UNIVERSITY OF
GLOUCESTERSHIRE

This is , post-print (final draft) version of the following published document, This is the peer reviewed version of the following article: Muse, Kate and McManus, F (2016) Expert Insight into the Assessment of Competence in Cognitive behavioural Therapy: a Qualitative Exploration of Experts' Experiences, Opinions and Recommendations. Clinical Psychology and Psychotherapy, 23 (3). pp. 246-259. doi:10.1002/cpp.1952 which has been published in final form at https://onlinelibrary.wiley.com/doi/10.1002/cpp.1952 This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited. and is licensed under All Rights Reserved license:

Muse, Kate ORCID: 0000-0001-5824-1841 and McManus, F (2016) Expert Insight into the Assessment of Competence in Cognitive behavioural Therapy: a Qualitative Exploration of Experts' Experiences, Opinions and Recommendations. Clinical Psychology and Psychotherapy, 23 (3). pp. 246-259. doi:10.1002/cpp.1952

Official URL: https://onlinelibrary.wiley.com/doi/10.1002/cpp.1952
DOI: http://dx.doi.org/10.1002/cpp.1952
EPrint URI: https://eprints.glos.ac.uk/id/eprint/10998

Disclaimer

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT.

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

Kate Muse[1] and Freda McManus[2]

[1] Psychology, Institute of Health and Society, University of Worcester, Worcester, UK

[2] Psychology, University of Stirling, Stirling, UK

To offer insight into how cognitive–behavioural therapy (CBT) competence is defined, measured and evaluated and to highlight ways in which the assessment of CBT competence could be further improved, the current study utilizes a qualitative methodology to examine CBT experts' (*N* = 19) experiences of conceptualizing and assessing the competence of CBT therapists. Semi-structured interviews were used to explore participants' experiences of assessing the competence of CBT therapists. Interview transcripts were then analysed using interpretative phenomenological analysis in order to identify commonalities and differences in the way CBT competence is evaluated. Four superordinate themes were identified: (i) what to assess, the complex and fuzzy concept of CBT competence; (ii) how to assess CBT competence, selecting from the toolbox of assessment methods; (iii) who is best placed to assess CBT competence, expertise and independence; and (iv) pitfalls, identifying and overcoming assessment biases. Priorities for future research and ways in which the assessment of CBT competence could be further improved are discussed in light of these findings.

Key Practitioner Message:

- A qualitative exploration of experts' experiences, opinions and recommendations for assessing the competence of CBT therapists.
- Semi-structured interviews were conducted and analysed using interpretive phenomenological analysis.
- Themes identified shed light on (i) what to assess; (ii) how to assess; (iii) who is best placed to assess; and (iv) common pitfalls.
- Priorities for future research and ways in which the assessment of CBT competence could be further improved are discussed in light of these findings.

Keywords: CBT, Cognitive–Behavioural Therapy, Competence, Clinical Skill, Assessment, Qualitative

Introduction

Cognitive–behavioural therapy (CBT) intervention competence can be defined as the degree to which a therapist demonstrates the general therapeutic and treatment-specific knowledge and skills required to appropriately deliver CBT interventions that reflect the current evidence base for treatment of the patient's presenting problem (Barber, Sharpless, Klostermann & McCarthy, 2007; Kaslow, 2004). Roth and Pilling (2007) provide a comprehensive framework outlining the key competences required to deliver effective CBT. These competences are grouped within five domains: generic therapeutic competences (e.g., mental health knowledge and patient engagement), basic CBT competences (e.g., knowledge of CBT principles and ability to explain CBT rationale), specific CBT techniques (e.g., guided discovery and thought records), problem-specific competences (e.g., interventions in disorder-specific treatment manuals) and metacompetences (e.g., ability to select and apply the most appropriate CBT method).

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

Despite the need for accurate and efficient methods of assessing CBT competence, progress has been somewhat limited. A recent review of the methods for assessing CBT competence identified 64 articles pertaining to a method of assessing competence in the provision of standard CBT interventions to adults experiencing mental health problems (Muse & McManus, 2013). Ten key methods for assessing CBT competence were identified from these articles, all of which require further evaluation or refinement. In principle, therapists' knowledge is the easiest aspect of CBT competence to assess. However, there are few standardized multiple-choice questions with demonstrated validity and reliability, and the validity of essays has not yet been established. The reliability, validity and utility of case reports have also been questioned, and there is a lack of standardized short-answer clinical vignettes with established psychometric properties. Establishing whether therapists have the skills to apply their knowledge in clinical situations has proved especially challenging. Some of the methods that have been used to assess clinical skill have significant limitations and thus are of limited usefulness. For example, patient outcome data provide an indirect, confounded and limited indication of therapists' competence. Questions have also been raised regarding the feasibility and validity of inferring therapist competence from patient surveys. Additionally, the ability to accurately rate one's own

CBT competence is highly questionable. Supervisory assessments have the potential to provide more robust assessment of therapists' clinical practice, although further psychometric evaluation and refinement of currently available measures is needed. The use of standardized role-plays also has potential, but little progress has been made in this domain to date. The 'gold standard' is assessor-ratings of therapists' in-session performance. However, this method is also problematic due to the lack of empirically evaluated rating scales with adequate reliability, validity and feasibility.

In sum, a range of different methods can be used to assess different aspects of CBT competence, all of which have their own inherent advantages and disadvantages. Given the limitations of the evidence base for the use of existing CBT competence assessment methods, there are currently no clear guidelines as to how CBT competence should be assessed. Thus, little is known about when,

why and how a given assessment method, or combination of methods, is selected and used in practice. The current study uses a qualitative methodology to shed light on this issue by examining experiences of, opinions about and recommendations for assessing the competence of CBT therapists from the perspective of CBT experts with a background in assessing competence.

