A systematic review of methods for assessing competence in cognitive–behavioural therapy

Kate Muse[a], Freda McManus [b]
[a] Department of Psychiatry, University of Oxford, United Kingdom
[b] Oxford Cognitive Therapy Centre, United Kingdom

Abstract

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 (Miller, G. E. [1990]. The assessment of clinical skills/competence/performance. *Academic Medicine*, *65*, 63–67) framework for assessing clinical skill. The advantages and disadvantages of each method are examined in relation to reliability, validity and feasibility. The limitations of the current evidence base are outlined and priorities for future research are highlighted. Tentative recommendations for assessing therapist competence are made within the context of the limited evidence base and need for feasibility in clinical practice settings.

Keywords: CBT Cognitive behavioural therapy Competence Clinical skill Assessment

1. 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 Behaviour 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 (McHugh & Barlow, 2010). 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; Schoenwald et al., 2011).

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, 2011). Third, as evidence suggests that therapist competence may play a role in determining treatment outcomes, at least in the context of depression (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, results from studies examining the relationships between CBT competence and patient outcome are variable, and the poor reliability of existing methods for 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 may facilitate much needed future research examining the association between competence and outcome in CBT and has the potential to provide 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 (McManus, Rosen, & Jenkins, 2010) and to enable research examining the acquisition of CBT skills, thus informing the training of therapists.

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 (Barber, Sharpless, Klostermann, & McCarthy, 2007; McGlinchey & Dobson, 2003; Sharpless & Barber, 2009; Waltz et al., 1993). Hence this review outlines and evaluates strengths and weaknesses of existing methods for assessing CBT competence in order to make recommendations about the most effective methods and identify priorities for future research into the development of reliable, valid and cost-effective methods of assessing CBT competence.

2. What is CBT competence?

Before examining the utility of different methods for assessing therapist[1] competence in CBT it is necessary to clarify what is meant by the term competence. In line with Barber et al.'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. Roth and Pilling (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 number and varying levels of competences outlined, the authors note that in its current format the framework cannot be used to assess competence, suggesting instead 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., skillful 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 and a hierarchical relationship between the constructs, with adherence being a perquisite for the competent delivery of CBT, while adherence does not necessarily imply competent delivery (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, and it may provide useful information for understanding difficulties in disseminating protocols (Moncher & Prinz, 1991; Perepletchikova & Kazdin, 2005; Schoenwald et al., 2011). 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). Thus the present review considers not only measures of competence, but also those that assess a combination of competence and adherence. Measures that assess adherence alone are however beyond the scope of the review.

3. Method

3.1. Search strategy and procedures

Articles included in the review met the following criteria: (i) English language publication pertaining to a quantifiable CBT competence measure, (ii) relate to individual face-to-face CBT (i.e., not group, family, online CBT) for an adult population experiencing mental health problems, and (iii) not a review or commentary. Peer-reviewed articles published[2] between January 1980 and July 2012 were searched using PsychInfo, Scopus, Science Direct, Psycarticles, Web of knowledge, Medline and PubMed databases using the broad strategy of including any of the following competence-related terms: "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 CBT terms: "CBT", "cognitive behavior/ur", "cognitive behavior/ural", "behavior/ural therapy", "behavior/ur therapy" or "cognitive therapy" (for full search strategies see Appendix 1 online supplemental materials). After removing duplicated references, the

[1] 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.

[2] Unpublished manuscripts or conference presentations were excluded from the search strategy to ensure that no unpublished assessment methods were included in the review. However, discussion of each of the methods for assessing competence that were identified in the search does include reference to a wider literature base, including a small number of unpublished manuscripts and conference presentations.

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 (Centre for Reviews and Dissemination, 2009) was employed independently by two reviewers (KM and FM), with the final selection being agreed by both reviewers. 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). Second, full copies of the remaining 127 studies were retrieved and assessed for eligibility, 64 of which met inclusion criteria[3] (see Fig. 1).

