which performance has been evaluated

Questions, comments, thoughts?



Lets practice using the ACCS!



1. Read over the supporting information
2. Watch the video and make notes
3. Rate the the skilfulness of the therapists' performance (numerical only)
4. Complete the 'usability' questionnaire

Notes:

- Use the **ACCS manual** as your guide
- Work alone, rather than conferring with colleagues (there will be time for this later!)



Reflection exercise

- What scores did you assign the therapist? Why did you provide these scores?
- What were the key strengths and learning needs of the therapist? How could the therapist further develop their CBT skills?
- Did you find yourself using the scale differently compared to this morning? If so, how? And why do you think this is?
- What were the main things you learned today? How will this be useful to your work with clients, supervision or CPD?

ACCS Evaluation Study

Aims of the study

- To examine whether the ACCS works in practice
- To identify any need for further refinement of the ACCS

What does participation involve?

As part of your CPD you will be asked to submit recordings of CBT treatment sessions to your clinical supervisors for feedback on your performance.

Participation involves **one of these recordings** being viewed and rated using the ACCS scale by three people:

1. Your supervisor within the First Step service
2. A Senior Psychotherapist within the First Step service
3. Yourself (i.e. a self-rating)

Points to Note

- All information which is collected from you would be kept **strictly confidential.** Identifying information (i.e. your name etc.) will be removed from all materials.
- Individual scores **will not** be used in the scale evaluation study (only averages, ranges etc. of the whole group).
- Your session recording **will not** be accessed or viewed by any individuals other than Cumbria Partnership NHS Foundation Trust staff employed within the First Step service.
- Completed ACCS ratings will be **sent directly to Emma Cowperthwaite without being shared with any other parties** (i.e. therapists / supervisors / senior psychotherapists within the service will not see the ratings completed by other members of staff).

Paperwork

1. Information sheet
2. Numerical ACCS Feedback form
3. Submission cover sheet (for supporting information)
4. ACCS Manual (should already have this)

I will also email these documents around for those of you who prefer to work electronically.

Questions



Before you leave please . . .

1. Hand in your ACCS ratings for the **first and second** role-play
2. Hand in your demographics forms and usability questionnaires
3. Complete & hand in your feedback forms

Thank you for your time



kate.muse@psych.ox.ac.uk

Appendix 2: Numerical ACCS Feedback form used to rate the therapist's performance within the role-play treatment sessions.

ACCS Feedback Form

Therapist's Name: _____ **Assessor's Name:** _____

Date of Submission: _____

For each item, please assign a rating of:

1. Limited	2. Basic	3. Good	4. Advanced

Domain 1 - Agenda Setting

1.1. Suitable Items	
1.2. Feasible Agenda	
Total Domain Score *(average item score)*	

See user manual page 7 – 8

Domain 2 - Formulation

2.1. Coherent Formulation	
Total Domain Score *(average item score)*	

See user manual page 9

Domain 3 – CBT Interventions

3.1. Appropriate Intervention Targets	
3.2. Choosing Suitable Interventions	
3.3. Rationale for Interventions	
3.4. Implementing Interventions	
3.5. Reviewing Interventions	
Total Domain Score *(average item score)*	

See user manual page 10 – 14

Domain 4 – Homework

4.1. Reviewing Homework	
4.2. Choosing Suitable Homework	
4.3. Rationale for Homework	
4.4. Planning Homework	
Total Domain Score *(average item score)*	

See user manual page 15 – 18

Domain 5 – Measuring Change

5.1. Choosing Suitable Measures	
5.2. Implementing Measures	
Total Domain Score *(average item score)*	

See user manual page 19 – 20

Domain 6 – Effective Use of Time

6.1. Pace	
6.2. Time Management	
6.3. Maintained Focus	
Total Domain Score *(average item score)*	

See user manual page 21 – 23

ACCS Feedback Form

Domain 7 – Fostering Therapeutic Relationship

7.1. Interpersonal Style	
7.2. Empathic Understanding	
7.3. Collaboration	
Total Domain Score *(average item score)*	

See user manual page 24 – 26

Domain 8 – Effective Two-way Communication

8.1. Patient Feedback	
8.2. Reflective Summaries	
Total Domain Score *(average item score)*	

See user manual page 27 – 28

Overall Session Score

Total Score (*add total domain scores*)

Global Performance Rating

Therapist's overall performance within the session was:

1. Limited | 2. Basic | 3. Good | 4. Advanced

ACCS Feedback Form

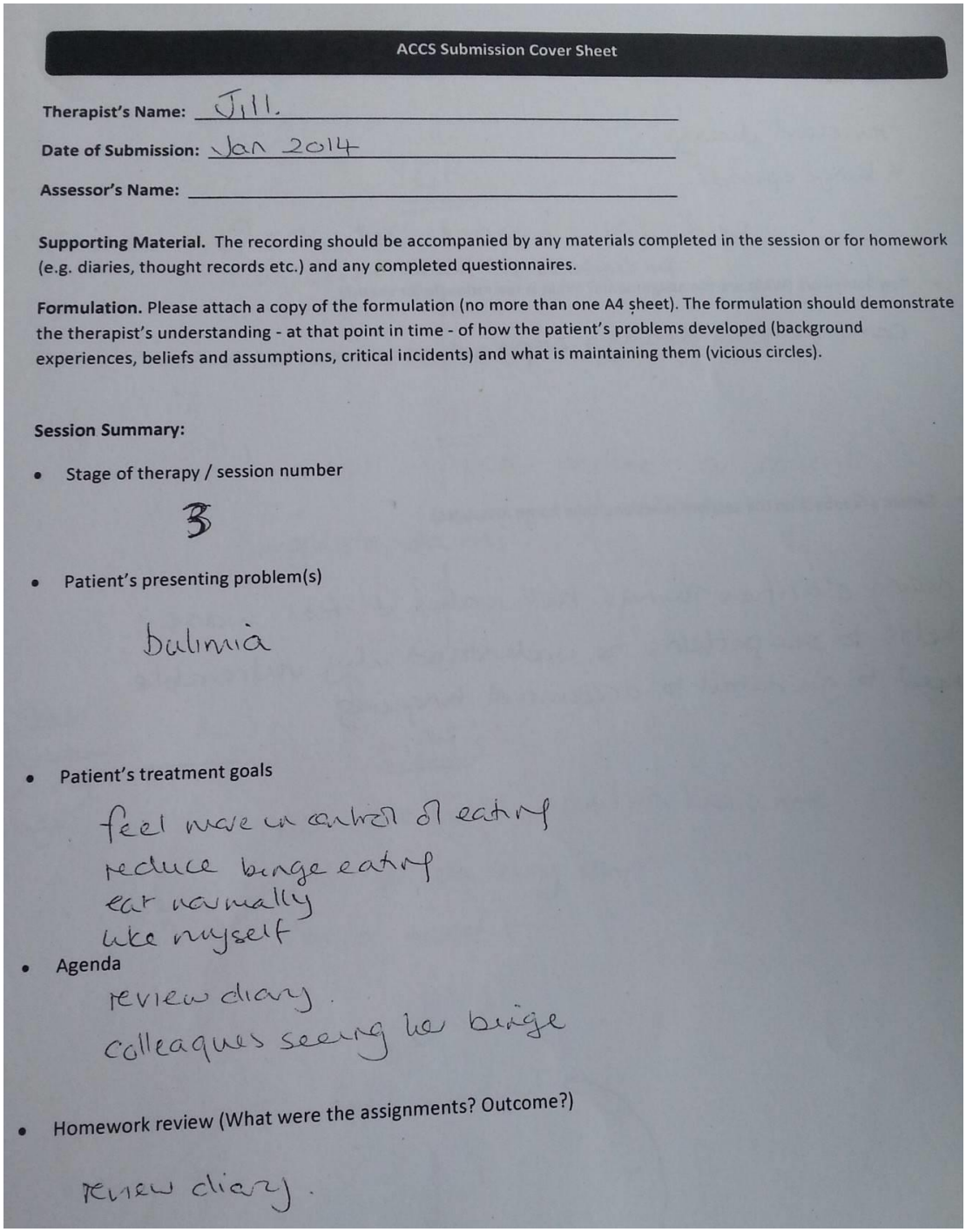
Patient Complexity Rating

How complex do you feel the patient was to work with?

1. **Very Straightforward** **2. Somewhat Straightforward** **3. Somewhat Complex** **4. Very Complex**

Appendix 3: Submission cover sheets provided for participants alongside the role-plays

1) Cover sheet to accompany role-play one (pre-training)



ACCS Submission Cover Sheet

Therapist's Name: Jill.

Date of Submission: Jan 2014

Assessor's Name: _____

Supporting Material. The recording should be accompanied by any materials completed in the session or for homework (e.g. diaries, thought records etc.) and any completed questionnaires.

Formulation. Please attach a copy of the formulation (no more than one A4 sheet). The formulation should demonstrate the therapist's understanding - at that point in time - of how the patient's problems developed (background experiences, beliefs and assumptions, critical incidents) and what is maintaining them (vicious circles).

Session Summary:

- Stage of therapy / session number

  3

- Patient's presenting problem(s)

  bulimia

- Patient's treatment goals

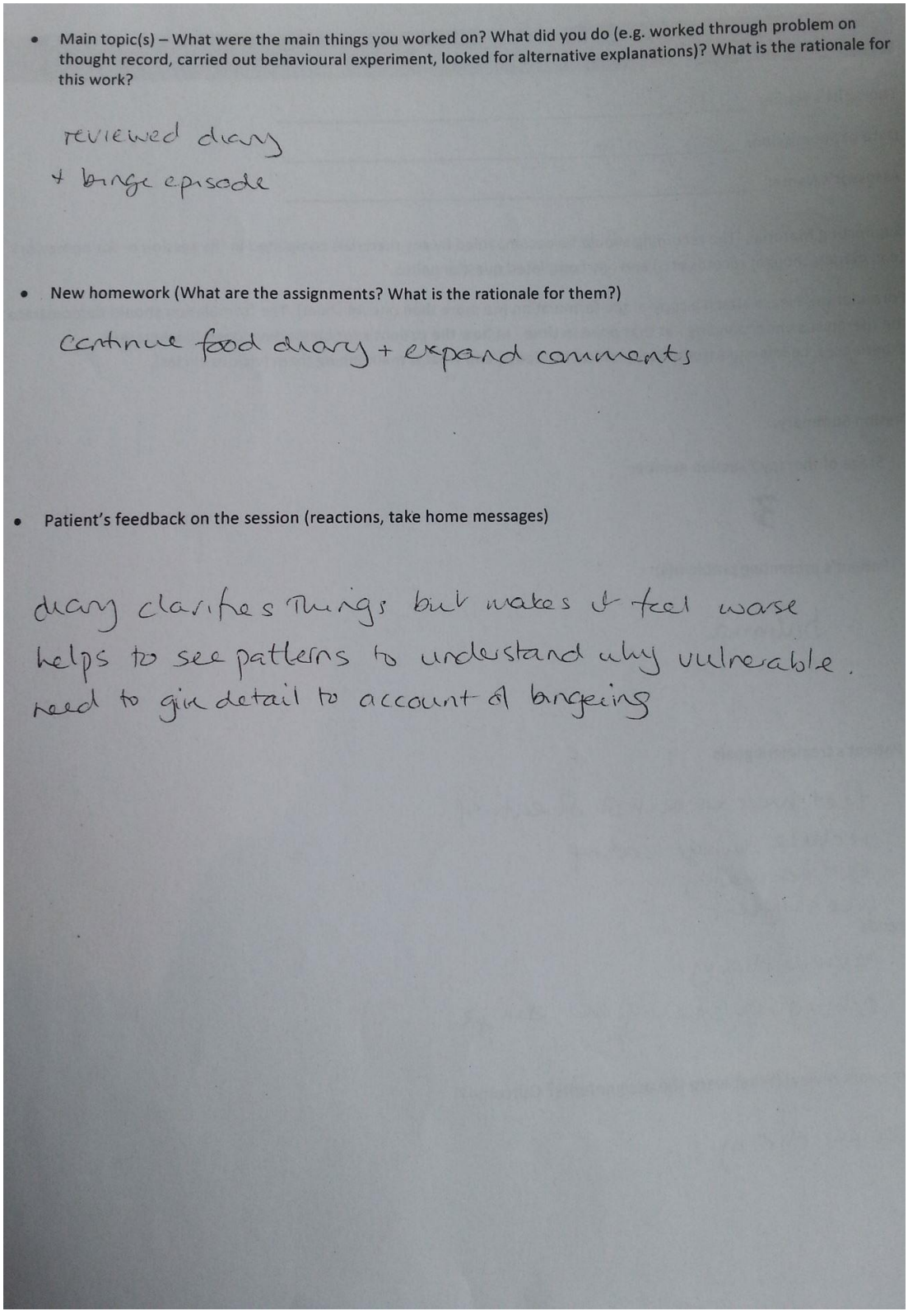
  feel more in control of eating
  reduce binge eating
  eat normally
  like myself

- Agenda

  review diary.
  colleagues seeing he binge

- Homework review (What were the assignments? Outcome?)

  review diary.



- Main topic(s) – What were the main things you worked on? What did you do (e.g. worked through problem on thought record, carried out behavioural experiment, looked for alternative explanations)? What is the rationale for this work?