A qualitative exploration of the way in which CBT competence is assessed has potential to contribute to the further development of CBT competence assessment in a number of ways. First, it has been argued that there is need for further refinement of existing measures and/or the development of novel tools for assessing CBT competence (Fairburn & Cooper, 2011; Muse & McManus, 2013). When developing assessment measures, it is important to consider pragmatics, such as feasibility, usability and acceptability as, without these features, measures are unlikely to be adopted in practice (DeVellis, 2012). In describing patterns that emerge repeatedly throughout the data set, qualitative enquiry identifies recurrent, unifying concepts that characterize the experiences of a number of individuals (Patton, 2002; Smith, Flowers & Larkin, 2009; Sofaer, 1999) and thus could usefully identify which methods experts generally view as feasible, acceptable and applicable across a broad range of situations. This would enable future efforts to refine current assessment methods to focus on those methods that are most likely to be widely implemented in practice. Drawing upon the collective experience of a number of individuals with practical involvement in assessing CBT competence may also highlight problems and pitfalls with existing tools, as well as yielding useful information as to how assessment measures could be further developed and refined.

Second, frameworks that delineate the knowledge and skills required to competently deliver CBT (e.g., Bennett-Levy, 2006; Rodolfa *et al*., 2005; Roth & Pilling, 2007) provide a comprehensive definition of CBT competence. However, they also draw attention to the fact that assessing CBT competence is far from straightforward. Indeed, the models suggest that assessors must consider multiple inter-related dimensions in order to determine a therapist's competence. Qualitative research provides rich and nuanced descriptions of phenomena and is, therefore, well suited to increasing understanding of complex, multifaceted phenomena, such as CBT competence (Patton, 2002; Smith *et al*., 2009; Sofaer, 1999). Thus, qualitative enquiry has the potential to highlight the ways in which assessors draw upon these multifaceted definitions when examining the competence of CBT therapists. Third, sharing the experiences of experts could improve awareness and understanding of the practical issues involved in assessing CBT competence. For example, qualitative enquiry could help to identify any difficulties that commonly occur during the assessment process, together with ways these problems can be overcome. Finally, as well as revealing commonalities between experts in the field, qualitative enquiry could also be helpful in highlighting areas where experts differ in their experiences, viewpoints or opinions (Patton, 2002; Smith *et al*., 2009; Sofaer, 1999). By exploring both convergence and divergence between expert opinion and experience, qualitative enquiry could also serve to stimulate discussion and debate about how to assess CBT competence.

To summarize, the current study utilizes qualitative methodology to explore the experiences, understandings and views of assessing the competence of CBT therapists from the perspective of 19 CBT experts with a background in assessing competence. In line with the open-ended nature of qualitative enquiry, the study had three broad aims: (1) to explore what participants looked for when assessing therapists' competence (i.e., how they defined CBT competence); (2) to examine the way in which participants went about assessing the competence of CBT therapists; and (3) to reveal any difficulties that participants experienced during the assessment process.

Method

Expert participants

'Experts' were broadly defined as experienced CBT therapists whose research or clinical activities focused on the assessment and evaluation of CBT competence. Individuals were identified as experts through their involvement in the development and provision of British Association for Behavioural and Cognitive Psychotherapies-accredited CBT training courses and/or recent publications related to training or assessment of CBT therapists. Given the absence of an ideal sample size for qualitative research (Smith *et al*., 2009) and the large variation in the number of participants included in previous interpretative phenomenological analysis (IPA) studies (range = 1–42; Reid, Flowers, & Larkin,

2005), the present study sought to recruit a comparatively large sample size in order to ensure that the study captured the experiences and opinions of a broad range of experts across the field. Twenty-six CBT experts from the UK were invited to participate, of which five did not respond and two declined. Thus, 19 participated. The 10 female and 9 male participants ranged in age from 40 to 63 years (M = 51, standard deviation [SD] = 7.45) and had been practising CBT for between 13 and 34 years (M = 23.68, SD = 6.66). Sixteen participants had treated over 200 CBT cases, with the remaining three having treated 50–200. Fourteen were clinical psychologists, four were mental health nurses and one was a psychiatrist. Participants were evenly distributed between those working primarily within training ($n = 7$), routine practice ($n = 7$) and research ($n = 6$).

Data collection

In line with recommendations of Smith *et al*. (2009) and Kvale (1996), individual semi-structured interviews consisting of open-ended questions and minimal prompts were used to invite participants to provide a narrative description and reflection on their experience of assessing CBT competence (questions outlined in Table 1). Interviews lasted between 28 and 60 min and were recorded and then transcribed in full.

Table 1. Summary of interview schedule

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

Note: At the end of the interview, all participants were also asked, 'is there anything we haven't spoken about which you feel would be relevant or important?'
CBT = cognitive–behavioural therapy.

†The schedule was used flexibly to guide the interview, with participants being encouraged to take a strong role in leading the discussion. Thus, the prompts outlined here serve as examples, with each interview differing in specific content depending on the experiences of the participant being interviewed.