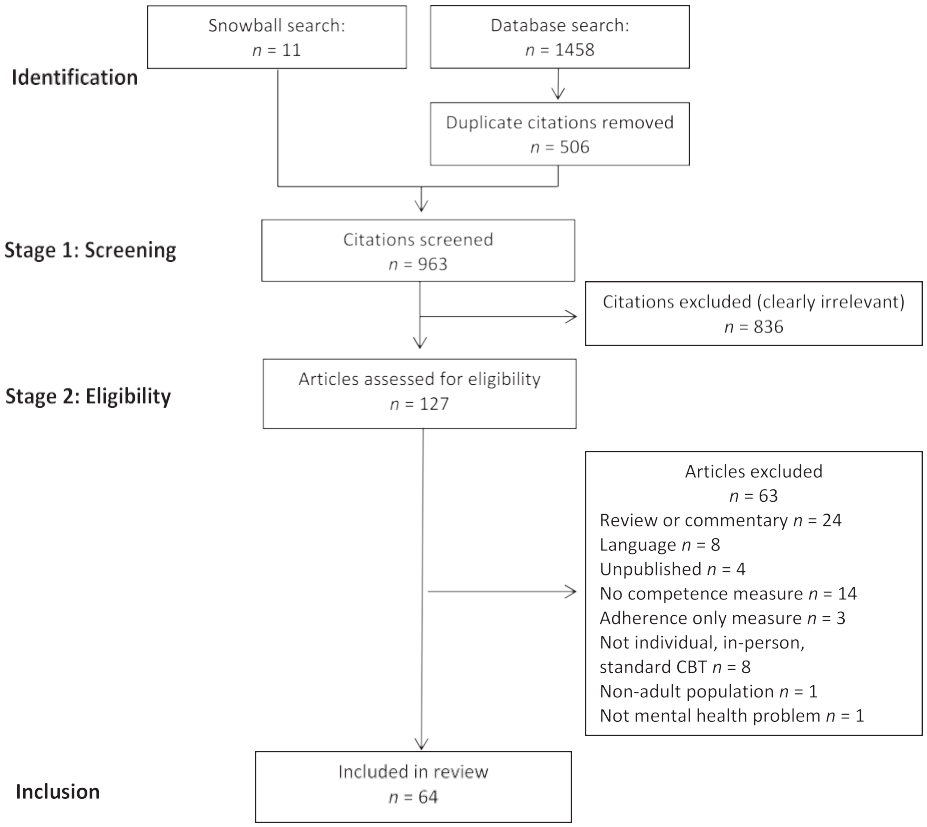

Fig. 1. Flow diagram of study selection procedures.

4. 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 Fig. 2). 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, validity and feasibility.

5. 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). CBT knowledge can be assessed using multiple choice questionnaires (MCQs) or essays (for a summary of studies employing knowledge assessments see online Supplemental materials Appendix 2). As well as assessing CBT knowledge, MCQs and essays can be used to assess knowledge of the clinical application of CBT and thus can also be assessments of practical understanding but for parsimony these methods are only discussed within this section.

5.1. Multiple choice questionnaires

MCQs commonly comprise of a lead-in statement or question followed by a number of responses, from which one or more correct answer(s) is selected. Although development costs can be significant, once developed MCQs are inexpensive to implement as they do not require expert assessors, put little burden on therapists and can assess a range of content. MCQs have been used to measure CBT knowledge acquisition following training (Decker, Jameson, & Naugle, 2011).

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

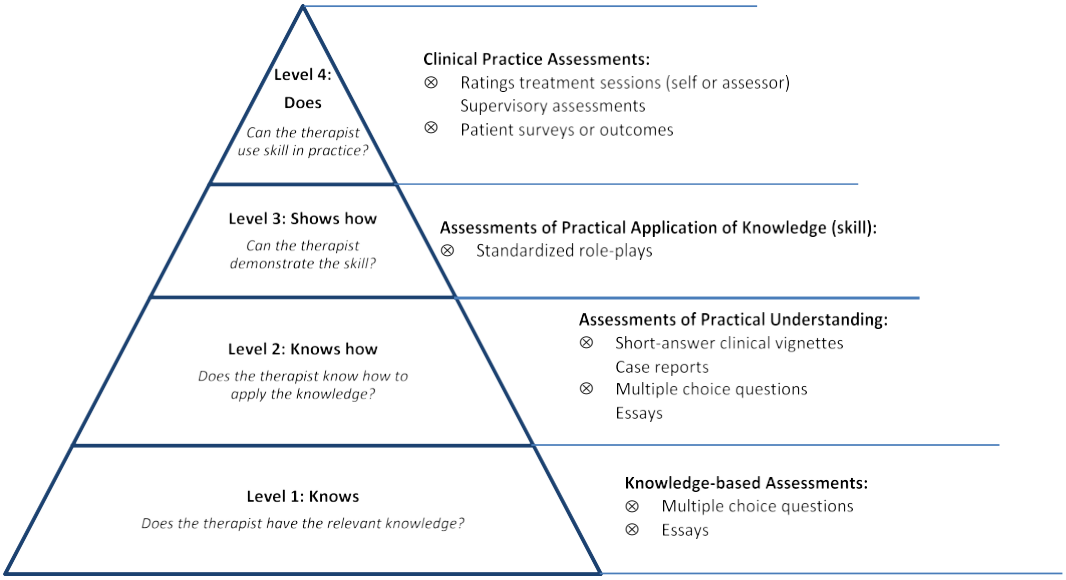

Fig. 2. A framework for CBT therapist competence measures, based on Miller's (1990) clinical skills hierarchy.

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, the CTAS has recently been used 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, but given its intended use with patients, the items assess relatively basic CBT concepts and thus it is a relatively superficial CBT knowledge assessment. The CBT-KQ has 26-items, each with four response options, and provides a more in-depth assessment of CBT knowledge, including the theoretical underpinnings of CBT as well as practical application. 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.