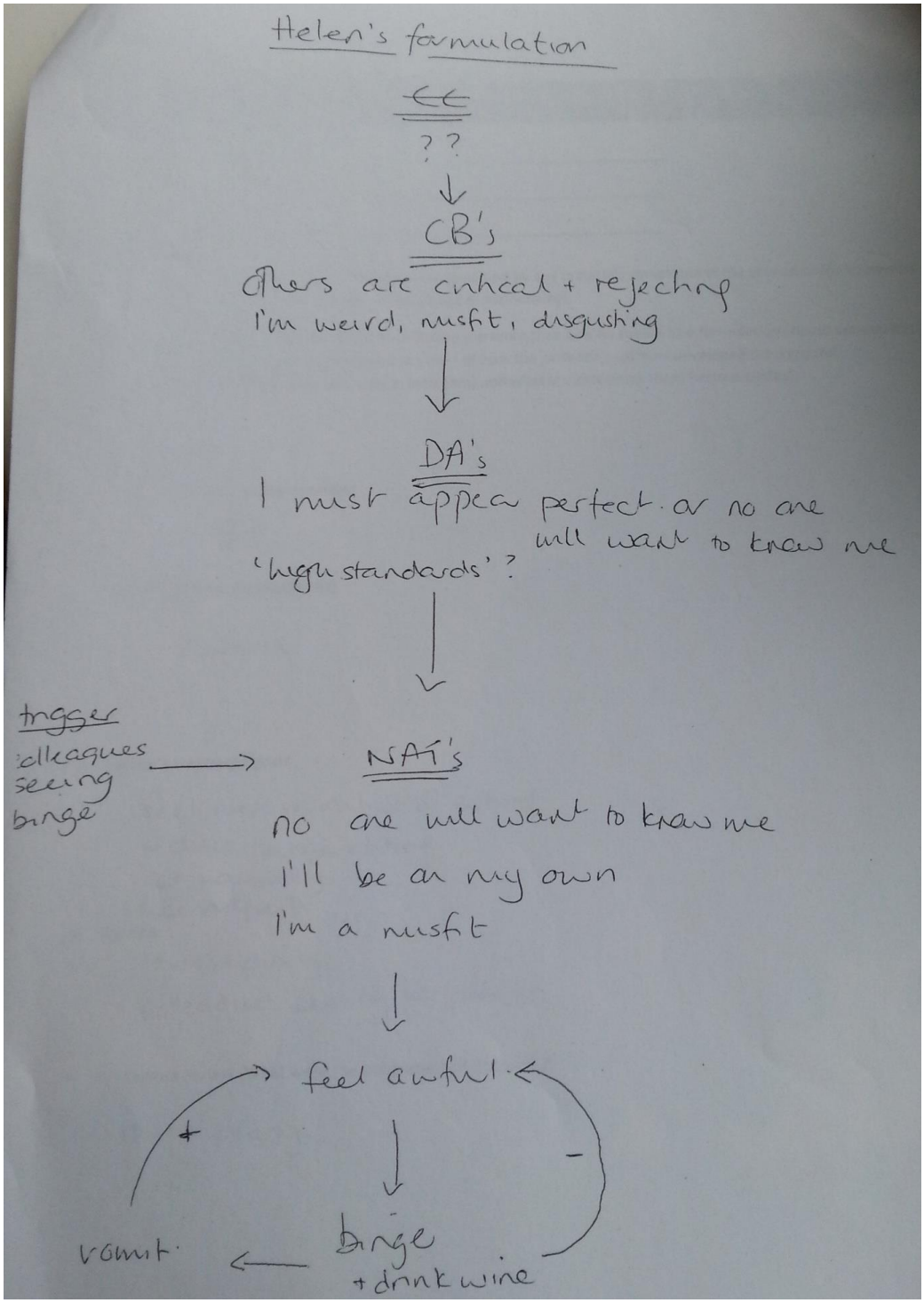
reviewed diary
+ binge episode

- New homework (What are the assignments? What is the rationale for them?)

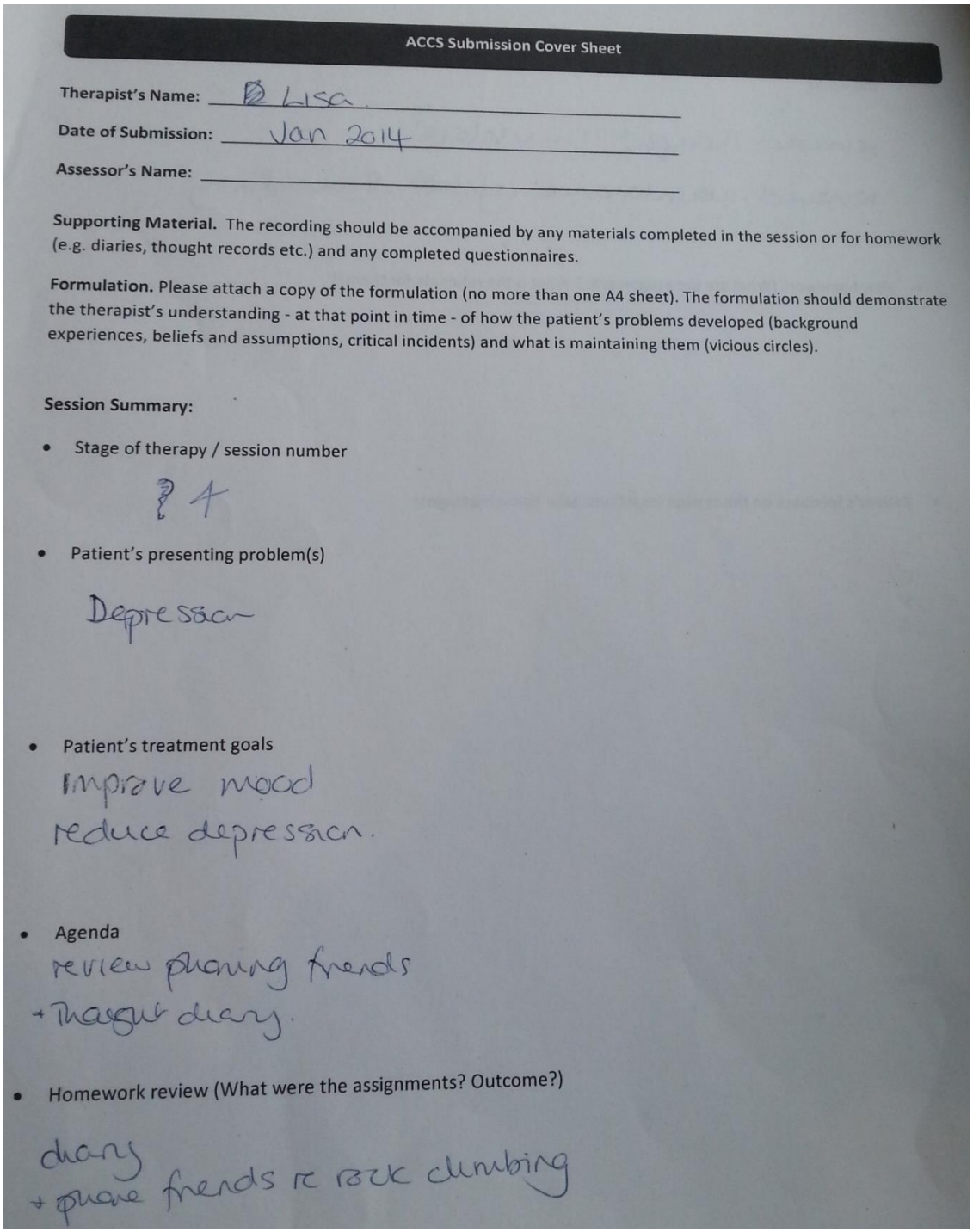
continue food diary + expand comments

- Patient's feedback on the session (reactions, take home messages)

