Volume 48 Issue 11 November 2010 ISSN 0005-7967

ELSEVIER

BEHAVIOUR RESEARCH AND THERAPY

An International Multi-Disciplinary Journal

Contents

Regular Papers

Continued on outside back cover

Abstracted/indexed in: Biosis Data, Elsevier BIOBASE/Current Awareness in Biological Sciences, Curr. Cont./Soc. & Beh. Sci., EMBASE/Excerpta Medica, Ind. Med., LLBA, MEDLINE, PASCAL-CNRS Data., Psychol. Abstr., PsycSCAN, PsycINFO, PsycLIT, Res. Alert, Soc. Sci. Cit. Ind., Social Work Res. & Abstr. Also covered in the abstract and citation database SCOPUS®. Full text available on ScienceDirect®.

ISSN 0005-7967
48(11) 1067–1160 (2010)

Behaviour Research and Therapy 48 (2010) 1123–1132

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

An evaluation of the effectiveness of Diploma-level training in cognitive behaviour therapy

Freda McManus [a,b,2,*], David Westbrook [a,1], Maria Vazquez-Montes [c,3], Melanie Fennell [b,2], Helen Kennerley [a,1]

[a] Oxford Cognitive Therapy Centre, Warneford Hospital, Oxford OX3 7JX, UK
[b] University of Oxford, Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK
[c] University of Oxford, Department of Public Health, Old Road Campus, Oxford OX3 7LF, UK

ARTICLE INFO

Article history:
Received 31 January 2010
Received in revised form
4 August 2010
Accepted 4 August 2010

Keywords:
Cognitive behaviour therapy training
Evaluation
Cognitive therapy rating scale
Dissemination

ABSTRACT

Background: As part of the UK government's initiative to Increase Access to Psychological Therapies (see http://www.iapt.nhs.uk/for full details of the IAPT programme) there has been an expansion in the provision of post-graduate Diploma training in cognitive behaviour therapy (CBT). Previous evaluations of such training programmes have yielded mixed results but have been limited by small sample sizes and/or limited assessment measures.

Aims: To evaluate the impact of a long-standing Diploma in CBT training programme on a variety of measures of CBT competence.

Method: Trainees' levels of CBT skill are compared at the beginning and end of CBT training. The effect of therapist factors such as age, professional background and gender on the development of CBT competence is also examined.

Results: Results show that trainees demonstrate higher levels of CBT skills after completing the training than they did before, with the majority achieving pre-determined criteria for competence. Trainees' gender was not related to their performance but trainees' age showed a negative association with CBT skill (older trainees performed worse). Trainees' professional background also had an impact on their level of CBT competence, with trainees who were clinical psychologists demonstrating the highest levels of competence across a range of measures.

Conclusions: CBT Diploma training leads to increases in the level of trainees' CBT competence, with the majority achieving the levels demonstrated in research trials by the end of training. Thus, this training is likely to lead to improved outcomes for patients. Further research is needed to determine the most efficient ways of enhancing CBT skills.

Introduction

Cognitive behaviour therapy (CBT) and CBT training are in a phase of unprecedented expansion. This is partly due to the increasing evidence for CBT's effectiveness with a variety of disorders (Roth & Fonagy, 1996; NICE, 2004a, 2004b). In addition, recent economic analyses have confirmed the considerable cost to the state of untreated mental illness (e.g., Layard, 2006) and have prompted initiatives to increase access to evidence-based psychological therapies. In October 2007 the UK government announced "an unprecedented, large-scale initiative for Improving Access to Psychological Therapies (IAPT) … within the English National Health Service. Between 2008 and 2011 investment in psychological therapies for these conditions will steadily rise to £173 million [approximately 260 million US dollars] per annum above existing expenditure. The extra investment is being used to train and employ at least 3600 new psychological therapists" (Clark et al., 2009, p. 910). Because of the evidence for CBT's efficacy, most of this funding will go towards training an additional 3600 CBT therapists (see http://www.iapt.nhs.uk/for full details of the IAPT programme). Training in 'high intensity' CBT (standard CBT, as opposed to 'low intensity' CBT/self-help) is being done by dramatically increasing the number of courses providing post-graduate Diploma (Masters) level training in CBT. This increase in funding represents a huge investment and hence it seems timely to consider the effectiveness of one such course that has been in existence for 18 years.

* Corresponding author at: University of Oxford, Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK. Tel.: +44 1865 738816; fax: +44 1865 738817.
E-mail address: freda.mcmanus@psych.ox.ac.uk (F. McManus).
[1] Tel.: +44 1865 738816; fax: +44 1865 738817.
[2] Tel.: +44 1865 223635; fax: +44 1865 793101.
[3] Tel.: +44 1865 289432; fax: +44 1865 289436.

0005-7967/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.brat.2010.08.002

The aim of CBT training is to increase therapists' CBT competence and the increased investment in CBT training is based on two premises: first, that increased training will lead to increased competence; and second, that increased competence will lead to improved outcomes for patients. Whilst some evidence links increased CBT competence with improved patient outcome (Kingdon, Tyrer, Seivewright, Ferguson, & Murphy, 1996; Kuyken & Tsivrikos 2009; Shaw et al., 1999; Strunk, Brotman, DeRubeis, & Hollon, 2010; Trepka, Rees, Shapiro, Hardy, & Barkham, 2004), there is much less consensus about the link between training and improved CBT competence. Previous research into the effectiveness of CBT training has yielded mixed results, with some studies showing no enduring effects of training (King et al., 2002; Walters, Matson, Baer, & Ziedonis, 2005); some showing limited effects of training (Mannix et al., 2006; Sholomskas et al., 2005); and others demonstrating significant effects on therapists' CBT skills and patient outcomes (Grey, Salkovskis, Quigley, Clark, & Ehlers, 2008; Westbrook, Sedgwick-Taylor, Bennett-Levy, Butler, & McManus, 2008).

The few previous evaluations of CBT Diploma training have yielded similarly mixed results. The smallest study reported no significant effect of training (Williams, Moorey, & Cobb, 1991), the most comprehensive reported a small effect size on a standard measure of CBT skills (Keen & Freeston, 2008), and a further small study reported a large effect size as well as an impact on patient outcomes (Milne, Baker, Blackburn, James, & Reichelt, 1999). This research suggests that Diploma training may have some positive impact on trainees' CBT skills, but the conclusions that can be drawn are limited by small sample sizes (n's range from 11 to 52), reliance on single measures and/or one-off measurements, and variable standards for defining competence. The small samples also limit the generalisability of findings and prohibit any analysis of the effects of therapist characteristics such as age, gender or professional background. The question of whether such therapists' characteristics are related to CBT competence, and thus to patient outcomes, has received little previous investigation. James, Blackburn, Milne, and Reichelt (2001) reported no relationship between trainees' age and the acquisition of competence in CBT, but found that therapists with greater previous experience of CBT showed higher levels of competence. In contrast, Sholomskas et al. (2005) found no impact of therapists' amount of previous experience on outcomes from different forms of CBT training. With regard to gender, James et al. (2001) reported that their finding that male therapists ($n = 9$), who started from a lower baseline, showed more improvement in competence than female therapists ($n = 11$) (who showed little improvement), was unclear and required further investigation in larger samples. With regard to professional background, Brosan, Reynolds, and Moore (2006) hypothesized that because of greater exposure to CBT methods in their professional training, clinical psychologists would demonstrate greater CBT competence than those from other professional backgrounds, but they found only limited support for this hypothesis.

