The Relationship between Competence and Patient Outcome with Low-Intensity Cognitive Behavioural Interventions

Amanda Branson

School of Psychology and CLS, University of Reading, Earley Gate, Reading

Pamela Myles

School of Psychology and CLS, University of Reading, Earley Gate, Reading

Mishka Mahdi

University College London Great Ormond Street Institute of Child Health, Gower Street, London

Roz Shafran

University College London Great Ormond Street Institute of Child Health, Gower Street, London

Background: Little is understood about the relationship between therapist competence and the outcomes of patients treated for common mental health disorders. Furthermore, the evidence is yet to extend to competence in the delivery of low-intensity cognitive behavioural interventions. Understanding this relationship is essential to the dissemination and implementation of low-intensity cognitive behavioural interventions. ***Aims***: The aim of this study was to explore the relationship between Psychological Well-being Practitioner (PWP) competence and patient outcome within the framework of the British government's Improving Access to Psychological Therapies (IAPT) initiative. ***Method***: Forty-seven PWPs treating 3688 patients participated. Relationships between PWP scores on three observed standardized clinical examinations and reliable change in patients' symptoms of anxiety and depression were explored at two time points: during the year-long training phase, and over a 12-month follow-up. ***Results***: Results indicated that patients treated by qualified PWPs achieved superior outcomes than those treated by trainees. Little support was found for a general association between practitioner competence in delivering low-intensity cognitive behavioural interventions and patient outcome, either during or post-training; however, significantly more patients of the most competent PWPs demonstrated reliable improvement in their symptoms of anxiety and depression than would be expected by chance alone and fewer deteriorated compared with those treated by the least

---

Correspondence to Amanda Branson, University of Reading, Earley Gate, Reading RG6 6AL, United Kingdom. E-mail: a.branson@reading.ac.uk

competent PWPs. **Conclusion**: Results were indicative of a complex, non-linear relationship, with patient outcome affected by PWP status (trainee or qualified) and by competence at its extremes. The implications of these results for the dissemination and implementation of low-intensity cognitive behavioural interventions are discussed.

Keywords: IAPT, training, depression, anxiety, low-intensity, PWP, psychological well-being practitioner

Introduction

In recent years there has been an exponential increase in evidence for the efficacy of psychological interventions in the treatment of anxiety and depression, and a paralleled increase in demand for these interventions (Williams, 2015).

Low-intensity interventions such as cognitive behaviour therapy (CBT) in the form of guided self-help (Cuijpers and Schuurmans, 2007) and computerized cognitive behavioural therapy (CCBT; see Ferriter et al., 2008) are recommended by the National Institute for Health and Care Excellence (NICE) for the treatment of mild to moderate anxiety and depression (NICE, 2011). These cognitive behavioural treatments offer a patient-led, lower dose of psychological treatment, with less contact time with the mental health practitioner than traditional CBT (Turpin, 2008), and are considered less burdensome for patients (Bennett-Levy, 2010).

In order to meet demand, the British government has to date invested in excess of £700 million in the Improving Access to Psychological Therapies (IAPT) initiative, a programme aimed towards establishing nationwide accessibility to evidence-based treatments, delivered by highly trained therapists. Further to training experienced therapists to deliver CBT, there has been the development of a new work force – Psychological Well-being Practitioners (PWPs) – to deliver evidence-based low-intensity interventions, such as those listed above. PWPs require less experience than CBT therapists; many are recent graduates with an interest in, and some experience in, mental health. In order to widen the pool of PWPs, training is also available to those with more extensive relevant experience, but who do not hold an undergraduate degree.

PWP training, provided in partnership between the National Health Service (NHS) and universities, combines curriculum-led teaching with clinical practice. PWPs are expected to master a broad array of competences in order to effectively deliver treatment (Roth and Pilling, 2008), whilst maintaining a high patient case-load. The clinical competence of PWPs during the study period was measured via observed standardized clinical examination (OSCE). Whilst used routinely in medical training, OSCE has only recently been applied to the measurement of competence in therapist training (e.g. Manring et al., 2003) and it is only routinely used within the IAPT framework. Although it has been suggested that OSCE may offer a more reliable approach to the measurement of competence than existing approaches such as the rating of therapy sessions (Fairburn and Cooper, 2011; Ginzburg et al., 2012; Muse and McManus, 2013), this is yet to be empirically tested.