diary clarifies things but makes it feel worse
helps to see patterns to understand why vulnerable.
need to give detail to account of bingeing


Helen's formulation
??
CB's
Others are critical + rejecting
I'm weird, misfit, disgusting
DA's
I must appear perfect or no one will want to know me
'high standards'?
trigger
colleagues seeing binge
NAT's
no one will want to know me
I'll be on my own
I'm a misfit
feel awful
+
-
vomit
binge
+ drink wine

2) Cover sheet to accompany role-play two (post-training)



ACCS Submission Cover Sheet

Therapist's Name: Lisa.

Date of Submission: Jan 2014

Assessor's Name: _____

Supporting Material. The recording should be accompanied by any materials completed in the session or for homework (e.g. diaries, thought records etc.) and any completed questionnaires.

Formulation. Please attach a copy of the formulation (no more than one A4 sheet). The formulation should demonstrate the therapist's understanding - at that point in time - of how the patient's problems developed (background experiences, beliefs and assumptions, critical incidents) and what is maintaining them (vicious circles).

Session Summary:

- Stage of therapy / session number

  ? 4

- Patient's presenting problem(s)

  Depression

- Patient's treatment goals

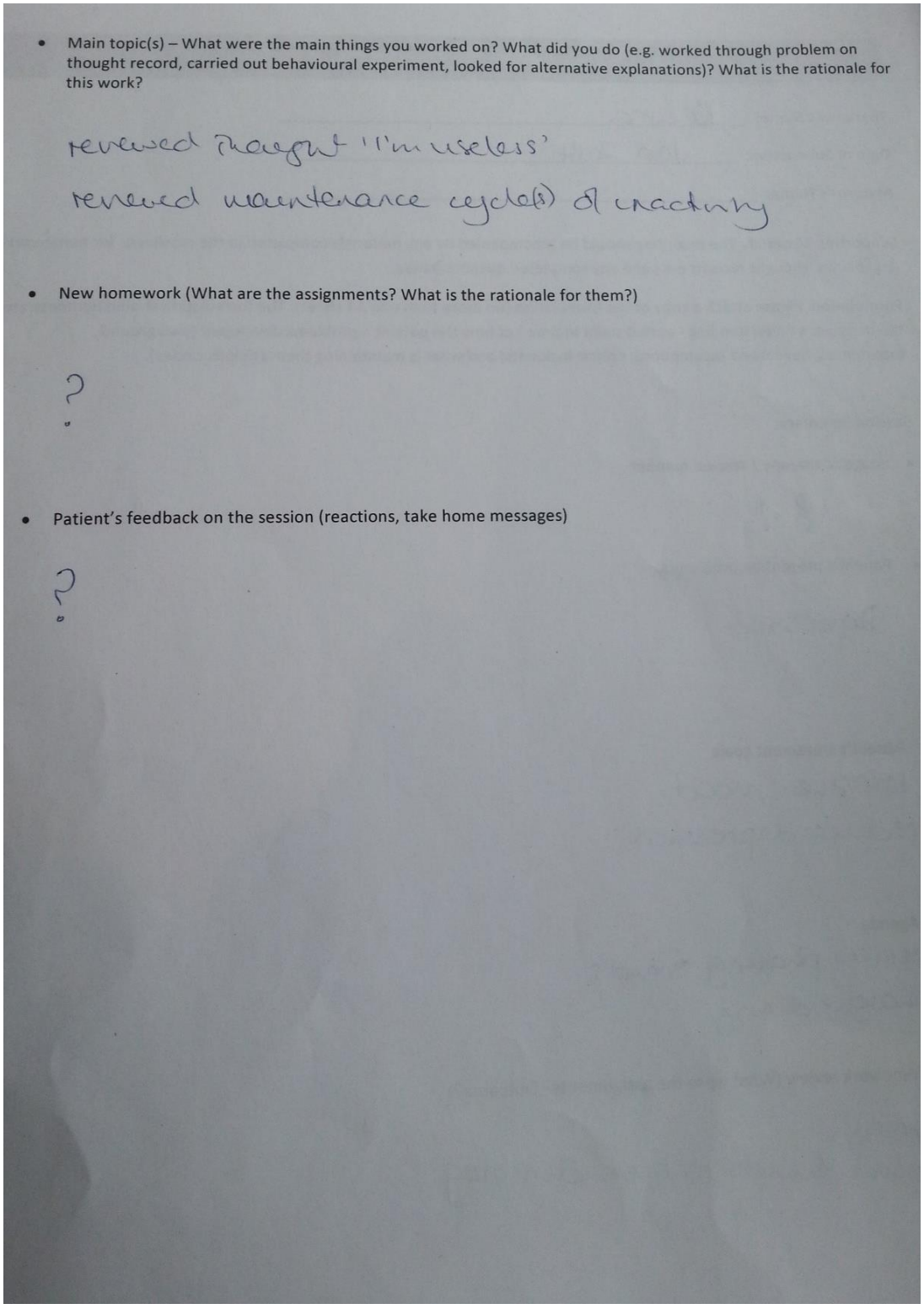
  Improve mood
  reduce depression.

- Agenda

  review planning friends
  + Thought diary.