A further motivation for evaluating CBT training programmes comes from the shift in focus of CBT research from efficacy research to effectiveness research (Hunsley & Lee, 2007; McManus, Grey, & Shafran, 2008; Stewart & Chambless, 2009). The value of CBT interventions shown to be efficacious in randomised controlled trials (RCTs) is limited if the results cannot be replicated in routine clinical practice. The effect sizes seen in RCTs have not always been replicated in routine clinical practice studies, and a lower level of therapist competence has been suggested as an explanation for the reduced effect sizes (Davidson et al., 2004; Kingdon et al., 1996; Shaw et al., 1999). Whilst previous evaluations of Diploma training have considered whether CBT skills are enhanced, none has examined whether trainees achieve pre-determined standards of competence such as those demonstrated by therapists in efficacy trials. Hence, it is important in assessing CBT effectiveness to know what proportion of the therapists being trained in routine clinical practice are attaining the standard of competence demonstrated by therapists in RCTs.

Given the current increased focus on the dissemination of CBT and the government's investment in CBT training, it is timely to conduct a more comprehensive evaluation of a CBT training programme. The current study aims to address some of the limitations of prior research by using a larger sample and a more comprehensive range of measures to assess the impact of training on therapists' CBT competence. Secondary aims are to compare the effects of training across different ages, professional groups and genders, and to examine the proportion of trainees achieving predetermined standards of competence.

Method

Training course

The Oxford Diploma course reflects a collaboration between Oxford University and the Oxford Cognitive Therapy Centre (part of Oxfordshire and Buckinghamshire NHS Foundation Trust). The course prospectus and details of the selection criteria and process can be found on the Oxford Cognitive Therapy Centre website (http://www.octc.co.uk/content.asp?PageID=53). The course leads to the award of a Master's level post-graduate Diploma in CBT from the University of Oxford. The course is accredited by the British Association of Behavioural and Cognitive Psychotherapists (BABCP) and has received favourable reports from the University's periodic review process. All course staff are BABCP-accredited CBT therapists and many are accredited as CBT supervisors and trainers. The course was established in 1992 and has been one of the most popular, longest running and largest courses of its type in the UK. The course is aimed at qualified mental health professionals with at least two years post-qualification experience and some prior experience of CBT. It is heavily over-subscribed, with two to three applicants for each place. Most successful applicants already have post-graduate qualifications and come from the core mental health professions of clinical psychology, psychiatric nursing, psychiatry and counselling.

The course consists of 36 days of training over three 12-week terms, with 4-week breaks between the first and second, and second and third terms. Trainees are usually employed in clinical settings and are released to attend the course one day a week during each of the three terms. Each training day comprises a 90 min group supervision session (one supervisor with three trainees) followed by a 5-h workshop. In addition to attending the 36 training days trainees are expected to set aside a further six to 8 h per week, throughout the duration of the course, for the private study and preparation that is required for successful completion (preparing for supervision, listening to and reflecting on therapy recordings, completing academic assignments).

In order to ensure exposure to a variety of experiences, trainees change supervisors and supervision groups each term. During the course, they are expected to treat at least six patients with CBT. In the first two terms they treat relatively straightforward cases (e.g. anxiety or depression); in the third, they are encouraged to practise their developing skills with more complex cases.

The 5-h workshops aim to train clinical skills within the framework of relevant theory and research. The focus is primarily on the application of CBT in clinical practice; demonstrations and experiential role-plays are used to help students develop their skills. The first term concentrates on fundamental CBT skills: assessment, goal setting and socialisation to the CBT model; structuring sessions; formulation; Socratic methods; identifying

and challenging automatic thoughts; identifying and re-evaluating dysfunctional assumptions; behavioural experiments; imagery techniques; techniques for changing core beliefs; and process issues and ending treatment. The second term concentrates on evidence-based disorder-specific protocols. Workshops cover: an overview of anxiety disorders; panic and agoraphobia; social phobia; obsessive-compulsive disorder, health anxiety; post-traumatic stress disorder; eating disorders; depression; hopelessness and suicidality; and psychosis. The third term addresses more complex transdiagnostic processes in CBT and covers the following topics: conceptualisation – applying theory to complex cases; schema theory; schema change methods; interpersonal issues in CBT; low self-esteem; avoidance in complex cases; advanced imagery techniques; childhood trauma and abuse; compassion-based methods in CBT; and an introduction to an evidence-based 'third wave' approaches.

Participants

Participants were the 278 trainees (72 male, 26%) who completed the course between 1998 and 2009. All trainees who undertook the training during this time are included, using data collected and stored as part of the routine running of the course. This means that the data is incomplete; not all records and recordings were retained beyond the 5-year period required by University regulations. For this reason, *n*'s are reported separately for each analysis.

Assessments and measures

It is recognised that there are difficulties in defining CBT competence and what constitutes a valid measure of it (Sharpless & Barber, 2009). For the present study, trainees' CBT knowledge and skill was assessed via supervisors' ratings of their CBT skills, based on recordings of therapy sessions (at six time points during the course), and via marks awarded for written assignments (three essays and two case reports). The primary measure of CBT competence was considered to be the supervisor ratings of trainees' actual clinical performance, as previous research suggests (a) that therapists' "self-reported perceptions of change are frequently not matched by actual behaviour change" (Beidas & Kendall, 2010, p. 15) and (b) that supervisors' ratings of clinical performance have the greatest validity because they bear the closest relationship with patient outcomes (Chevron & Rounsaville, 1983; Kuyken & Tsivrikos, 2009).

Ratings of clinical performance: Cognitive Therapy Scale (CTS: Young & Beck, 1980, 1988)

Trainees selected one of their recordings (audio or video) of a mid-treatment CBT session for submission for evaluation at the beginning and end of each term (six recordings in total). Each recording was evaluated on the Cognitive Therapy Scale (CTS) by their supervisor that term. The CTS was developed by Young and Beck (1980, 1988) to assess levels of competence in CBT. In completing the scale, an experienced cognitive therapist rates a recording of a therapy session on a number of components of cognitive therapy. There are several adaptations of the CTS in existence and the version used in the present study is the 13-item version that has the most thorough available manual (Young & Beck, 1988), providing detailed descriptions for each item, desirable therapist strategies, examples, and notes about special factors in rating. The first 11 items cover General Interview Procedures (Agenda, Feedback, Collaboration, Pacing), Interpersonal Effectiveness (Understanding, Interpersonal effectiveness) and specific Cognitive-Behavioural Techniques (Guided discovery, Strategy for change, Focus on key cognitions/behaviours, Application of cognitive and behavioural techniques, Homework). The twelfth item rates the therapist's skill in dealing with any special circumstances that arose (if applicable), and the thirteenth is an overall rating of skill. All items are rated on a seven point (0–6) Likert scale yielding total scores that range from 0 to 78, with higher scores representing a higher level of skill.