Further research to develop valid and reliable MCQ's for broader use needs to address a number of issues. First, it is important to utilise formats with sufficient contextual detail (e.g., case vignettes) to enable assessment of the higher-order cognitive processes that are characteristic of clinical practice (e.g., understanding, evaluation and application of knowledge) rather than simply testing recall of basic facts which bears little semblance to clinical practice (Case & Swanson, 2001). Second, the ability to obtain a correct answer due to recognition (i.e., 'cueing') needs to be addressed, for example by increasing the number of possible responses or using extended matching items (several questions with a long list of possible answers) (Schuwirth & van der Vleuten, 2004). Third, the bank of questions needs to be large enough to enable different versions of the test to be delivered so as to reduce practice effects on repeat testing. Fourth, although potentially costly at the developmental stage, the use of computer-based delivery and administration could ease implementation (i.e., automatic marking) and enable the inclusion of media within questions (e.g., video clips, formulation diagrams 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 in order to establish the predictive validity of MCQs.

5.2. Essays

Essay questions allow for more contextualised answers than MCQs and require effortful retrieval of information to construct an answer, making them more suitable for assessing higher-order cognitive processes and less subject to recognition effects (Epstein, 2007; Schuwirth & van der Vleuten, 2004). CBT training courses routinely use essays both to enhance and consolidate learning and to assess 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). Essay assessments in medical education settings have been criticised for low reliability and high labor costs (Epstein, 2007; Schuwirth & van der Vleuten, 2004). However, within CBT settings essay assessments have demonstrated acceptable levels of reliability (Keen & Freeston, 2008). This may be because of the use of structured marking procedures (e.g., clear making criteria, multiple markers, independent monitoring of the grading process) which are thought to improve reliability (Leigh et al., 2007). Research suggests that multiple essay assessments are necessary to yield a reliable assessment (e.g., Keen and Freeston (2008) reported that if one examiner was used seven essays were required or if two examiners were used five essays were required). Despite the need for multiple essays and expert assessors, McManus, Keen, and Freeston (under review) estimate a labor cost per therapist of between 3.96 h (one examiner/five submissions) and 7.92 h (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 conference presentation by McManus, Keen, et al. (2010) reported a significant relationship between therapists' essay scores and their patients' post-treatment depression scores ($r = -.18$, $p < .05$), thus providing preliminary evidence of the predictive validity of essays as a method of CBT competence assessment.

6. Level 2: assessments of practical understanding

CBT competence involves not only understanding of CBT theory and research but also the ability to use this knowledge to inform the implementation of CBT interventions (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.

6.1. 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 video vignettes. Although restricted in breadth, depth and complexity of clinical information, vignettes provide standardised, replicable information. Using open-ended questions precludes correctly answering questions due to recognition and allows more in-depth, complex responses. The downside of this however is increased difficulty establishing reliable assessment of responses and thus higher assessment costs (Schuwirth & van der Vleuten, 2004). Although short-answer clinical vignettes are widely used in training, they are rarely used as formal assessments of competence. The present review identified only one standardised short-answer clinical vignette: the Video Assessment Task (VAT: Myles & Milne, 2004). The VAT presents a video of a therapy session dialogue of a patient with panic disorder and asks therapists 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: $r = .94$ [$N = 19$]). Again the relationship to other measures of competence or to patient outcomes remains unknown.

6.2. Case reports

Case reports are written reports in which a therapist explains and reflects upon the process of assessment, formulation and treatment with a CBT patient. They are usually marked by an experienced therapist on a 0–100 scale (>50 indicating adequate competence) in relation to therapists' ability to demonstrate 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 BABCP CBT therapist accreditation process and are commonly used in CBT training settings (e.g., Barnfield, Mathieson, & Beaumont, 2007; Department of Health, 2011; Keen & Freeston, 2008; McManus, Westbrook, et al., 2010). Through promoting self-reflection and providing therapists with detailed feedback, case reports may provide a useful formative assessment tool. However, there are limitations in using case reports as a basis of formal, summative assessment of CBT competence. 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. Consequently it has been 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. (under review) and Keen and Freeston (2008) respectively) are necessary to obtain a reliable assessment of an individual therapists' competence, resulting in a labor cost of between 3.5 h and 21 h (McManus et al., under review). Furthermore, studies have reported no significant progression in 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. Finally, although the present review did not identify any published studies investigating the predictive validity of case reports, one conference presentation reported no significant relationship between therapists' case report scores and their

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, or that measurement error in assessing the various patient presentations in case reports obscures any such relationship. In sum, both the reliability and validity of case reports have yet to be established.

7. Level 3: Assessments of practical application of knowledge (skill)

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 apply this knowledge in clinical situations (Roth & Pilling, 2007). Furthermore Miller (1990) distinguishes between skill-based assessments which occur within carefully constructed clinical simulations ('shows how') and those which occur within real clinical practice settings ('does' independently in practice). First we review assessments using clinical simulations (standardised role-plays), with clinical practice-based assessments being reviewed in the final section.