- Homework review (What were the assignments? Outcome?)

  diary
  + phone friends re rock climbing



- Main topic(s) – What were the main things you worked on? What did you do (e.g. worked through problem on thought record, carried out behavioural experiment, looked for alternative explanations)? What is the rationale for this work?

revised thought 'I'm useless'

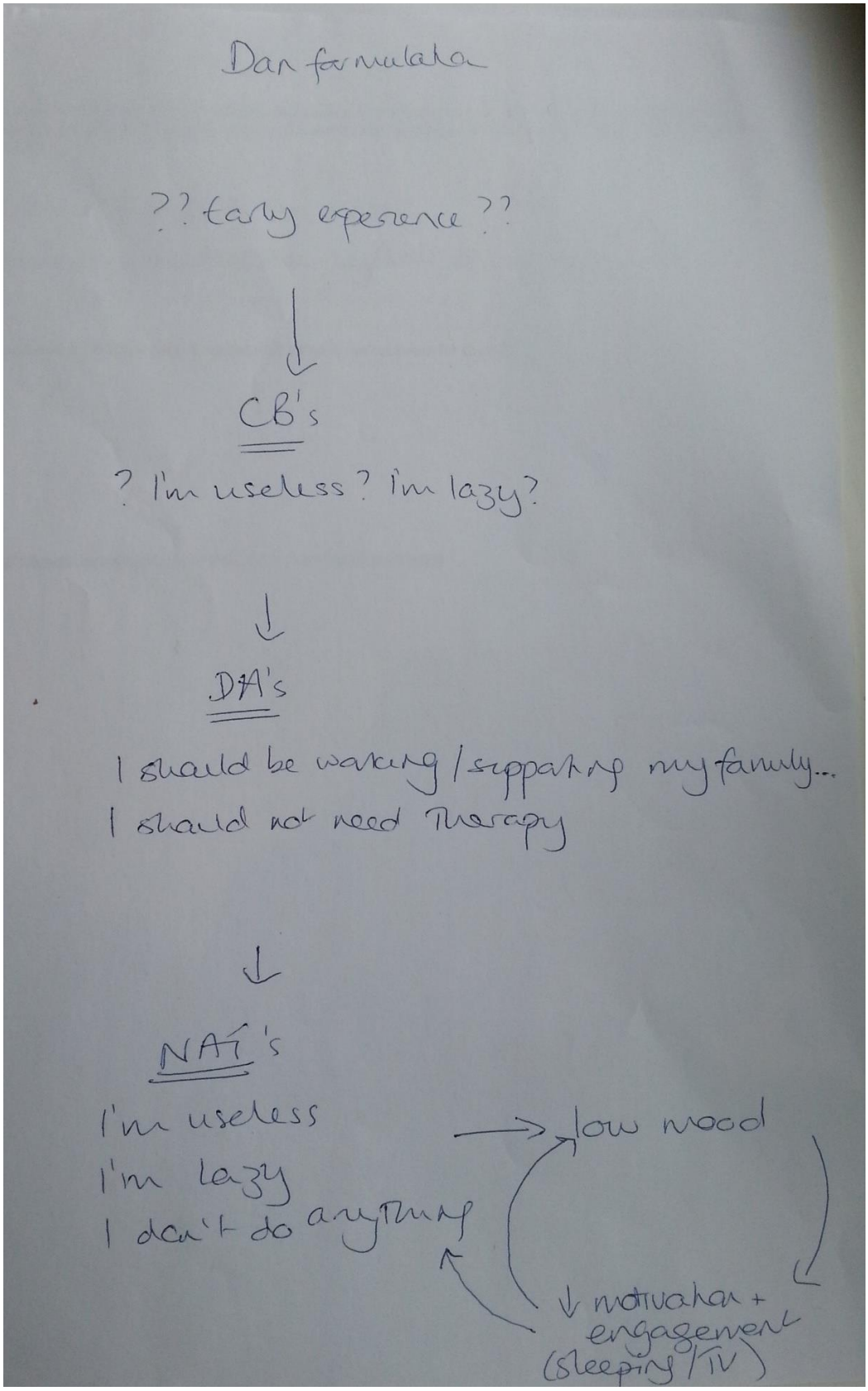
revised maintenance cycle(s) of inactivity

- New homework (What are the assignments? What is the rationale for them?)

?

- Patient's feedback on the session (reactions, take home messages)

?


Dan formulation
?? early experience ??
CB's
? I'm useless ? I'm lazy?
DA's
I should be working / supporting my family...
I should not need Therapy
NAT's
I'm useless
I'm lazy
I don't do anything
low mood
↓ motivation + engagement
(sleeping / TV)

Appendix 4: Feedback questionnaire to determine the usability of the ACCS

ID: _____ **Pre workshop / Post workshop** *(delete as appropriate)*

1. How **relevant** is the ACCS to the assessment of CBT competence?

Please Circle

1 not relevant
2 somewhat relevant
3 quite relevant
4 very relevant

2. How **clear** were the items in the ACCS?

Please Circle

1 not clear
2 somewhat clear
3 quite clear
4 very clear

3. Please rate the overall **style, appearance and layout** of the ACCS.

Please Circle

1 poor
2 fair
3 good
4 very good

4. How **easy** did you find the ACCS was **to use**?

Please Circle

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

5. What is your **overall satisfaction** with the ACCS?

Please Circle

1 not satisfied
2 somewhat satisfied
3 quite satisfied
4 very satisfied

6. What did you **most like** about the ACCS?

7. What did you **least like** about the ACCS?

Appendix 5: Workshop feedback questionnaire

ACCS Training Day- Feedback Form





We would be grateful if you would complete this feedback form before you leave.

The Workshop

Ratings:	Not at all		Moderately		Extremely
Was the content interesting?	1	2	3	4	5
Was the content useful?	1	2	3	4	5
Was the training well structured?	1	2	3	4	5
Was the presentation good?	1	2	3	4	5
Were the handouts useful and relevant?	1	2	3	4	5
Were you offered the opportunity to participate as much as you wished?	1	2	3	4	5
Did the workshop help you to understand how to use the ACCS scale?	1	2	3	4	5

1. The best things about the workshop were:

2. The least satisfactory things about the workshop were:

3. What advice would you give about how the workshop could be run better another time?

Appendix 6: Qualitative feedback about the ACCS scale

Qualitative feedback on the workshop: What participants 'most liked' about the ACCS.

	Pre-training		**Post-training**	
Theme [1]	No. endorsed *n* = 26 [2]	Participant quotations [1]	No. endorsed *n* = 31 [2]	Participant quotations [1]
Ease of use	12	• It is easy to understand and implement. • Clarity. • Clarity. • Very clear to understand. • Clear instructions. • Coloured boxes. • Layout. • Few items overlapping. Fewer items to rate. Less items were open to interpretation. • I like the plain language and accessibility in terms of understanding. What it says is what it is. • Easy to refer back to manual most items self-explanatory. • It just has 8 main features. • The clear separation of items into different domains.	14	• Feels like not too much overlap between items - good that breaks down by sub domains. • Structure + layout. • More flexible than CTS-R. • Easier than the CTS-R. Clear manual. • Clarity of competences. • Clear and easy to understand. • Very clearly explained items. • Layout. • Coloured boxes. • Easy to assess. • The easy use of scale. • More usable. • I thought it was easier to use than CTSR. • Layout, easy language, simple.