The CTS has been criticised for overlap between the items and multiple concepts being addressed by one item, different levels of inference being required to make ratings, neglecting certain areas of CBT competence (e.g., the role of emotion), and being overly focussed on the narrow protocol of CBT for depression (Barber, Liese, & Abrams, 2003; Blackburn et al., 2001, Jacobson & Gortner, 2000; Whisman, 1993). In addition, while some studies have reported adequate internal and inter-rater reliability and discriminant validity for the CTS (e.g., Dobson, Shaw, & Vallis, 1985; Trepka et al., 2004; Williams et al., 1991) others have questioned its reliability and validity (e.g., Jacobson & Gortner, 2000). Alternative measures for assessing CBT competence have been proposed (e.g., Barber et al., 2003; Blackburn et al., 2001), but the CTS remains the most widely used and validated instrument for assessing CBT competence.

Previous studies have used different cut-off scores on the CTS to represent a 'red line' value for defining competence. Originally a 'red line value' of a total score of 44 on the 11-item version of the CTS (corresponding to a mean item score of 4.0 or above), was suggested for monitoring the competence of therapists in treatment trials of CBT (Shaw, 1984). Subsequently, Dobson and Shaw (1988) analysed mean ratings in Hollon et al.'s (1992) treatment of depression trial to suggest the red line should be a total score of 40 or above on the 11-item version of the CTS (corresponding to a mean item score above 3.5), and this standard has been widely adopted in the field (e.g., Dimidjian et al., 2006; Shaw et al., 1999). However, the same total score of 40 or more has been used in studies using a 13-item version of the CTS, which then corresponds to a slightly more lenient standard for competence of a mean item ratings of 3.0 or above (e.g., Brosan et al., 2006; Trepka et al., 2004). For the purpose of the present study, only the first 11 items of the CTS were used to calculate mean scores (because the twelfth item is discretionary and not applicable in the majority of cases, and the thirteenth is a global rating which will reflect the previous items). Mean item scores are reported rather than total scores to facilitate accurate comparisons with studies using scales with varying numbers of items. All three previously suggested red line cut-offs for competence are examined.

It is worth noting that the calculation of cut-off scores on the CTS is somewhat arbitrary. For example, the cut-off 40 or above (on the original 11-item version of the CTS) was derived from the value that was one standard deviation below the mean score of a group of certified cognitive therapists (Shaw, 1984; Shaw et al., 1999). Thus, the 'cut-off' does not reflect an empirically-proven 'tipping-point' for therapist competence. A further limitation of the CTS is its failure to take any quantitative account of patient complexity in assessing therapist competence, and it has been suggested that it gives a more valid estimate of competence when multiple sessions are sampled across cases (Keen & Freeston, 2008).

Written assessments (essays and case reports)

Trainees are required to submit five written assignments during the Diploma: two case reports (maximum 4000 words) and three essays (two of a maximum of 3000 words and one extended essay of 10,000 words). Written assignments (essays and case reports) are double marked, blind as to author, on a 0–100 scale, with a mark of 50% or above constituting a pass. The first two essays are written on topics that aim to enhance and consolidate trainees' learning about basic CBT techniques (e.g., conceptualisation, the role of specific techniques or CBT components such as though records, behavioural experiments, Socratic questioning or homework). The final 10,000

word essay serves the purpose of a dissertation and is written on a CBT topic of the trainee's choice – this is often a literature review of CBT in a specific area (e.g., for a particular disorder or client group), or in a particular setting (e.g., with inpatients); or an investigation of a key debate in CBT or the utility of a new CBT technique (e.g. compassionate mind training). Marking criteria require the trainee to: demonstrate knowledge of CBT theory and literature; show understanding of the implications of theory and research for clinical practice; demonstrate critical judgement; follow a coherent line of argument and, ideally show some originality of thought rather than simply parroting the received wisdom. Each case report describes the process of assessment, formulation and treatment with an individual CBT patient. Case reports are assessed on the ability to: demonstrate knowledge of CBT theory, research and clinical literature; describe clinical problems, identify relevant goals and conceptualise problem development and maintenance in accordance with CBT models; design and carry out a treatment programme which follows logically from the conceptualisation and is clearly cognitive-behavioural in nature, and assess outcome on relevant measures; take account of factors contributing to or blocking change; reflect on therapy and learn from experience; and present material clearly, coherently and concisely.

Examiners

CTS ratings and the marking of written assignments were carried out by a pool of local CBT therapists over the 11 years during which the data were collected. In general, examiners were CBT therapists with many years of experience of treating patients in a variety of settings, including both randomised controlled trials and routine clinical practice in NHS services.

Procedure

Data were collected routinely as part of delivering the training program described above. Trainees completed academic assignments according to the following schedule:

a) Recordings 1–6 were submitted at the start and end of each of the three terms, and rated on the CTS by supervisors.
b) Case reports were submitted at the beginning and end of the third term.
c) The first essay was submitted at the beginning of the second term, the second essay at the beginning of the third term, and the final essay at the end of the academic year.

For written assignments, trainees received written formative feedback as well as a numerical mark; for CTS ratings, trainees received formative feedback both verbally and in writing, as well as numerical scores on the CTS.

Analyses

Consistency, reliability and validity

Consistency tests of the 11 items of the CTS were calculated at each of the six time points. As CTS ratings were made by the trainee's current supervisor, who therefore knew both the trainee's identity and the stage of training, a random selection were re-rated 'blind'. 40 recordings from the beginning and end of training (recordings 1 and 6) were re-rated by different supervisors who were 'blind' to the stage of training as well as the trainee's identity. The inter-rater reproducibility for 'blind' and supervisors' CTS ratings at times 1 and 6 was evaluated by means of the intraclass correlation coefficient (ICC) in SPSS 15.0, using a two-way random-effects ICC and absolute agreement on single measures. Shrout (1998) suggests that the reliability of the ICC can be classified as follows: ICC = 0.00–0.10 – virtually none; 0.11–0.40 – slight; 0.41–0.60 – fair; 0.61–0.80 – moderate; 0.81–1.0 – substantial. A similar analysis was used to evaluate the reliability of the two examiners' marks for the three essays and two case reports. The validity of the 11-item CTS was also evaluated. In this case linear regression models were fit at time 1 and 6 respectively, with supervisors' CTS ratings as the dependent variable and 'blind' CTS ratings as the independent variable.

Random-effect models

Univariate and multivariate linear random-effect models were fitted to investigate the association of baseline variables (gender, profession and age) with primary outcomes (CTS ratings), and secondary outcomes (essay and case report scores). The analysis was run in Stata 10.0, using the xtmixed command. Models with random intercept and fixed slope as well as models with both random intercept and random slope were fitted. The final model was selected as the model with minimum AIC value (Akaike Information Criterion).

Normality of the dependent variables is a requirement of random-effect models. Shapiro–Wilk normality tests for CTS ratings, and essay and case report scores, were calculated at each time point. If $p < 0.01$ (allowing for multiple testing) the null hypothesis of having a normal distribution was rejected.

Results

Participant characteristics

Participants were the 278 trainees (72 male, 26%) who completed the course between 1998 and 2009. Participants' mean age was 36.7 (SD = 6.3) years, median 35 years, and the range was 28–60 years. Of the 278 participants, 145 (52.2%) were clinical psychologists, 56 (20.1%) were nurses (mostly psychiatric nurses), 33 (11.9%) were psychiatrists, 14 (5.0%) were counsellors or counselling psychologists, 13 (4.7%) were occupational therapists and 9 (3.2%) were social workers. The remaining 8 were: 2 speech therapists, 2 general practitioners, 1 charity worker, 1 substance misuse therapist, 1 health psychologist and 1 forensic psychologist.