7.1. Standardised role-plays

Role-plays are artificial simulations of clinical scenarios 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., set an agenda or complete a thought record) within the role-play are then rated by an observer, either via 'live' observation or from a recording, using pre-defined criteria. Medical training settings routinely use role-plays to assess clinical competence in the form of objective structured clinical examinations (OSCEs), which are usually a series of brief encounters with standardised patients that are assessed by independent observers (Epstein, 2007; McNaughton, Ravitz, Wadell, & Hodges, 2008). Although role-plays are also widely used within CBT training settings, they are generally used to demonstrate or practice new skills, rather than as formal assessments. The only role-play identified by the present review was that used by Sholomskas et al. (2005) to evaluate the impact of training on therapists' ability to demonstrate key CBT interventions. This role-play consisted of a one hour treatment scenario with a scripted patient played by an experienced therapist. 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 4 below). The validity and reliability of this role-play have, 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 have the potential to provide standardised assessments which can be replicated or given in differing forms (e.g., pre- and post-training). Furthermore, role-plays overcome some of the practical difficulties faced by clinical practice-based assessments (e.g., informed consent, patient confidentiality) and can be used with therapists who are not yet working clinically. However, there is a danger that role-plays may lead to the simplification of clinical scenarios and presentations, giving rise to a less authentic subjective experience that is not representative of clinical practice (Sharpless & Barber, 2009). These issues may be particularly pertinent when using computer technology, such as virtual reality interactions or pre-recorded clinical scenarios. Furthermore, the feasibility of role-plays is undermined by the resource commitment required for development and implementation (Kaslow et al., 2009).

8. Level 4: Clinical practice assessments

The highest level of competence assessment in Miller's (1990) hierarchy is that of 'does'—being able to use independent judgment and critical thinking to appropriately and effectively deliver CBT interventions within the cultural and organisational context of clinical practice settings (Miller, 1990; Roth & Pilling, 2007). Methods of assessing competence within clinical practice are reviewed below and include: (1) assessor-ratings of treatment sessions, (2) supervisory assessments, (3) therapists' self-assessments, (4) patient surveys, and (5) patient outcome.

8.1. Assessor-rated treatment sessions

CBT competence can be 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 1) under two broad categories: (1) transdiagnostic scales which focus on competences which are not specific to any one diagnosis, and (2) disorder-specific scales which focus on competences required to deliver a disorder-specific protocol. A brief discussion of the individual scales is followed by an overall evaluation of the utility of this assessment method.

Table 1
Summary of assessor-rated measures of CBT competence within treatment sessions.

Scale	Skills assessed	Items[a]	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 = *poor* 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	
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 0 = *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[b]	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
Cognitive Therapy Scale-Psychosis: CTS-Psy (Haddock et al., 2001)	General therapeutic skills (as applied in psychosis)	1. Agenda 2. Feedback 3. Understanding 4. Interpersonal effectiveness 5. Collaboration	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
	CBT specific skills (as applied in psychosis)	6. Guided discovery 7. Focus on key cognitions 8. Choice of intervention 9. Homework 10. Quality of intervention [Overall rating]	
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	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

(continued on next page)

Table 1 (*continued*)

Scale	Skills assessed	Items[a]	Rating
		9. Linking sessions	
		10. Using the manual	
		11. Homework assignments	
Yale Adherence and Competence Scale (YACS: Carroll et al., 2000)	Substance use disorder general therapeutic skills	1. Assess alcohol use	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
		2. Assess cocaine use	
		3. Assess other substances	
	a) assessment	4. Assess psychopathology	
		5. Assess general functioning	
	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 behaviour 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	

[a] 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.

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

8.1.1. 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 1 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, 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)[4] 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; Dobson, Shaw, & Vallis, 1985; Friedman et al., 2004; Hollon et al., 1992; Jacobson & Gortner, 2000; Jacobson et al., 1996; Jarrett et al., 1999; Lopez & Basco, 2011; Shaw et al., 1999; Simons, Gordon, Monroe, & Thase, 1995; Simons & Thase, 1992; Simons et al., 2010; Strunk et al., 2010; Trepka et al., 2004; Vallis, Shaw, & Dobson, 1986; Vallis, Shaw, & McCabe, 1988; Ward et al., 2000), the CTS has also been used to assess the competence of CBT for 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, 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 & Wilson-Smith, 1988; Shaw et al., 1999). 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 versions with more items. 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 et al., 1985; McManus, Westbrook, et al., 2010; Shaw et al., 1999)). The scale was originally divided into 'general' and 'specific cognitive therapy'

[4] 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.