Specificity	4	• More specific areas within domains. • Breaking down competencies into specific examples. • The sub-components of the competence domains- they helped breakdown detailed feedback in respect to the therapists specific areas of strength and weakness. • More detail.	4	• Lots of specific information to base your score on. • Domains / sub-domains. • The descriptions. • More detail.
Inclusion of basic therapeutic skills	3	• More generalised • The generic nature is both positive and a drawback especially for trainees. • Increased focus on generalised processes.	1	• More general/broad in relation to CTR.
Comprehensive	2	• Wider range of domains and sub-domains. • Seems quite comprehensive in terms of looking at the various elements of CBT.	3	• Comprehensive. • Incorporates wider range of CBT interventions. • More coherent than the CTS-R.
Formative function	0	•	3	• Feels like a tool for learning and improving practice rather than just rating and scoring. • Makes you think much more critically about your own practice. • Formative. •

Relevant	2	• Relevance and 'fit' to CBT sessions. • Feels more relevant.	1	• It seems to be more relevant (up to date) and allows better flow in therapy rather than trying to make it "fit" for CTS-R.
Misc.	3	• 1-4 scale • Complexity of patient is a feature. • Formative aspects.	5	• Like that patient complexity is recognised. • Recognising complexity of patient. • 4 point scale. • It features 8 domains rather than 12. • Warm phrasing and reflects how therapy is- CTSR bit intimidating and rigid.

[1] themes provide a summary which reflects the core content of feedback, whilst quotations are direct quotes taken from participants' questionnaires.

[2] *n* = number of participants who highlighted one or more thing they 'most liked' about the ACCS scale.

Qualitative feedback on the workshop: What participants 'least liked' about the ACCS.

	Pre-training		**Post-training**	
Theme [1]	No. endorsed *n* = 26 [2]	Participant quotations [1]	No. endorsed *n* = 20 [2]	Participant quotations [1]
Detail: need for practice / time consuming	10	• Need practice. • Thing being the first time using the scale it felt slightly cumbersome but am sure this is just lack of familiarity. • Having a CTS-R head on not being so familiar with this scale - need more time to learn. • Lots to read. Probably improves as familiarise with using more. • Detail- lots to plough through. • Too much to read. • Initially a lot of information but recognise this will become easier with use. • Dense/wordy- but at this stage not familiar with it. • Length, but with more familiarity it would get easier. • The number of items within each domain.	6	• Detailed- will need more time to familiarise. • Several pages. • There is still a lot of it. • Dense. But tell with practice will know it more and feel more confident. • Time consuming which can be difficult within an IAPT service. • Time consuming- such time not always readily available in clinical setting.

Omitted items	3	• Issue re noting use of emotion. • Goals specific item? • One way to improve may be to add additional section which is protocol specific (for feedback purposes). This might not need to be marked so would maintain consistency of the scale. This extra box would just be comments but very specific in relation to overall session.	1	• I miss the specific items around eliciting and using behaviours, cognitions and emotions. I think they could be added in to 3.4 with some benefit.
Misc.	9	• Gap between 2 + 3 on rating scale. • There seemed to be a big jump between 3 and 4. • CBT interventions section quite broad. • Not sure if it offers anything more than CTS-R as still open to interpretation. • Descriptors. I.e. major substantive problems to good degree of skill. • It does not give as many specific examples of what would like to hear/ did you hear any of the following as in the CTS manual- this is one of the main advantages/helpfulness of the CTSR. • Lack of idiosyncratic examples included in manual for how to score each item. •	13	• Its ease of use seems to be dependent on the content of the session. I.e. in the two examples, the second was easier to rate as it was structured more clearly and had clear interventions. The first session (pre-training) was less structured and made the ACCS seem less clear/easy to use. The ACCS was also easier to use when therapist competence was higher and there was less ambiguity about their performance. • The simplicity of the client problems. Interesting to see rating of more complex client and the difficulties that arise. • Some items can appear vague or overlap in the context of particular sessions. • Label 3 as basic skill but with some major substantive problems. • Big gap between 3 + 4.

Misc. continued …		• There doesn't seem to be much space allocated to the type of client the therapist is working with. Sometimes a therapists' performance can be hampered or bolstered by the type of client e.g. reflective, obstructive. • Change- getting used to a new scale!		• That if something is missing from the ACCS that it may be seen as therapy done badly when there is a valid reason for it. Explaining this isn't actual the case on feedback form. • Not having a cut-off for a pass grade. • Makes you think much more critically about your own practice! • It is not as helpful as CTSR in terms of including helpful questions- examples of things would like to hear from therapist did you hear any of the following… • Some points difficult to be clear in my head what a lot of the differences are (e.g. 3.1 / 3.2). • Domain 3- a lot to think about in 5 boxes! • In terms of variance most scores are 3 +/- 1 point. • Change from what I'm used to.

[1] themes provide a summary which reflects the core content of feedback, whilst quotations are direct quotes taken from participants' questionnaires.

[2] *n* = number of participants who highlighted one or more thing they 'most liked' about the ACCS scale.

Appendix 7: Qualitative feedback on the workshop

Qualitative feedback on the workshop: The 'best things' about the workshop.

Theme [1]	No. endorsed *N* = 34 [2]	Participant quotations [1]
Opportunity to practice and engage in group discussions.	27	• To use a new scale and rate and compare others performance/opinions. Discussion points around CBT competences. • Case study- scoring. • Using videos and practicing. • Group discussions. • Group discussion. • Videos then using the ACCS. • Learning about the tool and trying it out. Discussing discrepancies between ratings and reasoning for those differing viewpoints. • Participation. Trying out the scale in practice. • Good to see others' scoring and rationales. • Being able to practice marking using scale and discussions with group following marking. • Practice using the scale. • Opportunity to use the ACCS. • Stimulating discussions. Getting to do the ratings myself. • Watching tapes and scoring/reflecting/discussing. • Plenty of time for debating/ discussion. Actually using the manual and practicing the tool, rather than too much theory. Viewing videos and reflecting on them. Interesting debate with regards to different scores given with the group.

Opportunity to practice and engage in group discussions. Continued …		• Appreciated facilitative approach and involvement- thank you! • Practicing using ACCS and reviewing afterwards • Inclusive. • Practice. • The 'doing' of it! • Interactive style of the workshop. • Worked examples using video, self-reflection and group reflection. • Good interaction. • Very practical and chance to practice. • Comparison with other raters. Reflection of scoring. • Doing the ACCS and reflection. • Useful to go through ratings in a group.
Presentation style	6	• Lovely presentation style by Kate and Freda. • Both presenters engaging and nice to listen to. • Presentations very enthusiastic / motivational. • Clear, informative. • Good pace. • Good presentation. Motivation from presenters.
Misc.	5	• The opportunity to discuss and contribute to the development of this tool. • Opportunity to reflect on CBT practice in a structured way. • Structure of the session. • PowerPoint hand-out good. • Time to understand / study measure. • All round very useful day. Look forward to using this in practice.