Data

The primary outcome measure was CTS ratings (measured at six time points for each individual). Essay and case report scores were also examined (measured at three and two time points respectively).

Consistency, reliability and validity

The CTS showed high internal consistency: Cronbach's alphas ranged from 0.84 to 0.89 over the six time points. The reliability of the CTS was higher at time 1 than at time 6 (ICC = 0.71 $p < 0.001$ and ICC = 0.47 $p < 0.05$ respectively). A more stable reliability was observed for essay scores 1–3 (ICCs = 0.69 $p < 0.001$, 0.78 $p < 0.001$, and 0.74 $p < 0.001$ respectively), and for case report scores 1–2 (ICCs = 0.71 $p < 0.001$, 0.73 $p < 0.001$ respectively). These ICCs fall in the moderate range (0.61–0.8 (Shrout, 1998)) and are comparable to those reported in other studies using the CTS or the revised version of the CTS (e.g. 0.59 reported by Vallis, Shaw, & Dobson, 1986, and 0.57 reported by James et al., 2001). The reliability of the CTS subscales at time 1 and time 6 was as follows. CBT subscale ICC = 0.67 $p < 0.001$ and 0.47 $p < 0.05$ respectively; Interpersonal Effectiveness subscale ICC = 0.67 $p < 0.001$ and ICC = 0.41 $p = 0.02$, respectively; General Therapeutic Skills subscale ICC = 0.66 $p < 0.001$ and ICC = 0.55 $p = 0.003$ respectively. The validity

analysis of CTS 1 and 6 showed a strong association of 'blind' and supervisors' CTS ratings at time 1 ($\beta = 0.7$, $p < 0.001$) and significant association of 'blind' and supervisors' ratings at time 6 ($\beta = 0.4$, $p = 0.031$). In addition, 'blind' CTS ratings were not significantly different to supervisors' ratings for either recording 1: mean 3.07 (0.88) vs mean 3.17 (0.84), $t(19) = 0.71$, $p = 0.49$ or recording 6: mean 3.82 (0.74) vs mean 4.06 (0.68) $t(19) = 1.50$, $p = 0.15$, confirming the validity of the supervisors' ratings (see Fig. 1).

Missing data

There was only a small proportion of missing data for essay and case report scores: three students (1.1%) were missing any essay score, and six students (2.1%) were missing any case report score. Therefore no further analysis of missing data was carried out for essays and case reports. For CTS ratings, data were missing for 4 students (1.4%) for the first CTS rating, 17 students (6.1%) for the second CTS rating, 33 students (11.9%) for the third CTS rating, 29 students (10.4%) for the fourth CTS rating, 48 students (17.3%) for the fifth CTS rating, and 43 students (15.5%) for the final CTS rating. The most significant 'chunk' of missing data was for the 24 students who undertook the course in 2002 – for this cohort all CTS data for recording six, and most for recording five, was missing. There were no significant differences between those with and without complete CTS data (i.e. all six CTS ratings) in age ($t(226) = 0.62$, $p = 0.62$), gender ($\chi^2(1) = 0.53$, $p = 0.48$), professional group ($\chi^2(3) = 2.88$, $p = 0.41$) or overall mean CTS rating ($t(276) = 1.46$, $p = 0.15$).

Random-effect models

CTS ratings

Normality was not rejected for CTS ratings at all time points except time 5 ($p = 0.009$). Removing one outlier (the lowest mean CTS rating at time 5) resulted in $p = 0.085$. A sensitivity analysis showed that excluding this value gives similar results as when this value is included. Models with both random intercept and slope presented lower AIC values than their counterpart models with only random intercept. This implies that there is not just variability in the starting CTS ratings but also in the individuals' trend over time.

Time was included in all the fitted models and was found to be strongly associated with CTS ratings (which increased with time). The coefficient of time was $\beta = 0.16$, $p < 0.001$, when included on its own in a univariate random-effect model (see Table 1). The significant increase in CTS ratings over time is confirmed by analysis of the 'blind' ratings of recordings one and six, which also showed a significant increase from recording one to recording six (mean 3.07 (0.88) vs 3.82 (0.74) $t(20) = 3.17$ $p = 0.005$).

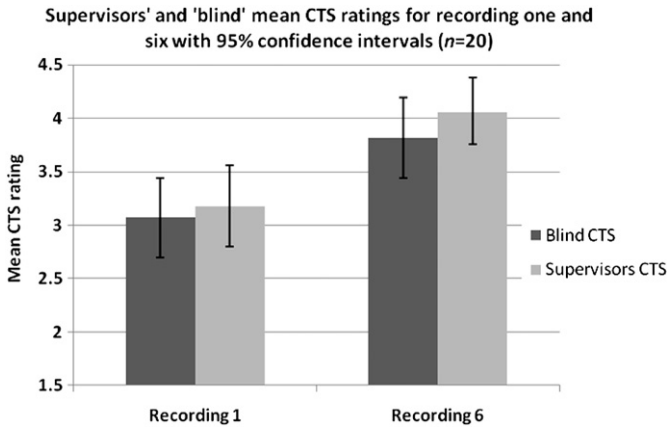

Fig. 1. Supervisors' and 'blind' CTS ratings with 95% confidence intervals.

Fitting models that included time and each of the other covariates returned the following. Gender was not significantly associated with CTS ratings. Compared to clinical psychology trainees, psychiatry and nursing trainees had a lower CTS ratings over time ($\beta = -0.19$, $p < 0.05$ & $\beta = -0.21$, $p < 0.01$, respectively). A strong association between CTS ratings and age was found ($\beta = -0.02$, $p < 0.001$) with younger trainees performing better. Dichotomising age at the median (35 yrs) gave a similar result ($\beta = -0.16$, $p < 0.01$). These results appear in more detail in Table 1.

Essay scores

Normality was not rejected for essay scores 1 or 2, nor case report scores 1 or 2. Essay 3 had a bimodal distribution that makes it difficult to transform the data so that normality is achieved. Thus the results for essay scores must be interpreted with care. In this case, the results for time combined with the rest of the covariates (gender, age and professional background) individually were similar to those obtained for CTS ratings. However, a backward stepwise analysis showed that while nurses generally scored lower than clinical psychologists, the nurses over 35, and female nurses, gained higher scores for their essays than trainees from a clinical psychology background. Details of this analysis are shown in Table 2.

Case report scores

There was no significant effect of Time for case report scores. Univariate random-effect models indicated that clinical psychologists gained higher scores for their case reports than trainees from all other professional backgrounds, and trainees who were younger than 35 gained higher marks than those who were older than 35. Backward stepwise analysis showed that while nursing staff generally scored lower than those from a clinical psychology background, the nurses who were older than 35 scored higher. Details appear in Table 3.