Evidence for the relationship between clinical competence and patient outcome is broadly supportive of a positive effect of training on therapist competence (e.g. Barnfield et al., 2007; Keen and Freeston, 2008; McManus et al., 2010), and of a positive effect of therapist competence on patient outcomes (e.g. Green et al., 2014; Okiishi et al., 2003; Saxon and Barkham, 2012); however, relationships appear to be complex (Branson et al., 2015; Muse and McManus, 2013) and have yet to be explored in relation to competence in low-intensity treatment, or via the use of OSCE. The aim of the current study was to provide a first

exploration of the relationship between clinical competence and patient outcome with low-intensity interventions using OSCEs.

It is thus hypothesized that:

- Training will improve the competence of PWPs, as measured by patient outcome. The proportion of patients achieving reliable improvement in their symptoms will be greater for qualified PWPs than for PWP trainees.
- There will be a positive association between PWP competence, as measured by OSCE, and reliable change in patients' symptoms of depression and anxiety.
- Patients treated by the most competent PWPs, as measured by OSCE, will be more likely to achieve reliable improvement in their symptoms than those treated by the least competent. Conversely, patients treated by the least competent PWPs will be more likely to experience a reliable deterioration in symptoms than those treated by the most competent.

Methods

Training course

The IAPT Graduate/Postgraduate Certificate in Low-Intensity Interventions at the University of Reading is a year-long training programme combining theoretical and experiential learning with routine clinical practice, in partnership with National Health Service (NHS) IAPT service sites. The University of Reading's PWP training programme is accredited by the British Psychological Society (BPS).

Training followed a national curriculum[1] consisting of four modules that encompassed: engagement and assessment of patients with common mental health problems; evidence-based low-intensity cognitive behavioural treatment; values, policy, culture and diversity; and working within an employment, social and healthcare context. Modules 1 and 2 were taught intensively to ensure that trainees had the core knowledge, skills and attitudes necessary to competently assess and treat patients. Trainees were required to pass Module 1 and two assessments prior to being deemed competent to treat patients. Modules 3 and 4 were shorter in duration and were merged into one academic term at the University of Reading. Following successful completion of modules 1 and 2, trainees were expected to maintain a high case-load of up to 250 patients per year,[2] whilst receiving case management supervision in the field to ensure treatment fidelity (Department of Health, 2010).

Participants

Patients with mild to moderate anxiety and/or depression were referred or self-referred to one of five Thames Valley Local Education and Training Board (LETB) IAPT service sites (formerly South Central Strategic Health Authority, SC-SHA). Patients experiencing mild to moderate symptoms typically receive low-intensity interventions; if treatment is unsuccessful, patients may transition (step up) to a full course of CBT delivered by a high-intensity therapist.

---

[1] The national curriculum has since been revised (see UCL, 2015); descriptions of the training course and assessments herein refer to the curriculum that was in place during the study period, 2008–2012.

[2] Trainees typically manage 80% of a full case-load by month 10 of training.

Patients were included in the study if their current course of treatment was low intensity, i.e. being treated by a PWP. This approach was taken to allow for direct comparison with national data (Gyani et al., 2013), and also to control for any affect associated with prior treatment. Additionally, patients must have received two or more sessions for which outcome data were available, their treatment must have been conducted by one participating PWP, and they must have been discharged from the low-intensity service. Change scores for 3688 patients meeting the above criteria were included in the analyses. Data were collected in two phases: patients treated by trainees during the 12-month training period ('during') ($n = 2009$) and patients treated by qualified PWPs in the 12-month period after training ('post') ($n = 1679$)[3].