Data analysis

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

In line with IPA guidelines (Smith *et al*., 2009), a cyclical analytic approach was employed using a qualitative analysis software package (NVivo). First, a close line-by-line analysis of the first interview transcript was conducted in order to ensure familiarity with the participant's account. Initial descriptive and conceptual comments were noted along with any statements that appeared particularly relevant or pertinent to the assessment of CBT competence. Second, any significant, illuminating or revealing extracts from the transcript were isolated and grouped together to create a list of emerging themes. These two stages were repeated for each transcript in turn, with the emerging thematic structure being updated and adapted in light of new information. Where a new theme was identified, all of the

transcripts were re-examined to identify whether similar ideas had emerged within any other participant's account. Third, the list of themes was reviewed, and themes identified within individual transcripts were compared in order to identify any meaningful commonalities or differences between participants' accounts. Where appropriate, themes were restructured, amalgamated and subsumed within a superordinate theme in order to create overarching 'higher-order' or 'cross-therapist' themes, which brought the lower-level categories together in a meaningful way.

The first author (K. M.) took the lead in analysing the data, and each stage of analysis was overseen by the second author (F. M.), who verified that the analysis had been systematically achieved and was supported by the data. Validity checks included re-examining all transcripts in light of the final thematic structure, examining the number of themes assigned to each transcript to ensure adequate representation of all participants, reviewing extracts assigned to each theme to ensure the analysis was representative of the data and inviting participants to review and comment on the results of the analysis (i.e., respondent validation).

Results

Four superordinate themes emerged from data analysis. These are outlined in Table 2 and described, interpreted and illustrated with exemplar quotations[1] below.

Table 2. Summary of themes

Superordinate themes	Subordinate themes
1. What to assess: the complex and fuzzy concept of CBT competence	1.1 More than a CBT technician
	1.2 The challenge of diversity under the broad CBT umbrella
	1.3 Generic versus disorder-specific competence
	1.4 Scales used to assess CBT competence have come to define it
2. How to assess CBT competence: selecting from the toolbox of assessment methods	2.1 Assessing knowledge and understanding
	2.2 You can talk the talk but can you walk the walk?
	2.3 Gold standard versus feasibility
3. Who is best placed to assess CBT competence: expertise and independence	3.1 Assessor experience and expertise
	3.2 The pros and cons of supervisors as assessors
4. Pitfalls: identifying and overcoming assessment biases	4.1 Some patients will make stars of us all
	4.2 Therapist self-presentation biases

CBT = cognitive–behavioural therapy.

What to assess: The complex and fuzzy concept of cognitive–behavioural therapy competence

In order to discuss participants' experiences of assessing the competence of CBT therapists, it was first necessary to establish what they understood CBT competence to be. In line with the concept of 'limited-domain intervention competence' (Barber *et al*., 2007; Kaslow, 2004), participants focused on therapists' ability to appropriately, effectively and skilfully deliver evidence-based CBT treatments. Hence, broader aspects of CBT competence (e.g., use of supervision and ethical practice) were typically not assessed. Discussion about what participants looked for when assessing therapists' ability to effectively deliver CBT yielded four sub-themes (outlined below), which highlight the challenges and complexities involved in defining and disentangling the fuzzy concept of CBT competence.

More than a cognitive–behavioural therapy technician

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

> There's an aspect of CBT that often does get dropped in favour of it being seen as a series of techniques … this sense of a fundamental philosophy … The tragedy is that people have started to see a lot of CBT as a series of activities—do a thought record, ermm, do this, do that . . . people lose sight of the fact that, ermm, well look why are you doing this? P11[2]

In assessing competence, participants looked for evidence that therapists were not overly procedural 'CBT technicians' (P5) mechanically applying CBT techniques. Participants identified three competences that differentiated a competent CBT therapist from a 'CBT technician'. First, CBT interventions should be embedded within general

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

[2] 'P' is used throughout the results as an abbreviation for 'participant', and the number represents the participant's number.

clinical skills that echo the 'generic therapeutic competences' outlined by Roth and Pilling (2007) (e.g., interpersonal/two-way communication skills, empathy, warmth, ability to foster and maintain a sound therapeutic relationship). Second, therapists should make treatment choices (e.g., which, when and how to deliver interventions) on the basis of a sound understanding of CBT theory and research. Third, a competent CBT therapist should have the capacity to adapt and respond to their patient's needs, both in terms of the individual personality of the patient and the features of the disorder with which they present. This flexibility should be apparent in their selection and delivery of interventions as well as in their interpersonal behaviour:

> I think a good therapist is a bit like a chameleon, you know. Depending on what environment they're in or what disorder they've got in front of them, they adjust their behaviour. P14

The ability to deliver CBT techniques within the context of all three competences in combination was viewed as indicative of CBT competence:

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

In sum, participants noted that it is important to avoid narrowing the focus of assessment too tightly on therapists' ability to perform a number of brief technical skills, thereby precluding the comprehensive assessment of therapists' ability to deliver techniques skilfully, flexibly and appropriately in line with the individual patient's formulation, CBT theory and research.