Patterns of change across training

To examine the pattern of change across the training course, a repeated measures ANOVA was carried out on the six CTS ratings across the course. There was a significant effect of Time, $F(5,163) = 53.34$, $p < 0.001$, with a large effect size ($\eta^2 = 0.21$, Cohen's $d = 1.25$). Mean CTS ratings with 95% confidence intervals are shown in Fig. 2. Post-hoc pairwise comparisons showed that there were significant increases in supervisors' CTS ratings within each term i.e., from recordings one to two (3.34 (0.62) vs 3.82 (0.69) $t(256) = 11.98$, $p < 0.001$), recordings three to four (3.84 (0.65) vs 4.15 (0.65) $t(229) = 7.63$, $p < 0.001$), and recordings five to six (3.95 (0.68) vs 4.23 (0.66) $t(204) = 5.34$, $p < 0.001$). Between terms one and two (recordings two and three) there was no significant change (3.83 (0.70) vs 3.84 (0.66) $t(232) = 0.25$, $p = 0.80$). Between terms two and three (recordings four and five), there was a significant decrease in CTS ratings (4.13 (0.66) vs 3.98 (0.68) $t(208) = 2.82$, $p < 0.005$).

'Clinically significant' change

While the above data address the question of whether the trainees' CBT skills improved from pre- to post-training, they do not address the question of what proportion achieved competence according to the various criteria. At a basic level, of the 235 trainees with data for the final CTS rating, by the end of the course 97.4% scored above the lowest competence criterion, i.e. having a CTS item mean of 3.0 or more; 87.1% scored above the middle competence criterion, having an item mean > 3.5; and 63.4% scored above

Table 1
Random-effect models (with both random intercept and slope) fitted to CTS ratings on baseline variables (gender, profession and age). Results with $p < 0.05$ reported.

Independent variables in the model	Intercept (variance)	Time's coefficient β (variance)	Other explanatory variable's coefficient	Correlation of intercept & time	Residual's variance
Time	3.34 (0.25)***	0.16 (0.01)***	–	−0.04	0.31
Time, age	4.12 (0.25)***	0.15 (0.01)***	Age: −0.02***	−0.04	0.29
Time, profession	3.43 (0.24)***	0.16 (0.01)***	Compared to clinical psychology: Psychiatry: −0.19* Nursing: −0.21**		
Time, gender	3.35 (0.25)***	0.16 (0.01)***	Gender: −0.01	−0.04	0.31
Time, age > median (35)	3.44 (0.25)***	0.16 (0.01)***	Age > med: −0.16**	−0.04	0.31
Backward stepwise analysis: final model					
Time, age > med, profession, gender, and their interactions	3.40 (0.24)***	0.16 (0.01)***	Compared to clinical psychology: Psychiatry: −0.16* Nursing: −0.18**	−0.04	0.31

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

the highest competence criteria, having an item mean > 4.0. However, this basic analysis does not take account of the fact that many of the trainees met these competence criteria at the beginning of the course. As a further way of assessing changes in CBT skills, we performed an analysis based on Jacobson's clinical significance approach (Jacobson, Roberts, Berns, & McGlinchey, 1999). This approach involves calculating for each individual whether their scores on the relevant measure have (a) changed enough for the change to be unlikely to be due to chance (using the 'reliable change index', RCI); and (b) crossed a criterion of 'normality' on that measure so that they can be seen as back in normal range. In our adaptation, we calculated a reliable change index for the CTS in the standard way using Jacobson's methodology. The RCI is dependent on the test–retest reliability of a measure and there are very few published data to guide us in estimating this for the CTS. However previous studies give values ranging from 0.68 to 0.96 for reliability (Shaw, 1984; Vallis et al., 1986, footnote 6, p. 383). Putting these estimates into Jacobson's formula gave estimates for reliable change on mean CTS score of between 0.3 and 1.0. We therefore chose the most stringent criterion and set the criterion for reliable change as 1.0 points. Thus anyone changing their mean score from the start to the end of the course by 1.0 or more points (in other words increasing total score on the 11-item CTS by 11 points) was considered to have changed reliably, for better or worse (i.e. the change was unlikely to be due to chance). Of the 230 trainees with ratings for recording one and recording six, 102 reliably improved (44.35%) and five (2.17%) reliably deteriorated, with the remaining 123 (53.48%) not demonstrating reliable change.

The equivalent of 'crossing to normal range' in clinical significance analysis for this study was crossing one of the thresholds of competence described above, i.e. obtaining a mean CTS score that was below the criterion at the start and above the criterion at the end of training. Table 4 shows these results for the 3 different competence criteria.

As with any clinical significance analysis, the above calculations only make sense for people who are below the criterion at the start (otherwise they cannot possibly cross the threshold in the right

Table 2
Random-effect models (with random intercept and fixed slope) fitted to trainees' essay scores on baseline variables (gender, profession and age). Results with $p < 0.05$ reported.

Independent variables in the model	Intercept (variance)	Time's coefficient β	Other explanatory variable's coefficient	Residual's variance
Time	61.37 (5.57)***	1.39***	–	5.23
Time, age	69.40 (5.26)***	1.64***	Age: −0.23***	4.98
Time, gender	62.42 (5.50)***	1.40***	Gender: −1.53 (0.078)	5.23
Time, profession	63.65 (4.84)***	1.39***	Compared to clinical psychology: Psychiatry: −2.03 (0.066) Nursing: −7.21*** Other: −3.79***	5.23
Time, age > median (35)	62.71 (5.46)***	1.39***	Age > med: −2.30**	5.23
Backward stepwise analysis: final model				
Time, age > med, gender, profession, and their interactions	65.91 (4.58)***	1.41***	Age > med: −1.94* Compared to clinical psychology: Nursing: −14.99*** Nursing × age > med: 6.39** Gender × nursing: 4.59*	5.23

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Table 3
Random-effect models (with random intercept and fixed slope) fitted to case report scores on baseline variables (gender, profession and age). Results with $p < 0.05$ reported.

Independent variables in the model	Intercept (variance)	Time's coefficient β	Other explanatory variable's coefficient	Residual's variance
Profession	65.12 (5.68)***	–	Compared to clinical psychology: Psychiatry: −2.97* Nursing: −4.95*** Other: −2.93*	6.11
Age > med	64.55 (5.98)***	–	−2.10*	6.10
Backward stepwise analysis: final model				
Time, age > med, gender, profession, and their interactions	65.66 (5.60)***	–	Compared to clinical psychology: Nursing: −9.16*** Nursing × age > med: 7.35*	6.11

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

direction). In order to give a more complete picture, Table 5 shows the proportions of all trainees who were competent by the different criteria at the start and at the end of training. McNemar tests on each of the three standards for competence separately showed that for all three standards, the proportion achieving that level at the End point was significantly higher than at the Start point (all p's < 0.001). A McNemar–Bowker test on the whole of Table 5 (i.e., across all competence categories) also showed that the Start and End proportions were significantly different, $p < 0.001$.

NB each category is inclusive, so those who 'Met 3.0 criterion' at the Start will also include those who met the 3.5 or 4.0 criterion (because anyone meeting the 3.5 or 4.0 criterion must also have met the 3.0 criterion).