Practitioners were PWPs trained on the University of Reading's Postgraduate Certificate in Evidence-Based Psychological Therapies, between academic years 2008/2009 and 2011/2012. All trainee PWPs ($n = 124$) were invited to take part. Of those, 79 were employed in participating SC-SHA services, 47 of whom consented to be involved. The mean age was 30.6 years ($SD = 7.69$), and 34 (72.3%) were female. Mean pre-training experience was 2.15 years ($SD = 2.22$); 36 (76.6%) were female. All participating practitioners were graduates; of those who gave information on the subject studied ($n = 46$), 37 (80.4%) studied psychology, four (8.7%) studied another social science, and five (10.9%) studied other subjects. Six (12.8%) achieved a first class pass in their degree, 34 (72.3%) were awarded an upper second (2:1), six (12.8%) a lower second (2:2) and one trainee received a third class pass (2.1%).

Chi-squared tests found that the distribution of males to females did not differ significantly ($\chi^2(4,47) = 1.93, p = .75$), nor did the distribution of degree award classifications ($\chi^2(12,47) = 12.73, p = .39$). One-way ANOVAs revealed no significant differences in age, or years of experience ($p > 0.05$ for all) across services.

Measures

Patient. At each point of contact, patients completed the depression module of the Patient Health Questionnaire (PHQ-9: Kroenke et al., 2001) and the Generalized Anxiety Disorder questionnaire (GAD-7: Spitzer et al., 2006).

The PHQ-9 measures the nine DSM-V (American Psychiatric Association, 1994) diagnostic criteria for depression on a 0–3 point Likert-style scale. It has been found to be both reliable and valid in the facilitation of recognition and diagnosis of depression within primary care, and can also be used to monitor symptom change over the course of treatment. A score greater than 9 indicates clinically significant symptoms of depression (IAPT, 2011) .

The GAD-7 is a 7-item self-report scale that has been found to be both reliable and valid as a tool for screening for and assessing the severity of GAD in clinical settings (Spitzer et al., 2006). It has also been found to measure symptoms of panic disorder, social anxiety disorder and post-traumatic stress disorder moderately well (IAPT, 2011). The GAD-7 is measured on a 0–3 point Likert scale; a score greater than 7 indicates clinically significant symptoms of anxiety.

Practitioner. Clinical competence was examined through three OSCEs, which were rated using criteria, and assessable learning outcomes set out in ReachOut, the national curriculum

---

[3] Due to the timing of data collection, follow-up patient data (post) were not available for the last two cohorts of trainees ($n = 22$).

for PWPs (Richards and Whyte, 2009). OSCEs took place on completion of each taught module, roughly 3, 6 and 9 months into training, respectively. ReachOut scales are competency based, each examining the assessable learning outcomes set out in the training manual. OSCE 1 examines patient assessment, including: knowledge and understanding of mental illness, diagnostic criteria and recognition of symptoms; competence with common factors such as patient engagement and the therapeutic alliance; information gathering including recognition and assessment of risk; information giving and shared decision making. OSCE 2 examines competence in treatment delivery, including: ability to demonstrate skills in planning and implementing a low-intensity treatment intervention such as behavioural activation, graded exposure, worry management, cognitive restructuring, and guided self-help; competence with common factors such as developing and maintaining the therapeutic alliance; also ongoing risk management. OSCE 3 examines the use of case management supervision, including: preparation for supervision, prioritization of cases for discussion in supervision, risk assessment, reflection on own competence and the ability to work within a multi-disciplinary team, with awareness of the boundaries of low-intensity treatment.

OSCEs are rated 0–100, with a pass mark of 50. Raters were clinicians who were also experienced PWP educators employed by the University; raters underwent annual rater-training. Following submission, raters blind-marked one OSCE to ensure acceptable inter-rater reliability (see below). A further sample of OSCE 1 (25.5%), OSCE 2 (17.0%) and OSCE 3 (8.5%) were moderated by the director of training, who is a British Association for Behavioural and Cognitive Psychotherapies (BABCP) accredited therapist, supervisor and educator. Samples of work were scrutinized by an external examiner in accordance with University regulations.