The challenge of diversity under the broad cognitive– behavioural therapy umbrella

Participants noted that the broad range of approaches under the 'CBT umbrella' has yielded disagreement about what CBT is, the way it should be applied and thus what competence is. As P12 commented, 'there are lots of different ways to get to the top of the mountain'. Participants observed varying opinions regarding best practice amongst therapists working within the same protocol. However, participants commented that disagreement is especially apparent when patients do not clearly fit a specific diagnosis, have co-morbid diagnoses or present with disorders for which there are not yet agreed protocols or for which there are several evidence-based protocols. For example, P3 commented on the debate surrounding the treatment of patients who do not clearly fit a specific diagnosis:

> With an ordinary clinical patient who may not come in with pure diagnosis … there are different views about how CBT should be done … ranging from people who probably advocate that, well, you do one protocol and then you do another … through to people who would use a more formulation driven CBT or schema-focused CBT to work with more or less anything. P3

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

Generic versus disorder-specific competence

Participants recognized the tension between assessing generic CBT competences and assessing competence within the context of a specific protocol (i.e., basic versus problem-specific competences; Roth and Pilling, 2007). Some argued that it is not possible to adequately assess CBT competence without focusing on a specific protocol:

> I can't see there being a useful single measure of competence in CBT. Now I can see that there are going to be commonalities … There might be, sort of, some core constructs that you'd want to assess that would be shared by these approaches. But there would also be some protocol-specific features that … wouldn't be relevant to another protocol. P13

Other participants questioned whether a protocol-specific approach was necessary or realistic outside a research context. Participants raised concerns that this approach would result in multiple scales or 'unwieldy instruments' (P11) and would be less applicable in routine practice where patient's co-morbidity and complexity are common and where therapists do not always deliver protocol-driven CBT:

> It's mainly those transdiagnostic things that we're concerned about. And I would think they [transdiagnostic things] would be the most important in assessing competence really because … if you've got those building blocks in place, it allows for

> flexibility and, when you meet a complex case or you meet somebody who doesn't fit in a particular box, you've still got your back to basics core skills that you can apply. P8

Thus, whilst all participants noted the importance of assessing generic CBT competences, there was uncertainty about whether the protocol should merely be taken into account during assessment or whether this should be a central focus, with opinions being somewhat dependent on the assessment context.

Scales used to assess cognitive–behavioural therapy competence have come to define it

Participants were aware that defining and operationalizing the skills involved in CBT competence were far from straightforward:

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

Rating scales, particularly the Cognitive Therapy Scale and the Revised Cognitive Therapy Scale (CTS-R: Blackburn *et al*., 2001; CTS: Dobson, Shaw, & Vallis, 1985), helped participants to operationalize the specific skills a competent CBT therapist should display. However, participants observed that these rating scales have now come to define CBT competence for many. It is widely recognized that assessment drives learning and that learners will employ strategies to optimize their chances of passing assessments (Van der Vleuten *et al*., 2010). It is, therefore, perhaps not surprising that participants highlighted the key danger in seeing the CTS/R as synonymous with CBT competence is that it can lead therapists to work towards achieving good CTS/R ratings (e.g. trying to "tick all of the scale's boxes" [P3]), rather than focusing on delivering quality CBT. As P7 stated, 'students sometimes are more concerned with getting a good CTS score than they are with doing a good CBT session'. Participants also identified the need for trainers and supervisors to recognize and resist the temptation to teach therapists how to pass assessments:

> One of the risks is the temptation to teach to your assessment, rather than teaching to need … students in fact will say 'teach me to pass this exam'. No I'll teach you to do therapy. If you do therapy you'll pass the test. P9

Whilst rating scales were viewed as necessary for assessing CBT competence, this theme highlights the importance of recognizing that these are merely tools that should not be the primary driving force behind CBT training or treatment delivery.

In summary, the first theme outlines the difficulties participants faced in deciding what to assess when assessing CBT competence. Challenges included the tensions between generic and CBT-specific skills, between protocol-specific interventions and across-protocol CBT skills, a lack of agreement in the field over how CBT should be performed and the danger of scales coming to define competence.

How to assess cognitive–behavioural therapy competence: Selecting from the toolbox of assessment methods

A number of different methods can be used to aid in the assessment of therapists' competence in delivering CBT. This theme provides insight into when, why and how assessment methods were selected and used by participants. The first sub-theme relates to how participants assessed therapists' knowledge and practical understanding of CBT, whilst the second sub-theme discusses how participants assessed whether therapists had the skills to apply this knowledge. The final sub-theme considers the disparity between the ideal approach to assessing CBT competence and what was considered feasible in practice.

Assessing knowledge and understanding

Four methods can be used to assess therapists' knowledge and practical understanding of CBT: multiple-choice questions (MCQs), short-answer clinical vignettes, case reports and essays (Muse & McManus, 2013). MCQs and short-answer clinical vignettes were not used by participants, whilst essays and case reports were used only within training settings or for accreditation purposes. Instead, participants primarily assessed therapists' CBT knowledge through informal verbal discussion and reviewing case formulations within supervision. This gave participants insight into whether therapists' clinical work was grounded in a sound understanding of underlying CBT theory, principles and research:

> You're thinking more often about the way which people discuss their cases, whether it evidences that they are able to formulate problems using the CBT model. That their discussion and thinking about the case … is informed by a clear CBT model and that the kinds of things they are doing fit with that. P8

Thus, although a number of standardized methods can be used to assess CBT knowledge, these were largely not used, especially outside of formal education or accreditation settings. Instead, participants relied on case discussion within supervision to assess therapists' knowledge of CBT theory and research and their ability to relate this knowledge to clinical practice.