Effect of training on CTS subscales

To examine the effect of training on the three subscales of the CTS a repeated measures MANOVA was used with subscale and time as within subjects factors. There was a main effect of Time, $F(1,231) = 173.6$, $p < 0.001$ with a large effect size ($\eta^2 = 0.43$); a significant effect of Subscale, $F(2, 230) = 83.08$, 23 $p < 0.001$ ($\eta^2 = 0.42$); and a significant Time × Subscale interaction, $F(2, 230) = 42.34$ $p < 0.001$ ($\eta^2 = 0.27$). Post-hoc tests revealed significant increases on all three subscales from recording one to six (see Table 6). To compare the impact of training on the three subscales, change scores were calculated and compared, confirming that trainees made significantly more improvement on the CBT subscale than on the general subscale ($t = 6.02$, $p < 0.001$) or on the Interpersonal Effectiveness subscale ($t = 7.43$, $p < 0.01$).

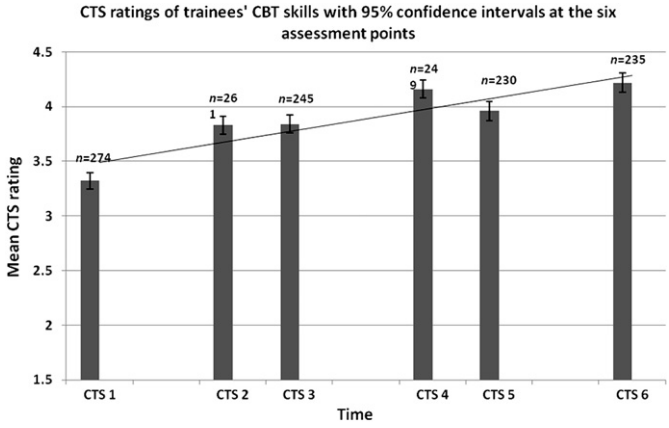

Fig. 2. Supervisors' CTS ratings of trainees' CBT skills with 95% confidence intervals at the six assessment points.

As a confirmatory analysis, univariate and multivariate linear random-effect models were fitted to investigate the effect of time as well as the association of baseline variables (gender, profession and age) to CTS subscale scores. Results from these analyses confirmed a significant effect of time on all three subscales, with the largest change on the CBT subscale followed by the General Therapeutic Skills subscale, then the Interpersonal Effectiveness subscale. As in the analysis of the CTS as a whole, the analysis of the subscales showed no effect of gender on any of the subscale scores, and the effects of age and profession were replicated across all subscales, with younger trainees performing better and trainees from a clinical psychology background performing better than those from nursing or psychiatry.

Discussion

Summary of findings

These results confirm that Diploma training is associated with an increase in scores on both academic assignments (essay scores) and on ratings of clinical skills (CTS ratings). The effect size of $d = 1.25$ on the primary outcome measure (CTS ratings) is larger than has been reported previously (e.g., Keen & Freeston, 2008). The analysis of the reliability of assessment confirms that the two independent markers' assessments of written work were significantly associated on all occasions, as were a sample of supervisors' CTS ratings compared with the ratings of supervisors who were 'blind' to the student and stage of training. It is interesting to note that there was no effect of training on trainees' performance on case reports but this may be due to the scheduling of the two case reports, in that they were submitted at the beginning and end of

Table 4
'Clinical significance' analysis of changes on CTS (only for those not reaching competence at start of course).

	Competence criterion (mean item score on CTS)		
	≥3.0	≥3.5	≥4.0
Reliably worsened	0 (0%)	0 (0%)	2 (1.0%)
No reliable change	13 (23.6%)	51 (37.8%)	94 (47.7%)
Reliably improved but not reached competence	0 (0%)	1 (0.7%)	11 (5.6%)
Reliably improved *and* reached competence	42 (76.4%)	83 (61.5%)	90 (45.7%)
Total n in analysis	55	135	197

Table 5
Proportion of trainees ($n = 232$) achieving competence on CTS at the start and end of the course.

End competence category	Start competence category *n* in cell (percent of relevant start category)			
	Below 3.0 criterion	Met 3.0 criterion	Met 3.5 criterion	Met 4.0 criterion
Below 3.0 criterion	2 (3.6%)	5 (2.8%)	3 (3.1%)	0 (0%)
Met 3.0 criterion	6 (10.9%)	16 (9.0%)	6 (6.2%)	3 (8.6%)
Met 3.5 criterion	17 (30.9%)	44 (24.9%)	27 (27.8%)	10 (28.6%)
Met 4.0 criterion	30 (54.5%)	112 (63.3%)	61 (62.9%)	22 (62.9%)
Total *n*	55	177	97	35

the third term of training, meaning any impact of training would have been from that term only.

The pattern of change in CBT competence across the course was consistent with expectations in that the highest ratings were at time six and the lowest ratings were at time one. As there was no control group, it is possible that the increase in competence was due to the passage of time or other factors. However, the patterns of change observed are not entirely consistent with this suggestion. For example, there were significant increases in CTS ratings within each term, but not between the terms (when there was no active training input). In addition, trainees' greater improvement on the CBT subscale of the CTS than on the General Skills or Interpersonal Effectiveness subscales may suggest a specific training effect. Interestingly, while there was no change in CTS ratings between terms one and two, there was a significant *decrease* in ratings between term two and three. The most likely explanation for this is the impact of increased patient complexity on the CTS ratings. It has been suggested that competence ratings may be influenced by the patient's suitability for CBT (James et al., 2001) and that certain patient characteristics require greater therapist skill (Whisman, 2008). In support of this suggestion, Strunk et al. (2010) found evidence that competence was more highly related to outcome for patients with greater severity of symptoms, earlier age of onset and (at trend level) chronicity. In the first two terms our trainees were instructed to select 'straightforward' patients whereas in the third term trainees were asked to select more complex cases. Hence, it may be that increased patient complexity makes it more difficult to demonstrate CBT competence and this accounts for the significant decrease in trainees' CTS ratings between the end of term two and the beginning of term three.

In examining the proportion of trainees who achieved 'clinically significant change', our results suggest that the majority of trainees reliably improve and reach competence. Less than 1% reliably deteriorate, and 45–76% reliably improved and reached the various criteria for competence. By all criteria for defining competence, significantly more trainees had attained competence by the end of the course than at the beginning. Using the most lenient criteria for competence (mean CTS rating ≥ 3.0) only 2.6% of trainees failed to reach competence by the end of training. This compares favourably with previous reports that up to 50% of sessions failed to meet this standard (Brosan et al., 2006; Trepka et al., 2004). Using the 'medium' standard of a mean score above 3.5, Shaw et al. (1999) reported that 27% of sessions in the Treatment of Depression Collaborative Research Program were below this level of competence, whereas in the current study only 12.9% of trainees' final assessments were rated below this standard. However, in Shaw et al.'s (1999) analysis the sessions evaluated were randomly selected, whereas in the current sample and in Brosan et al.'s (2006) sample the trainee/therapist *selected* the recording to be evaluated, most likely leading to their filtering out poorer performances. It is also worth noting that by Shaw's (1984) original 'red line value' of a total score of 44 on the 11-item version of the CTS, corresponding to a mean item score of 4.0 or above, there remains room for improvement in outcomes from CBT training, with 36.6% of the current sample not meeting this standard by the end of training on a self-selected recording.

Table 6
A comparison of the change achieved by trainees on the three subscales of the CTS ($n = 230$).