Results

Preliminary analyses

To investigate whether OSCE scores differed across service sites, a series of one-way ANOVAs were conducted. Significant differences in OSCE 1 scores emerged across service sites (F (4,46) = 4.001, mean square error (MSE) = 120.68, $p = .008$). This was explained by PWPs in service 3 achieving significantly higher OSCE 1 scores (mean = 82.38, SD = 10.36) than did service 5 (mean = 60.8, SD = 5.76). Scores on OSCE 2 and OSCE 3 did not differ significantly by service site (F (4,46) = 1.66, MSE = 156.46, $p = .16$; F (4,46) = 2.07, MSE = 108.94, $p = .10$ for OSCE 2 and OSCE 3, respectively).

Inter-rater reliability

Per cohort of PWPs ($n = 5$) one of each OSCE (1–3) was marked blind by three raters. Due to the naturalistic design, OSCEs were rated by different sets of raters, selected from the larger population of raters ($n = 11$). Where it is not possible to separate effects of rater, interaction between rater and assessment, and random error (Shrout and Fleiss, 1979) argue for the use of one-way random effects intra-class correlation (ICC) with absolute agreement. Inter-rater reliability for all assessments ranged from fair to substantial (Shrout, 1998). Reliability was greatest for OSCE 3 (ICC = .84, $p = .004$), followed by OSCE 2 (ICC = .71, $p = .019$);

reliability of OSCE 1 (ICC = .52, p = .029) was lowest, but still comfortably within the 'fair' range.

Practitioner competence

Mean OSCE scores were 73.3% (range: 52–98%, *SD* = 12.36), 70.6% (range: 50–100%, *SD* = 12.86) and 79.1% (range: 59–100%, *SD* = 10.91) for OSCE 1–3, respectively. As each OSCE measures a different, though overlapping set of competences, it was not appropriate to explore change in competence over time. One sample *t*-tests confirmed that PWP scores for each OSCE were significantly higher than the threshold of competence (50%) (t (46) = 12.95, $p < .001$; t (46) = 10.98, $p < .001$; t (46) = 18.29, $p < .001$ for OSCEs 1–3, respectively).

Patient outcome

Change in symptoms of anxiety and depression were calculated by subtracting final GAD-7 and PHQ-9 scores from first scores; a negative score reflects a reduction in symptoms. Jacobson and Truax's (1991) methodology was used to calculate statistically reliable change. The criterion for reliable change was calculated to be 3 on the PHQ-9, and 2 on the GAD-7. The indices are small, but this is likely to reflect low measurement error, first as inter-rater reliability reported in the validation studies (Kroenke et al., 2001; Spitzer et al., 2006) was good (Ye, 2010), and second because the standard deviation of mean first outcome measure scores and threshold for recovery were small (PHQ-9: mean = 11.81, threshold = 9, *SD* = 2.70; GAD-7: mean = 6.99, threshold = 7, *SD* = 2.09). Standard deviations are small as patients treated at step 2 typically present with symptoms in the mild to moderate range on the PHQ-9 and GAD-7, thus the difference between first score, and threshold for recovery is narrower than that of patients treated at step 3 who typically present with symptoms in the moderate to severe range.

Combining PHQ-9 and GAD-7 reliable change scores using the approach taken in an evaluation of year 1 IAPT patient data (Gyani et al., 2013), these data are presented in Table 1. The percentage of patients who experienced a reliable improvement in their symptoms of anxiety and depression was significantly higher than that demonstrated nationally (63.8%; Gyani et al., 2013) (t (46) = 2.91, p = .006).

Effect of time on patient outcome

Existing research has shown that training improves competence as measured by assessments (e.g. Barnfield et al., 2007; Keen and Freeston, 2008; McManus et al., 2010) and patient outcome (e.g. Westbrook et al., 2008) . Mann–Whitney *U*-tests were conducted to explore whether the outcomes of patients treated by trainee PWPs differed from those of patients treated by qualified PWPs. PWPs treating fewer than five patients within that time frame and those with data available for one time frame only were excluded. Results indicated that the proportion of patients who reliably improved was greater for patients treated by qualified PWPs (post-training) (mean = 0.68) than it was for patients treated by trainees (during training) (mean = 0.72, U = 1597500.00, p = .004, r = .05), whilst the proportion of patients who reliably deteriorated was greater for patients treated by trainees (during training) (mean = 0.10) than it was for patients treated by qualified PWPs (post-training) (mean = 0.06, U = 1621783.50, p = .002, r = .05).