subscales but research has not supported this division (Dobson et al., 1985; Vallis et al., 1986). 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 significant variance in clinician-rated depression, but not 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 or low quality (Vallis et al., 1986) and 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)) to 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 (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, expertise and training of raters, and use of aggregate ratings (for a summary of studies employing CTS see online Supplemental materials Appendix 3). 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 1 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 with- out rater training (*r* =.44) to good following rater training (*r* =.67) (Reichelt et al., 2003). However, Gordon (2006) found only fair agreement between raters (*ICC* = .38) and a discrepancy of more than five points in over half of rating pairs. Similarly, Keen and Freeston (2008) reported significant measurement error in CTS-R scores, with the confidence intervals for more than 90% of trainees' scores overlapping the pass/fail cut-off. Inter-rater reliability for some CTS-R items is poor, even following rater training (range *r* =.26–.62 (Reichelt et al., 2003)). Additionally, a number of CTS-R items do not increase significantly following training (Blackburn et al., 2001; Milne et al., 1999), indicating variable item sensitivity. In sum, initial evidence indicates that the CTS-R may not have fully overcome the limitations of the CTS. However, direct comparisons between the CTS and CTS-R, and 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 1 for scale summary) derived from the CTS and modified to assess CBT for cocaine abuse. The authors propose that the scale can also be used to assess CBT more broadly but the utility of the CTACS 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 et al., 2003, 2004)) and fair to very good for individual items (ICC =.36–.92 (Barber et al., 2003)). The CTACS has also been shown to discriminate CBT from supportive-expressive dynamic therapy or drug counseling (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 (r_s>.90 (Barber et al., 2003)) indicates that they assess the same construct and thus undermines the validity of retaining both ratings. Hence further

refinement and examination of the CTACS is needed.

8.1.2. Disorder-specific scales

Although transdiagnostic scales assess the competences which underpin most CBT interventions they do not assess the disorder-specific strategies and procedures thought to be central mechanisms of change in CBT (Fairburn & Cooper, 2011). Hence disorder-specific scales aiming to assess competence in the delivery of specific CBT protocols have been developed. The first of these scales was the 21-item Yale Adherence and Competence Scale (YACS: Carroll et al., 2000) which measures therapist adherence and skill in the implementation of behavioural treatments for addiction. It includes three 'general' subscales assessing competences common across modalities and three 'treatment specific' scales assessing competences specific to clinical management (CM), twelve step facilitation (TSF) 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). Next, a single-item measure was used to examine competence in the delivery of CBT for panic disorder (Multicenter Collaborative Study for the Treatment of Panic Disorder—Global Competence Item, MCSTPD-GCI; Huppert et al., 2001). 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 on the *MCSTPD-GCI*, which the authors suggest may be because a single item lacks the sensitivity to adequately assess competence. A further disorder specific scale is the 11-item Manual-Assisted Cognitive Behaviour Therapy-Rating Scale (MACT-RS: Davidson et al., 2004), which was proposed to assess competence in the delivery of manual-assisted CBT for the prevention of para-suicide. The scale developers report good inter-rater reliability (*r* =.66) and some predictive validity in that MACT-RS scores were significantly associated with observer-rated patient outcome, although this was only apparent at the 12-month follow-up and did not extend to patients' ratings or reductions in self-harm episodes (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., patient acquiescence). 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 these early disorder specific measures of therapist competence has yet to be established.

A more widely used disorder-specific scale is the 10-item Cognitive Therapy Scale-Psychosis (CTS-Psy: Haddock et al., 2001), an adaptation of the CTS for assessing competence in the delivery of CBT for psychosis. The CTS-Psy has also been used to assess CBT competence with complex cases more broadly (e.g., borderline personality disorder) where differences in patient presentation (e.g., instability of mood) are thought to similarly impact on the way CBT is delivered (Gordon, 2006; Haddock et al., 2003; Haddock et al., 2001). The CTS-Psy uses a modified scaling system in which six micro skills are provided per domain, each of which is marked as being appropriately included (scored one), inappropriately omitted (scored zero) or appropriately omitted (scored 1). The scale focuses on the same general CBT skills as 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 1 for scale summary). The CTS-Psy has been shown to discriminate between CBT and supportive counseling (Lewis et al., 2002; Tarrier et al., 2004; Valmaggia, Van der Gaag, Tarrier, Pijnenborg, & Sloof, 2005), and between therapists who have vs. have not received CBT training (Haddock et al., 2001). The scale authors 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). However, results from Haddock et al. (2001) should be interpreted cautiously as the sample size is limited (*N* = 14) and therapists had low levels of skill, with only 8/14 achieving a score >50%. Furthermore, Gordon (2006) reports only fair inter-rater agreement (*ICC* =.28) and a discrepancy between markers of greater than 5 points 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 due to high correlations between the two scales (Gordon, 2006), 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).

In summary, initial development of some disorder-specific scales has begun and the development of others is underway (e.g., social phobia [David M. Clark, personal communication]). However development is at an early stage and further research is needed to establish the psychometric properties and range of applicability of such scales. Additionally, while disorder-specific scales may be highly relevant for use within research settings where patients are

selected on the basis of diagnosis, the applicability of diagnosis-specific scales is less clear cut in wider clinical practice or training settings where clinicians deliver a variety of CBT protocols to patients experiencing a range of mental health problems (Barber et al., 2007).