CTS subscale	Mean (*SD*) on recording one (beginning of training)	Mean (*SD*) on recording six (end of training)	*t*(229)
CBT	3.21 (0.72)	4.19 (0.71)	15.06*
GTS	3.47 (0.74)	4.23 (0.77)	12.36*
IE	3.75 (0.77)	4.33 (0.80)	8.00*

*$p < 0.001$. CBT = Cognitive-behavioural skills subscale; GTS = General Therapeutic Skills subscale; IE = Interpersonal Effectiveness subscale.

Effect of therapists' characteristics

Trainees' gender did not have any influence on training outcomes but trainees' age had an effect, with older trainees performing worse than younger trainees on all assessment types. There was also an effect of professional background with trainees from a clinical psychology backgrounds performing better than those from either psychiatry or nursing on all assessments of CBT competence. The exception to this pattern was that older female nurses scored higher than clinical psychologists on both essays and case reports, but not on CTS ratings. The finding that clinical psychologists perform better than other professional groups on academic assignments is perhaps unsurprising given that the marking of the essays and case reports has largely been done by clinical psychologists, so there may be a transfer of academic standards. However, this does not account for the better performance of clinical psychologists on CTS ratings as these ratings are made by a larger pool of supervisors from a range of professional backgrounds, including nursing and psychiatry. The superior performance of clinical psychologists is likely to be a factor of their having had more previous exposure to CBT methods and models in their core professional training – clinical psychology training in the UK requires two degrees in comparison to only one being required for nursing or psychiatry, and clinical psychology training concentrates on training in psychological interventions (psychological therapies) whereas medical and nurse training is necessarily broader. The superior performance of clinical psychologists may also reflect the fact that there are significant barriers to entry to clinical psychology training (only 23–29% of applicants are successful in gaining a place), which may mean more filtering out of weaker candidates (see Clearing House for Post-graduate Courses in Clinical Psychology http://www.leeds.ac.uk/chpccp/BasicNumbers.html for more detail).

The poorer performance of older trainees requires further investigation, as one previous study reported no relationship between trainees' age and the acquisition of competence in CBT training (James et al., 2001) and several studies have reported no relationship between therapists' age and patient outcomes (e.g., Beck, 1988; Greenspan & Kulish, 1985). However, in the current study, with the exception of older female nurses (who performed better on written assignments), older trainees (36 years and older) performed worse than younger trainees on all measures of CBT competence. It may be that it simply is harder 'to teach an old dog new tricks', but there are three other, highly speculative, possible explanations for the poorer performance of older trainees. First, there may be a selection bias with the more capable/CBT competent students managing to gain a place on the course at a younger age. Second, it is our impression that older trainees may be less motivated to learn 'standard' CBT, and have less opportunity to practise it. The older trainees generally come to the course with more previous clinical experience and often have a non-CBT therapeutic model that they have been working within for some

time. Thus they are more inclined to want to integrate CBT methods into their pre-existing framework, in contrast to the younger trainees who seem to be more motivated to learn standard CBT 'from the ground up'. Finally, older trainees are often in more senior roles meaning they spend less time in direct clinical work so have less opportunity to practise their developing CBT skills, and they tend to see more complex cases, possibly making it harder for them to demonstrate competence on the CTS. While the above discussion may throw some light on why the performance of older trainees is poorer, and why clinical psychologists perform better than other professional groups, it does not account for the superior performance of older, female nurses on academic assignments. This finding has not been previously reported and requires replication and further investigation.

Interpretation and conclusions

In general, the results of the current study support the recent expansion of Diploma-level CBT training in the UK – the general pattern of results confirms that such training does indeed lead to increased CBT competence, with the majority of trainees meeting pre-determined standards of competence by the end of training. This is encouraging as previous research has suggested that clinicians who demonstrate levels of fidelity comparable with levels seen in benchmark efficacy trials have comparable outcomes (Henggeler, Melton, Brondino, Scherer, & Hanley, 1997). So the finding that Diploma-level CBT training increases competence, with the majority achieving the levels demonstrated in trials, in combination with previous research demonstrating an association between greater CBT competence and improved patient outcomes, suggests that the government's investment in providing CBT training will lead to improved outcomes for patients. In addition, a concern for the dissemination of CBT has been that the results of CBT efficacy trials will not be replicated in routine clinical practice due the therapists having lower levels of CBT skills. This concern is not supported by the current data in which evaluations of therapists' skills at the completion of their training compare favourably with both previous studies and with the competence criteria used in efficacy trial.

It has been noted that there "is an urgent need for research on efficient ways of disseminating treatment procedures" (Shafran et al., 2009, p. 905). The results of this uncontrolled single sample study can say little about the most effective type of training. However, the results are consistent with previous suggestions that CBT training must be comprehensive and include supervision in order to have an enduring effect on competence (Brosan et al., 2006; Mannix et al., 2006; Sholomskas et al., 2005). In their review of issues in the dissemination of CBT, Taylor and Chang (2008) conclude: "the conventional wisdom is that workshops alone are not sufficient to help even experienced therapists achieve competence and that follow-up supervision is necessary" (p. 41). Consistent with this suggestion, Diploma-level training is a comprehensive program that includes case supervision and yields significant increases in CBT competence as measured by ratings on the CTS. Similarly, previous reports have suggested a dose–effect relationship of CBT training with studies with fewer hours training reporting more limited impact than those with more hours (e.g., Rakovshik & McManus, 2010; Shafran et al., 2009). It is worth noting that the programme evaluated here involved approximately 235 h of workshops and small group supervision.

Limitations

The current study overcomes some limitations of previous evaluations of CBT Diploma training by using a larger sample size and a greater variety of measures of CBT competence. Nevertheless, the uncontrolled retrospective nature of the current study means that results cannot be definitive. It is not known what changes in CBT competence, if any, would have been seen in this group of clinicians in the absence of training. However, given that several previous studies have failed to demonstrate any effect of training on CBT competence (e.g., Miller & Mount, 2001; Morgenstern, Blanchard, Morgan, Labouvie, & Hayaki, 2001 – see Beidas & Kendall, 2010 for a review), or much smaller effects sizes than those reported here, it seems unlikely that the improvements in competence seen here are due to the passage of time alone. It is also worth noting that for the majority of assessments, assessors were not blind to the stage of training, which may have influenced their assessments in either direction, particularly towards the end of the training course. Supervisors may make less generous CTS ratings towards the end of the course due to their expectations of what the trainee should be able to demonstrate by this stage being higher. Alternatively, they may make more generous ratings of competence towards the end of the course due to the demand characteristic of the expectation that trainees would pass the course. The weaker relationship between supervisors' and 'blind' CTS ratings on recording 6 than on recording 1 suggests supervisors' ratings may be less reliable at the end of training than at the beginning. Or it may be that the CTS is less reliable when used with more highly skilled therapists, or with therapists treating more complex cases as happens in the final term of this course. However, even 'blind' assessments confirm a significant increase in CBT training from pre- to post-training suggesting that this type of training programme is effective in enhancing CBT therapists' competence and thus is likely to lead to improved clinical outcomes for patients receiving CBT in routine clinical practice settings.