Table 1. Frequency and percentage of patients reliably improving and reliably deteriorating in their symptoms of depression (PHQ-9) and anxiety (GAD-7) by time frame

| | | Reliable improvement | | | | Reliable deterioration | | | | Combined reliable improvement[1] | | Combined reliable deterioration[2] | |
| | | PHQ-9 | | GAD-7 | | PHQ-9 | | GAD-7 | | PHQ-9 and/or GAD-7 | | PHQ-9 and/or GAD-7 | |
	Total patients	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
During training	2009	1206	60.0%	1051	52.3%	166	8.3%	130	6.5%	1359	67.6%	208	10.4%
Post-training	1679	1098	65.4%	967	57.6%	113	6.7%	66	3.9%	1212	72.2%	112	7.3%
All time points	3688	2304	62.5%	2018	54.7%	279	7.6%	196	5.3%	2571	69.7%	330	8.9%

[1]To be considered as showing combined reliable improvement, a patient needed to show reliable improvement on the PHQ-9 and/or the GAD-7; if reliable improvement occurred on one measure only, there should be no deterioration on the other.
[2]To be considered as showing combined reliable deterioration, a patient needed to show reliable deterioration on the PHQ-9 and/or the GAD-7; if reliable deterioration occurred on one measure only, there should be no improvement on the other.
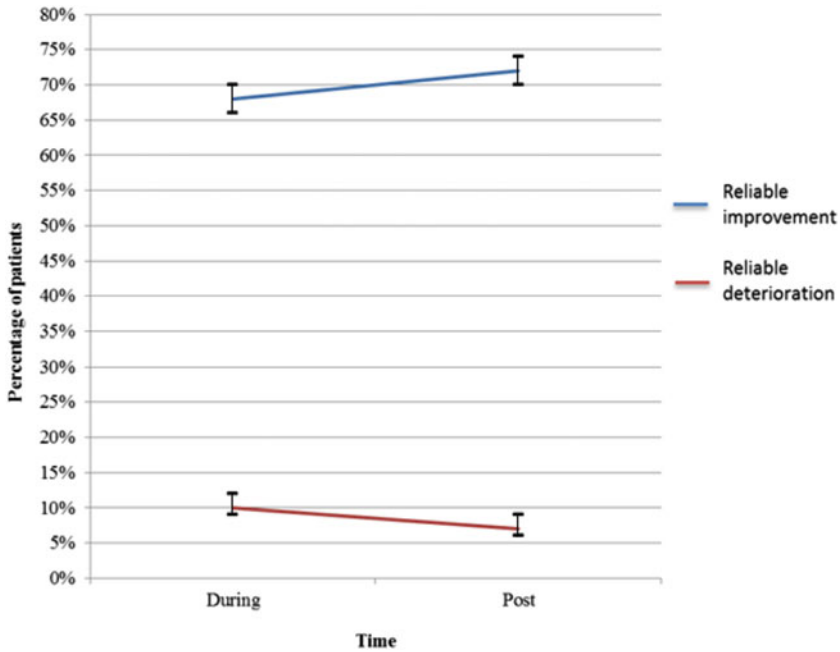
Figure 1. (Colour online) Graph showing mean percentage of patients reliably improving and deteriorating by time frame, with upper and lower bound confidence intervals.

Effect of competence on patient outcome

Spearman rho correlations between OSCEs 1–3, and reliable change in patients' symptoms (see Table 2) were inconsistent, and predominantly in the modest range. A small number of associations reached significance: OSCE 1 was negatively associated with reliable change on the PHQ-9 for patients treated during training ($r = -.341$, $p = .03$), and positively with the proportion of patients reliably improving in their symptoms of anxiety and depression[4] ($r = .405$, $p = .01$). No associations reached significance for OSCE 2 or OSCE 3. When all patients were included (all times), significant negative associations emerged only between OSCE 3 and reliable change on the GAD ($r = -.358$, $p = .02$) and the proportion of patients reliably deteriorating ($r = -.441$, $p = .003$).