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

Cognitive–behavioural therapy knowledge was viewed as necessary but not sufficient to infer competence. Instead, examining therapists' ability to apply this knowledge was the central focus of competence assessments:

> To me that is the essence of competence—it's not just knowing the theory, it's actually knowing how to use it. P5

A variety of methods can be used to assess CBT skills, namely, patient surveys, self-assessment, patient outcome, supervisory assessments, standardized role-plays and assessor-rated treatment sessions (Muse & McManus, 2013). Although participants occasionally administered patient satisfaction surveys, these were rarely relied upon to infer competence. Self-assessment was also not used for summative assessments as it was viewed as inaccurate and prone to self-presentation biases. This view is supported by research demonstrating that therapists' self-appraisals typically differ from assessors' appraisals (Brosan, Reynolds, & Moore, 2008; Mathieson, Barnfield, & Beaumont, 2010; McManus *et al*., 2011). Self-assessment was, however, used to enhance therapists' capacity for self-reflection, a strategy that is supported by an increasing body of literature highlighting self-reflection as an important mechanism for therapist development (Bennett-Levy, 2006; Laireiter & Willutzki, 2003). Participants were aware of the difficulties in using patient outcome data to infer therapists' competence, particularly that it does not take into account treatment-specific factors (i.e., improved outcome may not be due to receiving CBT) and is confounded by patient variables (e.g., patient difficulty and comorbidity). Despite the perceived fallibility of this method, participants felt it was still important to examine patient outcomes alongside other assessment methods, as the primary objective of CBT is to reduce the impact of patients' distressing symptoms.

Participants routinely assessed competence on the basis of observations made during supervision. Although best practice guidelines recommend viewing session material within supervision, case discussion alone was typically relied upon to infer competence. This supports Townend, Iannetta and Freeston's (2002) finding that the use of direct observation is rare in routine practice. There was, however, a sense of scepticism about whether what a therapist says they do—*talking the talk*—is an accurate representation of what they actually do in practice: *walking the walk* (P11). This scepticism was in part due to recognition that supervisees may not always recognize or choose to report important information during supervision. A study by Ladany, Hill, Corbett and Nutt (1996) found that supervisees often fail to disclose key session material, providing confirmation of this observation. However, participants also considered talking about and demonstrating skills to be distinct aspects of competence, akin to Bennett-Levy's (2006) declarative and procedural knowledge systems:

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

Thus, there was a preference for assessments based on direct observations of therapists' clinical performance, such as role-plays and assessor-rated treatment sessions. Role-plays were typically used as informal learning tools rather than as formal skills assessments. Participants commented that role-plays could overcome some of the difficulties faced when using session recordings (e.g., reducing the impact of patient variability, removing bias in the selection of recordings and overcoming practical and ethical difficulties faced by recording and storing clinical sessions). However, they also recognized that, as role-plays are constructed, artificial simulations, therapists' ability to perform within a role-play may not translate to an ability to perform within a real clinical encounter:

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

Some participants expressed a preference for assessing competence on the basis of real treatment sessions 'I think real practice is, is essential really. You don't pass your driving test without driving on the road' (P9). This typically involved rating audio or video recordings of treatment sessions using standardized rating scales. In particular, participants used the CTS and CTS-R (CTS-R: Blackburn *et al*., 2001; CTS: Dobson *et al*., 1985), although it was

perceived as 'a bit of a wobbly instrument' (P11). Criticisms included a lack of applicability beyond depression, failure to account for disorder/protocol-specific competences or stage of therapy, under emphasis of the therapeutic relationship, unclear instruction manual, ambiguity and lack of behavioural specificity in scale item descriptors, a high degree of assessor inference, poor inter-rater reliability and a lack of opportunity to provide formative feedback. In addition to session recordings, participants noted the importance of having contextual information (e.g., session number, presenting problem, goals, formulation, materials used during the session and patient outcome data) and sometimes viewed case reports alongside treatment sessions to provide contextual detail and insight into therapists' understanding of CBT.

In sum, assessing therapists' CBT skills was considered central to the assessment of CBT competence. Although participants often used case discussion within supervision to infer CBT skills, direct observations of therapists' performance were thought to be more reliable. Role-plays were identified as a potentially useful method of observing therapist performance, although participants were uncertain whether performance in role-plays would translate to performance within real clinical practice settings. Instead participants relied upon ratings, on standardized scales, of therapists' performance during recordings of treatment sessions.

Gold standard versus feasibility

A multi-method approach involving multiple samples of an individual's performance is widely advocated to ensure that reliable and generalizable inferences can be made about an individual's competence (Kaslow *et al*., 2009; Sharpless & Barber, 2009; Van der Vleuten *et al*., 2010). Consistent with this, participants considered the 'gold standard' to be a multi-method assessment involving multiple observations of therapist skill (e.g., role-plays and recorded treatment sessions) sampled across a range of patients presenting with a variety of problems, in combination with assessment of theoretical knowledge (e.g., case discussion, case reports and essays) and examination of patient outcome data. Participants commented that this approach assesses multiple aspects of CBT competence (i.e., knowledge, application of knowledge and skill) and compensates for inherent weaknesses of each method. Although participants advocated a multi-method approach, they also recognized that it was prohibitively time-consuming and expensive and thus of limited feasibility:

> The gold standard … is going to be observation by a number of expert observers of a number of samples of work with different methods of assessing them. So some observations of live practice, some observations of theoretical knowledge. And that's just not going to be practical for your local services. P1

Indeed, the way participants assessed competence in reality was often in stark contrast to this 'gold standard' approach. For example, some participants reported having access only to a limited number of 'snapshots' (P9) of therapists' performance, whilst many had no option but to rely on case discussion alone. Participants identified the need for feasible and cost-effective multi-method assessment packages to bridge the gap between what is optimal and what is feasible in practice.

Who is best placed to assess cognitive–behavioural therapy competence: Expertise and independence

Discussion about who is best placed to conduct CBT competence assessments centred on two sub-themes. The first relates to the amount of experience and understanding of CBT assessors require, whilst the second considers what level of assessor independence is most beneficial. Both sub-themes highlight the complexity of this issue and reflect the uncertainty within the field about who is best placed to assess CBT competence.