[1] themes provide a summary which reflects the core content of feedback, whilst quotations are direct quotes taken from participants' questionnaires.

[2] *n* = number of participants who highlighted one or more 'best thing' about the workshop.

Qualitative feedback on the workshop: The 'least satisfactory things' about the workshop.

Theme [1]	No. endorsed *n* = 12 [2]	Participant quotations [1]
Difficulty concentrating when watching/rating a long treatment session at the end of the day.	5	• Viewing a whole tape after lunch was tough! • One hour video in afternoon- difficult to concentrate. • My own concentration level. • The tapes were helpful- but last session on depression in graveyard hour was difficult! • Lengthy video at end of day not best time to concentrate.
Environmental factors	4	• Heat of room! • Lighting in venue. • Lack of air in the room! • Training environment (hot and stuffy).
Misc.	6	• Volume of paperwork. • Found the group a little large. • Could have done with more time. • Long video finished early as something was wrong with the tape. • Still a lot to become familiar with in a short time! • Videos had fairly straightforward / compliant clients. May be helpful to have less 'textbook' videos.

[1] themes provide a summary which reflects the core content of feedback, whilst quotations are direct quotes taken from participants' questionnaires.

[2] *n* = number of participants who highlighted one or more 'least satisfactory thing' about the workshop. Additionally, four participants said that there was "nothing" that was not satisfactory.

Qualitative feedback on the workshop: Advice about how the workshop could be run better another time.

Theme [1]	No. endorsed *N* = 14 [2]	Participant quotations [1]
More group discussion	2	• More groups/pair discussion. • Works well- maybe could have compared scores for tape 2 also.
Clearer paperwork	2	• Handouts (for exercises) bit confusing / too much- maybe colour co-ordinated or titled different OR clearer OR stapled together to make it simpler? • Give all separate sheets for feedback or video after lunch.
More frequent, but shorter breaks.	3	• Maybe a shorter lunch break. • More frequent short breaks. • Shorter lunch break.
Shorter videos	2	• Video 2 too long for me (but necessary). • Shorter therapy clips.
Misc.	7	• More complex clients • Larger print on handouts? • If it was run over two days- bring a video of your own. • Blank ACCS feedback forms to take away would be helpful. • Encourage people to score a tape using the new scale before the session? • Not sure- a lot of video to assess which is important but quite draining too. • Smaller groups.

[1] themes provide a summary which reflects the core content of feedback, whilst quotations are direct quotes taken from participants' questionnaires.

[2] *n* = number of participants who offered one or more piece of advice about how the workshop could be run better another time. Additionally, three participants said that there was "nothing" they would advise changing.

CHAPTER EIGHT:

Concluding Comments

Overview of Thesis

Cognitive Behavioural Therapy (CBT) has been shown to be effective in treating a wide range of psychological disorders (see Stewart & Chambless, 2009 for a review). The strong evidence-base for CBT has promoted an increased focus on the dissemination of CBT into routine practice settings. As discussed in chapter one, identifying an optimal strategy for assessing the competence with which CBT is delivered is crucial to the success of current drives to expand CBT training and service provision as it provides a means of assessing the training of new CBT therapists and ensuring the quality of treatment provision within routine clinical practice. Comprehensive assessments are also important within the context of the development and maintenance of CBT skills as they promote self-reflection and provide targeted feedback about therapists' strengths and learning needs. Effective tools for assessing competence in delivering CBT also play a crucial role in the empirical evaluation of CBT as research trials cannot draw valid conclusions regarding the efficacy of CBT protocols unless the competence with which the protocols are delivered can be established. Despite the importance of developing effective methods of assessing CBT competence, progress in this field has been somewhat limited to date. Thus, there are at present no evidence-based best practice guidelines outlining the way CBT competence should be assessed. Furthermore, many of the assessment measures that are currently available have been widely criticised, indicating a need for improved tools for assessing CBT competence. To begin addressing this issue, this thesis first focused on reviewing and evaluating current methods of assessing CBT competence (chapters one and two) and then moved on to building upon existing measures for assessing CBT competence in order to provide an improved assessment tool (chapters three to seven). This final chapter provides concluding comments that draw together the results from these thesis chapters and discusses the findings within the broader context of the assessment of CBT competence.

Reviewing and Evaluating Current Assessment Methods

Chapters one and two focused on reviewing and evaluating current methods of assessing CBT competence. These chapters reported results from a systematic review of current methods of assessing CBT competence and a qualitative exploration of experts' experiences of assessing CBT competence. Several recommendations for assessing therapists' competence can be made on the basis of the results of these studies. These recommendations should, however, be viewed as

tentative due to the limited evidence-base identified in the literature review and limitations in generalising results drawn from a qualitative exploration of a selective group of CBT experts to the broader CBT community.

It is clear from the findings outlined that a standardised 'one size fits all' approach cannot be taken when determining how best to assess CBT competence. Instead, tools will need to be developed, selected and implemented according to the specific context and purpose of the assessment. The literature review aims to help researchers, practitioners and trainers alike to make an informed decision about which method or combination of methods to use by providing a summary of the advantages and disadvantages of each of the different available methods. However, the fact that all of the ten assessment methods identified in the review have their own inherent limitations highlights the need for caution in drawing inferences about an individual's competence on the basis of any single assessment method. In fact, given the complex, multi-faceted nature of CBT competence, and the limitations of all methods of assessing competence, multi-method assessments involving multiple direct observations of therapist skill, knowledge-based assessments, and examination of patient outcome may be the ideal way of assessing CBT competence. This conclusion is somewhat troubling when viewed in the context of the results of the qualitative study, which revealed that the way experts typically assessed competence was in stark contrast to this ideal multi-method approach. It will be important to overcome the issues identified by experts in the field as barriers to adopting such 'best practice' approaches, both by developing more feasible and cost-effective assessment packages and by encouraging a shift in culture whereby the allocation of resources to the assessment of CBT competence is more clearly recognised and thus prioritised.

As well as providing insight into which methods can be used to assess CBT competence, findings from these studies also raise a number of questions about the way in which assessments should be implemented. Many of these questions relate to the material that form the basis of assessments. Is it more effective to base assessments on a full course of therapy, on single therapy sessions, or on multiple evaluations of specific aspects of treatment? Should the material be randomly selected from a larger pool of material or should therapists self-select the material they submit for assessment? Does the specific disorder(s) the patient presents with or the complexity of the case presented within the assessment material have an impact on assessment results? These studies also raise questions about who is best placed to conduct competence assessments. Are independent, external assessors the most appropriate people to assess competence or is it more useful to use supervisors who have an ongoing relationship with the therapists? What degree of experience and understanding of CBT is required to assess therapists'

competence? Do assessors require training in order to conduct competence assessments? Furthermore, results from the qualitative exploration indicate that the way in which competence methods are implemented may also have unintended consequences. Different implementation procedures may, for example, impact differently on the level of anxiety therapists' experience about being assessed. More concerning is the idea that assessment procedures could also influence both the quality of therapy delivered and the breadth and quality of the training therapists receive. It is it is hoped that raising these questions will encourage those involved in the assessment of CBT competence to give careful consideration to the way in which assessments are implemented as well as serving to stimulate further research, discussion and debate.