8.1.3. 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 directly assessing therapist competence. All of the multi-item scales provide a total score identifying overall skill as well as scores for individual items. 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, for example between different training courses or treatment trials, and are not as subject to practice effects as some other methods. However, there are a number of issues which undermine their utility. Issues relating to the scales themselves are that the ability of many of the scales to provide valid and reliable measures of competence outside of the controlled research settings for which they were developed has not yet been established. Difficulties conceptualizing and defining 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 standardizing assessors' interpretation of scale items through the provision of training (Barber et al., 2007; Gordon, 2006; Reichelt et al., 2003). However, resource constraints may mean it is not always feasible to obtain ratings from multiple trained assessors. Hence improved definitional clarity and item refinement would be useful. Furthermore, there is a lack of consensus about whether adherence and competence should be measured concurrently or independently. Some scales only assess 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 items. Although conceptually distinct, in practice there is much overlap between the constructs (Waltz et al., 1993). Differentiating adherence from 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 are rarely 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 a high degree of 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 (2012) report low agreement between CTACS ratings made by supervisors and independent judges, with supervisors' ratings being more positive. 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' ratings may be more accurate 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 is 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). A second issue in implementation is that, because rating treatment sessions is time consuming and thus costly, competence is typically inferred from rating one or two isolated treatment sessions per therapist. However, recent studies suggest that a much larger 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 (2012) report that between 24 and 30 treatment sessions must be sampled in order to make a reliable judgment about a CBT therapist's competence. Weck, Bohn, Ginzburg, and Ulrich (2011) 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, as therapists often self-select the sessions to be evaluated there may be sampling biases in which they submit their best sessions (or the converse) or sessions with patients who are especially suitable, 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. (under review) for example estimate the labor cost of providing a reliable assessment of clinical skill using the CTS to be between 30 h (one assessor/15 ratings [2 h/rating]) and 76 h (two assessors/19 ratings [2 h/rating]) 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.

8.2. Supervisory assessments

Supervisory assessments of therapist competence are completed retrospectively on the basis of the supervisor's observation of the therapists' performance in supervision over an extended period of time. These assessments may therefore be more representative of therapists' general competence than measures based on a small, usually insufficient, number of isolated treatment sessions. It is considered best practice for supervisors to directly observe session material within supervision, however this may be 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. As they are quick to administer and require few resources, supervisory assessments are commonly used in routine practice and training settings, but these are often informal and unstandardised. Only two standardised supervisory rating scales were found: 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 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. The psychometric properties of the SRF have yet to be evaluated so its reliability and validity as a measure of therapist competence is not known. The ETBF provides global competence ratings of (1) overall evaluation of competence, (2) general cognitive therapy skills, (3) flexibility and (4) general therapeutic/interpersonal skills. Ratings are 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, anchored on where the therapist is perceived to be 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*). Thus the scale does not assess whether a therapist has reached a specified competence threshold and ratings are dependent upon assessors' perceptions of an 'average' therapist. Kuyken and Tsivrikos (2009) report the ETBF to have good inter-rater reliability ($k = .80$), and variable 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 ($r_s = .32$, $p < .15$). Additionally, ETBF ratings were associated with patient outcomes (Kuyken & Tsivrikos, 2009). 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.

8.3. Therapist self-assessment

Self-assessments are rarely employed as formal competence assessment tools within research and training settings, where 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 implemented 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 modeled 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 com- pared to independent and supervisory ratings (Brosan et al., 2008; Carroll et al., 2000; Mathieson et al., 2010; McManus et al., 2011), calling into question the accuracy of therapist self-assessment. Whilst uncertainty regarding the accuracy of therapists' self-perceived competence undermines the utility of self-assessments as a formal competence measure, this need not deter from 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).

8.4. 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 sub-scales (α=.86 and .91 respectively) and 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, 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 an individual therapist's competence.

8.5. Patient outcome

While it is not directly a measure of therapist competence, patient outcome data (i.e., any standardised patient symptom/functioning measure) can be used to infer the competence of the therapist. 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 an individual therapist from patient outcome has significant limitations. First, this method fails to account for patient variables known to impact on patient outcome (e.g., patient severity and complexity, chronicity, life circumstances) (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), and support the predictive validity of measures of therapist competence.

9. Discussion

9.1. Summary of assessment methods

This review identified ten key methods for assessing CBT competence 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 operationalise 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 has been questioned and there is a lack of standardised short-answer clinical vignettes with established psychometric properties. 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, it is questionable whether these methods should be used 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 are thus a priority for 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, it is not a direct measure of therapist behaviour and 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 for assessors to rate therapists' skill (e.g. CTS/R, YACS) 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 CBT competence and the ability to accurately rate one's own CBT 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 needed 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 outside of the controlled research trials for which they were developed has not yet been established. Difficulties defining and disentangling competence, poor inter-rater reliability and limited feasibility also remain concerns. The question of whether the scales should assess disorder-specific protocols or the generic CBT skills that underpin most CBT interventions remains to be addressed. Although disorder-specific scales reflect the nuances of a specific evidence-based 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, and are not always implementing a specified disorder-specific protocol. In sum, although it has proved difficult to operationalise 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.