Assessor experience and expertise

Opinions about the level of CBT expertise required to assess CBT competence varied widely. Participant 13 argued that 'someone intelligent, conscientious but not necessarily a therapist' could assess CBT competence, providing the measure being used adequately operationalized the expected behaviour and that the novice received appropriate training. In line with recent research (Weck *et al*., 2011), others held the contrasting viewpoint that, whilst novices can identify implementation of techniques (i.e., rate adherence), CBT expertise is necessary to identify whether techniques were skilfully implemented and appropriate given the individual patient's presentation:

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

This view seemed to rest on the belief that assessing CBT competence is a complex skill that needs to be fostered

through experiential training and requires experience and expertise in CBT delivery. Performance-based assessments (e.g., role-plays and treatment sessions) were regarded as involving a particularly high level of assessor skill, as was providing developmental feedback (as opposed to simply scoring scale items). P6 reflects on the skills involved in assessing competence:

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

In sum, all participants highlighted the need for adequate assessor training, regardless of CBT expertise. However, opinions about how much CBT experience assessors require ranged from inexperienced novices to experts in the field, revealing uncertainty regarding the optimal level of therapeutic experience required to assess CBT competence.

The pros and cons of supervisors as assessors

Participants routinely assessed their own supervisees' competence. This approach was viewed as beneficial because it provided additional contextual information (e.g., treatment context and therapist's broader work) and facilitated delivery of appropriate feedback. However, participants acknowledged that supervisory assessments may be less objective due to the impact of social demands and relationship dynamics:

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

As highlighted by supervisors within medical assessment settings, participants raised particular concerns about the undesirable consequences of negatively evaluating supervisees (Dudek, Marks, & Regehr, 2005). Questions regarding supervisors' objectivity may be warranted, as agreement between supervisory and independent competence ratings is often low, with supervisors generally providing more favourable ratings (Borders & Fong, 1992; Dennhag *et al*., 2012; Martino *et al*., 2009). Although supervisory assessments are often a practical necessity, some viewed potential supervisor biases as insurmountable and therefore recommended that independent assessors be used. Others, however, argued that such difficulties could be circumvented by implementing reliability checks by independent assessors. Participants also noted that the decision to use supervisors or independent assessors may depend on the purpose of the assessment. For example, supervisors who have knowledge of the broader picture and a relationship with the therapist may be advantageous within formative assessments, whilst objective, independent assessors may be more appropriate for summative assessments.

Pitfalls: Identifying and overcoming assessment biases

Participants identified a number of pitfalls to be aware of in assessing CBT competence. These are presented below as those which relate to (i) patients and (ii) therapists.

Some patients will make stars of us all

Participants noted that it is more difficult for therapists to deliver effective CBT and therefore to demonstrate competence with more complex cases. Conversely, straightforward cases were viewed as being less taxing for therapists. As P11 acknowledged, 'some patients will make stars of us all'. Little research has been conducted to examine this issue; however, participants postulated a number of reasons why patient difficulty may impact on therapists' competence. For example, patients with only one 'straightforward' problem (e.g., panic disorder) would be easier to treat than patients with multiple comorbid problems or more complex difficulties (e.g., borderline personality disorder) that typically require a much higher level of general therapeutic skill as well as the skilful delivery of a broader range of complex CBT interventions. Participants also noted that patients do not always 'play the game' (P8) in that they may talk too little or too much, not be psychologically minded, have low motivation, be reluctant to engage in treatment or be experiencing severe social or environmental adversity, all of which present challenges for therapists implementing standard CBT protocols:

> If you have a really straightforward patient you try the first and obvious strategy, it may work and you move on … Whereas if you have a moderately complicated patient … strategy one and two may not work, you may have to try three or

> maybe even four, err, and do that with some evidence and to keep going, you know, rather than bottling it. And that becomes a much higher level skill. P6

Participants were aware that patient difficulty and complexity could lead to unfair comparisons, with those working with more challenging cases being at a disadvantage. Thus, participants highlighted the importance of assessors taking patient difficulty into account when making judgements about a therapist's competence. Participants also noted that this bias could be minimized by using standardized patients (e.g., role-plays) or by including multiple assessments of therapists' ability to deliver CBT across a range of different patient presentations and complexity.

Therapist self-presentation biases

A further theme that emerged was that therapists may present themselves in a biased manner within competence assessments. Three possible self-presentation biases were discussed. First, participants noted that therapists often put a positive spin on their performance in an effort to hide their mistakes:

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

This bias was thought to be driven by a fear of exposing oneself as incompetent and presented itself within self-assessments and verbal and written case discussion and resulted in therapists' selectively submitting recordings of their best treatment sessions for evaluation. Second, participants commented that therapists' self-confidence colours their own perception of their performance and thus any reports of their performance (e.g., within case discussion and self-assessment). Therapists were seen as largely falling into one of two groups: (i) those who are overconfident and unaware of their flaws and (ii) the self-critics:

> Some of it just reflects confidence. There are some people who would score themselves really quite highly and I think 'what are you on?' and other people will score them- selves very low … they are the self-critics. P6

Such a self-confidence bias could explain why therapists have been found to both underestimate and overestimate their own competence (Brosan *et al*., 2008; Mathieson *et al*., 2010; McManus *et al*., 2011). Third, participants observed that the anxiety therapists' experience when their performance is being scrutinized (e.g., treatment sessions and role-plays) can inhibit their performance. Although other participants commented that performance anxiety does not mask competence but demonstrates a limit to it, a competent therapist should be mindful of and able to manage their own affect.