Results from the literature review and qualitative exploration also draw attention to a number of significant limitations with the reliability, validity, and usability of existing tools, thus highlighting the need for improved assessment measures. Of particular importance is the improvement of measures which assess whether therapists can demonstrate the skills necessary to effectively deliver CBT. Three key avenues for further investigation were identified. First, research could further evaluate and improve upon the scales used to rate competence based on supervision sessions. Second, research could develop measures for assessing therapists' performance within standardised role-plays and could also examine the degree to which therapists' ability to perform within artificial clinical scenarios translates to an ability to perform within real clinical encounters. Third, research could develop more comprehensive, valid, reliable and usable scales for assessors to rate therapists' performance within treatment sessions. Regardless of the approach taken, it is important that efforts focus on the development of tools that are feasible, easy to use, and are perceived as credible, useful and valid by those involved in the assessment of CBT competence.

Developing an Improved Assessment Tool

The remaining chapters focused on further developing existing methods of assessing CBT competence in order to provide an improved assessment tool. Specifically, these chapters focused on the development and evaluation of a novel CBT competence rating scale: the Assessment of Core CBT Skills (ACCS). The ACCS provides an assessment framework for assessors to provide formative and summative feedback regarding therapists' performance within observed CBT treatment sessions and for therapists to rate and reflect on their own performance. The central aim of the ACCS is to build upon existing rating scales (particularly the Cognitive Therapy Scale- Revised- CTSR; Blackburn et al., 2001) to provide a tool with improved validity, reliability and usability. In particular, the ACCS aims to: break down broad aspects of CBT competence into

discrete components, provide clearer behavioural anchors for scale points, reduce the degree of ambiguity and assessor inference required, update the content of the scale in light of recent advancements in CBT practice, include additional aspects of CBT competence (e.g. measuring change), increase capacity for formative feedback, and incorporate the use of supporting materials. It is also hoped that the ACCS will be more broadly applicable across Axis 1 disorders and across a range of behavioural and cognitive interventions.

Evaluation of the ACCS demonstrated that the scale had good face validity, content validity, and usability and was viewed by assessors and therapists as a relevant, user-friendly, and useful formative learning tool with clear and specific scale items. A pilot evaluation of the ACCS also showed that both the self-rated and assessor-rated versions of the ACCS demonstrated good internal consistency, inter-rater reliability, and discriminant validity. Additionally, whilst assessor training did seem to be beneficial, simply providing assessors with the ACCS manual was found to be sufficient to achieve acceptable levels of reliability and usability. Thus, the ACCS appears to provide an appropriate, useful and feasible tool for assessing CBT competence. In addition, the ACCS was found to be correlated with but distinct from the CTS-R, was comparable to the CTS-R in terms of internal consistency and discriminant validity, and improved upon the CTS-R in terms of inter-rater reliability. Thus the ACCS appears to provide an acceptable alternative to the CTS-R.

There are, however, a number of limitations to the evaluation of the ACCS presented in this thesis. A replication and cross-validation of the evaluation of the validity, reliability and usability of the ACCS with an independent sample needs to be conducted to compensate for any study specific effects or potential biases in the samples included in the studies. The psychometric properties of the ACCS also need to be evaluated within the context of treatment sessions involving the delivery of behavioural interventions (e.g. behavioural activation) as well as the delivery of CBT to individuals with more complex presentations. It will also be important to collect ratings from a larger sample of therapists and to use this to complete exploratory factor analysis in order to determine the structure of the latent variables underlying the ACCS. Examination of the predictive validity of the ACCS scale (i.e. examining whether the measure predicts improvements in patients' symptoms) will also be a vital further step in evaluating the ACCS. This would also provide insight into which aspects of CBT competence are, in practice, necessary to achieve good patient outcomes. Finally, further consideration needs to be given to the development of a threshold score on the ACCS, above which a therapist can be considered competent. However, research in this domain would need to account for the fact that the level of competence and thus the score on the ACCS that is required may vary according to the context and purpose of the assessment.

It is also important to recognise the limitations of the ACCS in assessing CBT competence. As rating treatment sessions is a method of assessing CBT competence which is already widely accepted and used in practice, it is hoped that the ACCS could be quickly and easily implemented within a range of clinical and training contexts. However, the feasibility of this relatively time consuming and costly method of assessment within the resource constraints of many training settings and routine clinical services will present a challenge to the more widespread use of the ACCS scale. The practical and ethical difficulties that often serve as service level barriers to recording and viewing treatment sessions may also be problematic. A number of aspects of competence are also not assessed by the ACCS. This is because the ACCS is specifically designed to assess whether a therapist has demonstrated the core generic and CBT specific skills required to deliver effective 1:1 CBT to adults with mental health problems within a given treatment session. First, the ACCS does not assess therapists' knowledge or understanding of CBT, aspects of competence which can instead be assessed using multiple choice questionnaires, essays, case reports or clinical vignettes. Second, as the ACCS focuses on 'intervention competence', broader professional skills (e.g. ethical practice, effective use of supervision) are not covered. Third, the scale focuses on competences which are transdiagnostic, rather than skills which are specific to a particular disorder or treatment protocol. Fourth, as the ACCS assesses core CBT skills evident during active, mid-treatment therapy sessions, it does not assess therapists' assessment or relapse-prevention skills. Fifth, the ACCS is for use with individual face-to-face CBT (i.e. not group, family, online CBT) for an adult population experiencing mental health problems.

Despite its limitations, the initial evaluation of the ACCS presented in this thesis is promising and indicates that the ACCS is suitable for assessing the competence with which CBT is delivered. Indeed, the ACCS can be used for a wide variety of purposes. For example, the ACCS can be used as a summative tool for establishing whether therapists have reached an adequate standard of competence. This is particularly useful for accreditation purposes, for evaluating the impact of training, and for selecting and monitoring research trial therapists. The ACCS can also be used to provide detailed formative feedback to aid therapists' development and progression both within supervision and training settings and as part of continuing personal development programmes. Additionally, the ACCS can also be used by therapists to rate their own competence. Although the accuracy of self-ratings is questionable, self-ratings can be a useful tool for fostering self-reflection and highlighting on-going learning needs. As such, it is hoped that the ACCS will provide a useful addition to the toolbox of assessment methods that can be used to assess CBT competence.

References

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. doi: 10.1017/S1352465801004040

Stewart, R. E., & Chambless, D. L. (2009). Cognitive–behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology, 77*(4), 595-606. doi: 10.1037/a0016032