9.2. 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.

9.3. 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 adequate assessment of and feedback on all aspects of therapist competence. A number of CBT training courses currently use a variety of different assessment methods in order to establish the competence of their trainees (Keen & Freeston, 2008; 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 labor cost per trainee to deliver CTS, case report and essay assessments that meet standard reliability criteria would be between 35.5 h and 63.5 h of assessor time. Thus future research could usefully investigate the optimal combination of methods necessary to provide a thorough assessment of CBT competence.

9.4. Feasible measures

It is prohibitively resource intensive to achieve reliable assessments of CBT competence using many of the methods discussed in the present review. Indeed high cost and labor 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). Future research needs to strike a balance between the need for reliable and valid assessments of CBT competence and the limits on resource availability, in developing cost-effective methods of assessing competence which can be utilised across a range of settings.

9.5. Implementation protocols

There is currently great variation in the way in which CBT competence assessments are implemented, which influences the reliability and validity of the assessments. Protocols outlining the way in which competence assessment methods should be implemented 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 vs. non-random selection [e.g., of non-responders]). Second, it is necessary to determine who is best placed to assess competence. This includes the level of independence of the assessor, 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. This is especially relevant for higher level assessments where a right or wrong answer is not clearly distinguishable. For example, multiple choice questionnaires could be marked accurately and reliably by an assessor with no understanding of CBT (or a computer), whereas determining whether a therapist has selected and skillfully delivered appropriate intervention strategies given the idiosyncratic presentation of their patient is a more complex task.

9.6. Benchmarks

There is a need for validated benchmarks which outline what outcome is required on a given 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, advanced beginner, competent, proficient and expert. This framework outlines stage-specific clinical benchmarks and could 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 the assessment e.g., what is required to pass an introductory CBT training program may be lower than the requirement for accreditation as a CBT therapist.

9.7. Formative and summative assessment

Different assessment methods may be required for different assessment purposes, such as 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 validity and reliability is the 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.

9.8. Conclusions

It is clear that there are significant limitations in the evidence base for the use of existing CBT competence assessment methods. This means that it is currently not possible to make evidence-based recommendations about how best to assess CBT competence. In order to be able to make such recommendations further psychometric evaluation and refinement of existing measures and/or the development of novel assessment tools with validated benchmarking and clear implementation protocols is needed. With regard to priorities for future research, 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 are promising potential alternatives or additions for assessing the higher levels of 'skill' and 'does'. 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.

This research was supported by a grant from the British Association for Behavioural and Cognitive Psychotherapies (BABCP) awarded to Kate Muse and Freda McManus

References[5]

*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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1037/0735-7028.38.5.493.

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. http://dx.doi.org/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.http://dx.doi.org/10.1891/088983907780851586.

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

*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. http://dx.doi.org/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.

*Blackburn, I. M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., et al. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy*, *29*(4), 431–446. http://dx.doi.org/10.1017/S1352465801004040.

*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. http://dx.doi.org/10.1016/j.janxdis.2007.10.002.

*Bouchard, S., Gauthier, J., Nouwen, A., Ivers, H., Vallieres, A., Simard, S., et al. (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. http://dx.doi.org/10.1016/j.jbtep. 2006.08.002.

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

*Brosan, L., Reynolds, S., & Moore, R. G. (2007). Factors associated with competence in cognitive therapists. *Behavioural and Cognitive Psychotherapy*, *35*(02), 179–190. http://dx.doi.org/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. http://dx.doi.org/10.1017/S1352465808004438.

*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. http://dx.doi.org/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., et al. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, *57*(3), 225—238. http://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. http://dx.doi.org/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. http://dx.doi.org/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. 2*. (pp. 757–792) Dordrecht: Kluwer.

Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., et al. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, *1*(2), 81–91. http://dx.doi.org/10.1080/10503309112331335511.

*Crits-Christoph, P., Siqueland, L., Chittams, J., Barber, J. P., Beck, A. T., Frank, A., et al. (1998). Training in cognitive, supportive–expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology*, *66*(3), 484–492. http://dx.doi.org/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. http://dx.doi.org/10.1017/S0033291703001855.

*Davidson, K., Tyrer, P., Tata, P., Cooke, D., Gumley, A., Ford, I., et al. (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. http://dx.doi.org/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. http://dx.doi.org/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*. http://dx.doi.org/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. http://dx.doi.org/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., et al. (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. http://dx.doi.org/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. *The British Journal of Clinical Psychology*, *24*(4), 295–300. http:// dx.doi.org/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. http://dx.doi.org/10.1093/clipsy.bpi046.