Further analyses were conducted to explore non-linear effects of competence on patient outcome. Quadratic regression models were found to be non-significant for the relationship between OSCE 1 ($R^2 = 0.012$, $F(2,44) = 0.27$, $p = .85$; $R^2 = 0.018$, $F(2,44) = 0.393$, $p = .68$), OSCE 2 ($R^2 = 1.3$, $F(2,44) = 3.149$, $p = .05$; $R^2 = 0.085$, $F(2,44) = 2.054$, $p = .14$) and OSCE 3 ($R^2 = 0.067$, $F(2,44) = 1.586$, $p = .22$; $R^2 = 0.026$, $F(2,44) = 0.567$, $p = .57$) and PHQ-9 reliable change and GAD-7 reliable change, respectively.

---

[4] Negative correlations between reliable change and OSCE scores represent a reduction in symptoms as OSCE scores increase, whilst a negative reliable deterioration score indicates a reduction in patients deteriorating as OSCE scores increase.

Table 2. Correlations (Spearman's rho) between OSCE scores and patient change variables across time frames

During (n = 39)	OSCE 1	OSCE 2	OSCE 3	PHQ-9 change[1]	GAD-7 change[1]	Reliable improvement[2]	Reliable deterioration[2]
OSCE 1	1	.341*	.353*	−.220	−.341*	.405*	−.290
OSCE 2		1	.514**	.301	.295	-.002	.205
OSCE 3			1	.018	−.095	.243	−.270
PHQ-9 change				1	.826**	−.687**	.486**
GAD-7 change					1	−.695**	.538**
Reliable improvement						1	−.474**
Reliable deterioration							1

Post (n = 38)	OSCE 1	OSCE 2	OSCE 3	PHQ-9 change[1]	GAD-7 change[1]	Reliable improvement[2]	Reliable deterioration[2]
OSCE 1	1	.123	−.292	.276	.309	−.218	.323
OSCE 2		1	.078	−.104	−.134	.144	−.222
OSCE 3			1	−.021	−.243	−.052	−.227
PHQ-9 change				1	.632**	−.782**	.516**
GAD-7 change					1	−.711**	.346*
Reliable improvement						1	−.449**
Reliable deterioration							1

All time points (n = 46)	OSCE 1	OSCE 2	OSCE 3	PHQ-9 change[1]	GAD-7 change[1]	Reliable improvement[2]	Reliable deterioration[2]
OSCE 1	1	.334*	.307*	−.098	−.180	.169	−.112
OSCE 2		1	.350*	.019	.018	.069	−.015
OSCE 3			1	−.147	−.358*	.256	−.441**
PHQ-9 change				1	.725**	−.672**	.546**
GAD-7 change					1	−.737**	.516**
Reliable improvement						1	−.427**
Reliable deterioration							1

*$p < 0.01$ (2-tailed); **$p < 0.001$ (2-tailed); PWPs treating less than 5 patients were excluded.
[1]Mean reliable PHQ-9/GAD-7 change scores calculated by applying Jacobson and Truax's (1991) reliable change formula to the difference between patients first and last PHQ-9/GAD-7 scores.
[2]Proportion of patients showing a reliable improvement/deterioration on the PHQ-9 and/or GAD-7, with no deterioration on the other measure.

It has been suggested that there may be a threshold of competence beyond which superior outcomes are reached (McManus et al., 2010). To explore this, OSCE performance was divided into competence quartiles, and classified as 'high' (>75th percentile), 'middle' (between 25th and 75th percentiles) and 'low' (<25th percentile) (see Table 3). Chi-squared analyses were conducted and adjusted standardized residuals were explored to explain significant results.