Participants highlighted several strategies for minimizing the impact of therapist self-presentation biases. These included providing a supportive assessment environment and viewing therapists' performance (e.g., session recordings and role-plays) rather than relying solely on self-assessment, case reports or case discussion. Participants also suggested routinely recording treatment sessions so this becomes the norm rather than the exception, thus making the process less anxiety provoking and enabling random selection of recordings for assessment.

Discussion

Summary of results

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

Refining the toolbox of assessment methods

A broad range of methods can be used to assess CBT competence (Muse & McManus, 2013). However, the current

study suggests that only a narrow selection is implemented in practice, thus highlighting the importance of focusing future efforts on the development of methods that are perceived as practical, credible and valid. When evaluating therapists' knowledge and practical understanding of CBT, assessors only employed structured measures within formal educational settings, with those in routine practice settings relying instead on case discussion within supervision. Standardized supervisor measures could, therefore, provide a useful means of assessing therapist knowledge in routine practice. Broader assessments that assess both CBT knowledge and skills could also be useful. For example, case summaries could provide a useful adjunct to in-session performance ratings as they offer insight into whether therapists' clinical work is grounded in a sound understanding of underlying CBT theory, principles and research as well as providing therapeutic context. Clinical skill was viewed as being at the heart of delivering competent CBT and was typically assessed using recordings of therapists' in-session performance. Given the complexities involved in specifying the skills involved in the competent delivery of CBT, rating scales that behaviourally operationalize these skills were seen as providing a helpful framework for rating therapists' performance. However, results of the current study are consistent with previous suggestions that there is a need for more comprehensive rating scales with improved validity, reliability and usability (Fairburn & Cooper, 2011; Muse & McManus, 2013). Observations during supervision were also frequently used as a basis for assessment of CBT skills, despite typically being based on second-hand reports, rather than live performance. Given participants' strong preference for assessing clinical skill on the basis of direct performance-based observations, discussion-based supervisory assessments may be limited in their ability to serve as summative assessments. However, supervisory assessments still provide useful formative feedback and thus play an important role in training and continuing professional development (Dochy & Moerkerke, 1997; Milne, 2007; Van der Vleuten *et al*., 2010). Although not widely used at present, structured role-play assessments may warrant further exploration. Medical training settings routinely assess clinical skill using objective structured clinical examinations comprising a series of short encounters with standardized patients (Epstein, 2007; McNaughton, Ravitz, Wadell, & Hodges, 2008), and a similar approach has been advocated for assessing CBT skills (Fairburn & Cooper, 2011). However, research in this domain would need to address participants' concerns that therapists' ability to perform within artificially constructed role-plays may not directly translate to their ability to perform within real clinical encounters.

Choosing the right tool for the job

Selecting an assessment strategy is a complex process that will inevitably entail some compromises. The impact and relative importance of these compromises, and thus the degree to which any given method is appropriate, will, however, depend on the specific circumstances of the assessment. Results from this study highlight a number of factors that need to be taken into account when selecting appropriate strategies for assessing CBT competence. It is important to weigh up the inherent strengths and weaknesses of each assessment method, as well as the validity, reliability, usability and feasibility of the specific tools available. The assessment context will impact on the utility of any given assessment method and will, therefore, also need to be considered. For example, protocol-specific measures may be required within the context of research trials, whereas transdiagnostic measures may be more appropriate in some routine practice and training settings. The purpose of assessment should also guide the choice of assessment strategy. For example, the inclusion of formative feedback is not necessary for accreditation purposes but is essential for skill development within training settings. It will also be necessary to carefully consider the experience, expertise and independence of those who conduct the assessments.

Multi-method competency assessment programmes

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

Combining formative and summative assessments

Methods with high validity and reliability are the primary concern for 'high stakes' summative assessments that provide an overall judgement of competence for qualification or accreditation purposes. In contrast, formative assessments provide corrective feedback in order to promote self-reflection and guide future learning and so require methods that

provide in-depth feedback regarding specific aspects of competence. It has, therefore, been suggested that different assessment methods may be required for formative and summative purposes (Epstein, 2007), and results from this study concur. For example, self-assessment was not viewed as robust or accurate enough for summative purposes. Nevertheless, comparing therapists' own ratings of their in-session performance with an assessor's rating provided a useful tool for enhancing self-reflection (Bennett-Levy, 2006; Laireiter & Willutzki, 2003). That being said, participants did not typically employ independent formative and summative assessments, suggesting the need for assessment tools that can serve both functions. The inclusion of formative feedback may be particularly helpful in promoting the development of competent, reflective practitioners, within the context of both training and ongoing continuing professional development (Dochy & Moerkerke, 1997; Milne, 2007; Van der Vleuten *et al*., 2010), and is typically well received by those being assessed (Govaerts, van der Vleuten, Schuwirth, & Muijtjens, 2007; Shute, 2008; Van der Vleuten *et al*., 2010).