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

Epstein, R. M. (2007). Assessment in medical education. *The New England Journal of Medicine*, *356*(4), 387–396.

http://dx.doi.org/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.
*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. http://dx.doi.org/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. http://dx.doi.org/10.1017/S2045796011000199.
*Friedman, E., Thase, M., Kornblith, S., Wisniewski, S., Briggs, M., Rush, A., et al. (2004). The implementation of cognitive therapy in STAR*D. *Cognitive Therapy and Research*, *28*(6), 819–833. http://dx.doi.org/10.1007/s10608-004-0668-1.
*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. http://dx.doi.org/10.1192/bjp.bp.107.043570.
*Gordon, P. K. (2006). A comparison of two versions of the cognitive therapy scale. *Behavioural and Cognitive Psychotherapy*, *35*(3), 343–353. http://dx.doi.org/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. http://dx.doi.org/10.1136/bmj.38636.593461.68.
*Haddock, G., Barrowclough, C., Tarrier, N., Moring, J., O'Brien, R., Schofield, N., et al. (2003). Cognitive—behavioural therapy and motivational intervention for schizophrenia and substance misuse. *The British Journal of Psychiatry*, *183*(5), 418–426. http://dx.doi.org/10.1192/bjp. 183.5.418.
*Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., et al. (2001). An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy*, *29*(2), 221–233. http://dx.doi.org/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. http://dx.doi.org/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., et al. (1992). Cognitive therapy and pharmacotherapy for depression: Singly and in combination. *Archives of General Psychiatry*, *49*(10), 774—781. http://dx.doi.org/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. http://dx.doi.org/10.1037/0022-006X.69.5.747.
*Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., et al. (1996). A component analysis of cognitive–behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*(2), 295—304. http://dx.doi.org/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. http://dx.doi.org/10.1016/s0005-7967(99)00029-7.
*James, I. A., Blackburn, I. M., Milne, D. L., & Reichfelt, F. K. (2001a). Moderators of trainee therapists' competence in cognitive therapy. *The British Journal of Clinical Psychology*, *40*, 131–141. http://dx.doi.org/10.1348/01446650116 3580.
James, I. A., Blackburn, I. M., & Reichelt, F. K. (2001b). *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. *Archives of General Psychiatry*, *56*(5), 431–437. http://dx.doi.org/10.1001/archpsyc.56.5.431.
Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist*, *59*(8), 774–781. http://dx.doi.org/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. http://dx.doi.org/10.1037/a0015833.
Kazantzis, N. (2003). Therapist competence in cognitive–behavioural therapies: Review of the contemporary empirical evidence. *Behaviour Change*, *20*(1), 1–12. http://dx.doi.org/10.1375/bech.20.1.1.24845.
*Keen, A., & Freeston, M. (2008). Assessing competence in cognitive–behavioural therapy. *The British Journal of Psychiatry*, *193*(1), 60—64. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1037/0022-0167.43.1.10.

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

*Lewis, S., Tarrier, N., Haddock, G., Bentall, R., Kinderman, P., Kingdon, D., et al. (2002). Randomised controlled trial of cognitive–behavioural therapy in early schizophrenia: Acute-phase outcomes. *The British Journal of Psychiatry*, *181*(43), s91–s97. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1017/S1352465808004220.

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

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *The American Psychologist*, *65*(2), 73–84. http://dx.doi.org/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. *The British Journal of Clinical Psychology*. http://dx.doi.org/10.1111/j.2044- 8260.2011.02028.x.

McManus, F., Rosen, K., & Jenkins, H. (2010b). 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. (2010c). An evaluation of the effectiveness of diploma-level training in cognitive behaviour therapy. *Behaviour Research and Therapy*, *48*(11), 1123–1132. http:// dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1017/S1352465801001047.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247–266. http://dx.doi.org/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. (2001). *The development of two new instruments for the evaluation of foundation level training in cognitive behaviour therapy*. Paper presented at the Annual Conference for the British Association of Behavioural and Cognitive Psychotherapy, Strathclyde University; Glasgow.

*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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1093/clipsy.bpi045.

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. http://dx.doi.org/

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.cpr.2009.08.011.
Schoenwald, S., Garland, A., Chapman, J., Frazier, S., Sheidow, A., & Southam-Gerow, M. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 32–43. http://dx.doi.org/10.1007/s10488-010-0321-0.
*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. http://dx.doi.org/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., et al. (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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.cpr.2008. 09.008.
*Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology*, *67*(6), 837–846. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1037/0022-006x.63.3.369.
*Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K., et al. (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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1176/appi.ap. 27.3.154.
*Tarrier, N., Lewis, S., Haddock, G., Bentall, R., Drake, R., Kinderman, P., et al. (2004). Cognitive—behavioural therapy in first-episode and early schizophrenia. *The British Journal of Psychiatry*, *184*(3), 231–239. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1037/0022-006X.61.4.620.

*Ward, E., King, M., Lloyd, M., Bower, P., Siddald, B., Farrelly, S., et al. (2000). Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy, and usual general practitioner care for patients with depression: Clinical effectiveness. *British Medical Journal*, *321*(7273), 1383–1388. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. (2011c). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *The Journal of Nervous and Mental Disease*, *199*(4), 276–279. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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*. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1017/S0141347300014075.

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