A further limitation of this study arises from a limitation of the field. Several recent reviews have highlighted that we do not yet have a consensus about how CBT competence is conceptualized and defined, let alone how it can be effectively measured (e.g., Rakovshik & McManus, 2010; Shafran et al., 2009; Sharpless & Barber, 2009). And several limitations of the current 'gold standard' for assessing CBT competence ('expert' assessor ratings on the CTS) have been noted above. While other scales for assessing therapists' CBT competence have been developed (e.g., Barber et al., 2003; Blackburn et al., 2001) they have yet to undergo extensive validation. Further studies are warranted to more extensively validate existing measures of CBT competence (e.g., to establish empirically derived 'cut-offs') and to keep pace with developments in CBT (e.g., well validated disorder or protocol specific measures of CBT competence). Additionally it may also be necessary to triangulate various measures of competence, across sessions, settings and patients, to have any definitive measurement of competence.

Acknowledgements

The authors are grateful to Kim Morris-Smith, Jean Kennedy and Kate Muse for assistance with data collection and entry.

References

Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research, 13*, 205–221.

Beck, D. F. (1988). *Counselor characteristics: How they affect outcomes*. Milwaukee, WI: Family Service America.

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: a critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science & Practice, 17*, 1–30.

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*, 431–446.

Brosan, L., Reynolds, S., & Moore, R. G. (2006). Factors associated with competence in cognitive therapists. *Behavioural and Cognitive Psychotherapy, 35*, 179–190.

Chevron, E., & Rounsaville, B. J. (1983). Evaluating the clinical skills of psychotherapists: a comparison of techniques. *Archives of General Psychiatry, 40*, 1129–1132.

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.

Clearing House for Post-Graduate Courses in Clinical Psychology. Retrieved 30.07.10, from http://www.leeds.ac.uk/chpccp/BasicNumbers.html.

Cognitive Therapy in Oxford diploma course. (n.d.). Retrieved 30.07.10 from http://www.octc.co.uk/content.asp?PageID=53.

Davidson, K., Scott, J., Schmidt, U., Tata, P., Thornton, S., & Tyrer, P. (2004). Therapist competence and clinical outcome in the prevention of parasuicide by manual assisted cognitive behaviour therapy trial: the POPMACT study. *Psychological Medicine, 34*, 855–863.

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*, 658–670.

Dobson, K. S., & Shaw, B. F. (1988). The use of treatment manuals in cognitive therapy: experience and issues. *Journal of Consulting and Clinical Psychology, 56*, 673–680.

Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24*, 295–300.

Greenspan, M., & Kulish, N. M. (1985). Factors in premature termination in long-term psychotherapy. *Psychotherapy, 22*, 75–82.

Grey, N., Salkovskis, P., Quigley, A., Clark, D. M., & Ehlers, A. (2008). Dissemination of cognitive therapy for panic disorder in primary care. *Behavioural and Cognitive Psychotherapy, 36*, 509–520.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multi-systemic therapy with violent and chronic offenders and their families: the role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821–833.

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*, 774–781.

Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38*, 21–33.

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*, 103–117.

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application and alternatives. *Journal of Consulting & Clinical Psychology, 67*, 300–307.

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

Keen, A. J. A., & Freeston, M. H. (2008). Assessing competence in cognitive-behavioural therapy. *The British Journal of Psychiatry, 193*, 60–64.

King, M., Davidson, O., Taylor, F., Haines, A., Sharp, D., & Turner, R. (2002). Effectiveness of teaching general practitioners skills in brief cognitive behaviour therapy to treat patients with depression: randomised controlled trial. *British Medical Journal, 324*, 947–950.

Kingdon, D., Tyrer, P., Seivewright, N., Ferguson, B., & Murphy, S. (1996). The Nottingham study of neurotic disorder: influence of cognitive therapists on outcome. *British Journal of Psychiatry, 169*, 93–97.

Kuyken, W., & Tsivrikos, D. (2009). Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychotherapy and Psychosomatics, 78*, 42–48.

Layard, R. (2006). The case for psychological treatment centres. *British Medical Journal, 332*, 1030–1032.

Mannix, K. A., Blackburn, I. M., Garland, A., Gracie, J., Moorey, S., Reid, B., et al. (2006). Effectiveness of brief training in cognitive behaviour therapy techniques for palliative care practitioners. *Palliative Medicine, 20*, 579–584.

McManus, F., Grey, N., & Shafran, R. (2008). Cognitive therapy for anxiety disorders: current status and future challenges. *Behavioural and Cognitive Psychotherapy, 36*, 595–704.

Miller, W. R., & Mount, K. A. (2001). A small study of training in motivational interviewing: does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy, 29*, 457–471.

Milne, D., Baker, C., Blackburn, I. M., James, I. A., & Reichelt, F. K. (1999). Effectiveness of cognitive therapy training. *Journal of Behavior Therapy and Experimental Psychiatry, 30*, 81–92.

Morgenstern, J., Blanchard, K. A., Morgan, T. J., Labouvie, E., & Hayaki, J. (2001). Testing the effectiveness of cognitive-behavioral treatment for substance abuse in a community setting: within treatment and posttreatment findings. *Journal of Consulting and Clinical Psychology, 69*, 1007–1017.

National Institute for Clinical Excellence. (2004a). *Anxiety: Management of anxiety (panic disorder, with or without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care* (Clinical guideline 22). London: NICE.

National Institute for Clinical Excellence. (2004b). *Depression: Management of depression in primary and secondary care* (Clinical guideline 23). London: NICE.

Rakovshik, S., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: a review of current empirical findings and theoretical guidance. *Clinical Psychology Review, 30*, 496–516.

Roth, A., & Fonagy, P. (1996). *What works for whom? A critical review of psychotherapy research*. New York: Guilford Press.

Shafran, R. S., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., et al. (2009). Mind the gap: improving the dissemination of CBT. *Behaviour Research and Therapy, 47*, 902–909.

Sharpless, B. A., & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review, 29*, 47–56.

Shaw, B. F. (1984). Specification of the training and evaluation of cognitive therapy for outcome studies. In J. B. W. Williams, & R. L. Spitzer (Eds.), *Psychotherapy research: Where we are and where we should go* (pp. 92–128). New York: Guilford.

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*, 837–846.

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*, 106–115.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research, 7*, 301–317.

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*, 595–606.

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*, 429–437.

Taylor, C. B., & Chang, V. Y. (2008). Issues in the dissemination of cognitive-behaviour therapy. *Nordic Journal of Psychiatry, 62*, 37–44.

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*, 143–157.

Vallis, M. T., Shaw, B. F., & Dobson, K. S. (1986). The cognitive therapy scale: psychometric properties. *Journal of Consulting & Clinical Psychology, 54*, 381–385.

Walters, S. T., Matson, S. A., Baer, J. S., & Ziedonis, D. M. (2005). Effectiveness of workshop training for psychosocial addiction treatments: a systematic review. *Journal of Substance Abuse Treatment, 29*, 283–293.

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*, 569–579.

Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114*, 248–265.

Whisman, M. A. (2008). *Adapting cognitive therapy for depression: Managing complexity and co-morbidity*. New York: Guilford.

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 Psychotherapy, 19*, 373–376.

Young, J., & Beck, A. T. (1980). *Cognitive therapy scale: Rating manual*. Unpublished manuscript, University of Pennsylvania, Philadelphia, PA.

Young, J., & Beck, A. T. (1988). *Revision of cognitive therapy scale*. Unpublished manuscript, University of Pennsylvania, Philadelphia, PA.