Recognizing assessment consequences

Examining both the intended and unintended consequences of assessing competence is a necessary and important aspect of test evaluation (Linn, Baker, & Dunbar, 1991; Messick, 1994). Results from this study draw attention to a number of unintended negative effects that can occur as a result of assessing CBT competence. For example, the process of being assessed was recognized as anxiety provoking and thus requires a supportive assessment environment. Teaching therapists how to pass assessments, rather than how to deliver effective CBT, was also identified as a pitfall to be avoided. As competence frameworks (e.g., Roth & Pilling, 2007) are much broader than the tools used to assess CBT competence, focusing on the content of assessment tools could have a drastic impact on the breadth and quality of training. Of particular concern were participants' reports that striving to fulfil assessment criteria can sometimes lead to poorer treatment delivery. These issues do not appear to be restricted to the context of assessing CBT competence, as research within wider educational settings shows that assessments have a strong influence on learning and the provision of training, causing both trainers and trainees to focus on assessment requirements (Alderson & Wall, 1993; Prodromou, 1995). It is therefore imperative that those involved in assessing CBT competence are alert to such unintended consequences.

Limitations

Conclusions from this study must be interpreted in the light of limitations inherent in the qualitative method used. The retrospective nature of the study means participants' accounts may have been influenced by factors such as memory biases. In line with the principles of IPA, this study offers in-depth insight into the experiences of a limited group of individuals, rather than seeking to provide a representative sample (Brocki & Wearden, 2006). However, it is important to recognize that, although participants were from a range of academic, training and routine practice settings and included both men and women, the current sample is somewhat homogenous in that all were from the UK, most were clinical psychologists and many had undertaken CBT training around the 1980s and 1990s. It is also important to recognize that, in accordance with IPA principles, the researchers played a dynamic role in generating and analysing the data. Hence, it is possible that another researcher with different personal characteristics and theoretical beliefs would have facilitated a different interview and offered a different interpretation of the data.

Future directions

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

Acknowledgements

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

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

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*(2), 115–129. DOI:10.1093/applin/14.2.115.

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

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

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

Blackburn, I. M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy*, *29*(4), 431–446. DOI:10.1017/S1352465801004040.

Borders, L. D., & Fong, M. L. (1992). Evaluation of supervisees. *The Clinical Supervisor*, *9*(2), 43–51. DOI:10.1300/J001 v09n02_05.

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

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

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

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

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

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

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

Epstein, R. M. (2007). Assessment in medical education. *New England Journal of Medicine*, *356*(4), 387–396. DOI:10.1056/NEJ Mra054784.

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

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

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

Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency assessment toolkit for professional psychology. *Training and Education in Professional Psychology*, *3*(4), 27–45. DOI:101037/a0015833.

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

Ladany, N., Hill, C. E., Corbett, M. M., & Nutt, E. A. (1996). Nature, extent, and importance of what psychotherapy trainees do not disclose to their supervisors. *Journal of Counseling Psychology*, *43*(1), 10–24. DOI:101037/0022-016743110.

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

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

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

Mathieson, F. M., Barnfield, T., & Beaumont, G. (2010). Are we as good as we think we are? Self-assessment versus other forms of assessment of competence in psychotherapy. *The Cognitive Behaviour Therapist*, *2*(01), 43–50. DOI:10.1017/S1754470

X08000081.
Milne, D. (2007). Evaluation of staff development: The essential 'SCOPPE'. *Journal of Mental Health*, *16*, 389–400. DOI:10.1080/ 09638230701382818.
McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*, *65*(2), 73–84. DOI:10.1037/a0018121.
McManus, F., Rakovshik, S., Kennerley, H., Fennell, M., & Westbrook, D. (2011). An investigation of the accuracy of therapists' self-assessment of cognitive–behaviour therapy skills. *British Journal of Clinical Psychology*. DOI:10.1111/j.2044-8260.2011.02028.x.
Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13–23. DOI:10.3102/0013189x023002013.
Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive–behavioural therapy. *Clinical Psychology Review*, *33*(3), 484–499. DOI:10.1016/j.cpr.2013.01.010.
Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
Prodromou, L. (1995). The backwash effect: From testing to teaching. *ELT Journal*, *49*(1), 13–25. DOI:10.1093/elt/49.1.13.
Reid, K., Flowers, P., & Larkin, M. (2005). Exploring lived experience. *Psychologist*, *18*(1), 20–23.
Rodolfa, E., Bent, R., Eisman, E., Nelson, P., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice*, *36*(4), 347–354. DOI:10.1037/ 0735-7028.36.4.347.
Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.
Sharpless, B. A., & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review*, *29*(1), 47–56. DOI:101016/j. cpr.200809008.
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. DOI:10.3102/ 0034654307313795.
Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. London: Sage.
Sofaer, S. (1999). Qualitative methods: What are they and why use them? *Health Services Research*, *34*(5), 1101–1118. DOI:10.1093/intqhc/14.4.329.
Townend, M., Iannetta, L., & Freeston, M. H. (2002). Clinical supervision in practice: A survey of UK cognitive behavioural psychotherapists accredited by the BABCP. *Behavioural and Cognitive Psychotherapy*, *30*(04), 485–500. DOI:10.1017/ S1352465802004095.
Van Der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, *39*(3), 309–317. DOI:101111/ j.1365-2929200502094x.
Van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: Building blocks for theory development. *Best Practice & Research. Clinical Obstetrics & Gynaecology*, *24* (6), 703–719. DOI:10.1016/j.bpobgyn.2010.04.001.
McNaughton, N., Ravitz, P., Wadell, A., & Hodges, B. D. (2008). *Psychiatric education and simulation: A review of the literature* (Vol. *53*). Ottawa, ON, CANADA: Canadian Psychiatric Association.
Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*(4), 620–630. DOI:10.1037/0022-006X.61.4.620.
Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *Journal of Nervous and Mental Disease*, *199*(4), 276–279. DOI:10.1097/NMD.0b013e3182124617.