Table 3. Distribution of OSCE scores by performance group, sample sizes and χ^2 statistics

	OSCE 1			OSCE 2			OSCE 3		
	Low	Middle	High	Low	Middle	High	Low	Middle	High
Mean OSCE score (%)	58.0	74.7	95.0	54.6	73.1	94.8	65.1	80.6	97.8
Minimum/ maximum (%)	52–63	64–90	93–98	50–59	60–68	90–100	59–71	72–96	97–100
PWP n	10	32	5	11	32	4	10	32	5
Patient n	406	2659	623	502	2437	749	634	2273	781
Reliable change (PHQ-9)	$\chi^2(4) = 2.73, p = .60$			$\chi^2(4) = 4.48, p = .35$			$\chi^2(4) = 14.60, p = .006$		
Reliable change (GAD-7)	$\chi^2(4) = 14.29, p = .006$			$\chi^2(4) = 9.27, p = .06$			$\chi^2(4) = 18.31, p = .001$		

Significant differences were found between observed (f_o) and expected (f_e) frequencies across groups of OSCE performance.

OSCE 1: Significantly more patients experienced a reliable deterioration in their symptoms of anxiety if treated by PWPs in the low competence group ($f_o = 33, f_e = 21.6, Z = 2.7$) whilst significantly fewer than would be expected by chance deteriorated if treated by the middle competence group ($f_o = 124, f_e = 141.3, Z = -2.8$).

OSCE 2: Differences in reliable change in patients' symptoms of depression by OSCE competence quartiles did not approach significance and no standardized residuals reached significance. However, differences in reliable change in patients' symptoms of anxiety by OSCE competence quartiles approached significance; this was explained by a significantly greater number of patients experiencing a reliable deterioration in their symptoms of anxiety if treated by a PWP in the low competence group ($f_o = 38, f_e = 26.7, Z = 2.4$).

OSCE 3: When treated by the low competence group, significantly fewer patients than would be expected demonstrated a reliable improvement in their symptoms of depression ($f_o = 360, f_e = 396.1, Z = -3.3$), and more patients experienced a reliable deterioration ($f_o = 62, f_e = 48, Z = 2.3$). Similarly, when treated by the low competence group, fewer patients than expected reliably improved ($f_o = 321, f_e = 346.9, Z = -2.3$) and more reliably deteriorated ($f_o = 51, f_e = 33.7, Z = 3.4$) in their symptoms of anxiety. Conversely, significantly more patients treated by the high performing group showed a reliable improvement in their symptoms of depression ($f_o = 515, f_e = 487.9, Z = 2.3$) and fewer than expected showed a reliable deterioration in their symptoms of anxiety ($f_o = 26, f_e = 41.5, Z = -2.8$).

Discussion

This is the first study to explore the relationship between low-intensity practitioner competence and patient outcome in IAPT services. More than two-thirds of patients treated by PWPs during the study period experienced a reliable improvement in their symptoms of depression and

anxiety. These findings are consistent with national outcomes in IAPT; approximately two-thirds of referrals that completed a course of treatment demonstrated significant improvement in their condition (NHS Digital, 2017). Furthermore, partial support was found for the hypotheses; patients treated by qualified PWPs were more likely to experience reliable improvement in their symptoms, and less likely to experience a deterioration than those treated by trainees. However, in relation to practitioner competence, there was little association between OSCE scores and patient outcome, with a relationship only becoming evident at the extremes. More patients of the elite few in the high competence groups reliably improved and fewer deteriorated than statistically expected, whilst the pattern was reversed in the low competence group. Meanwhile the majority of PWPs in the middle groups achieved outcomes that were statistically as expected.

An encouraging finding of the current research is the effect of training with the outcomes of patients treated by qualified PWPs (post-training) being superior to those of patients treated by trainees (during training). Whilst preliminary, this finding supports a growing body of evidence for the effectiveness of therapist training, including brief training (e.g. Westbrook et al., 2008) and that conducted via the internet (e.g. Kobak et al., 2017; Rakovshik et al., 2013; Sholomskas et al., 2005).

The non-linear, complex relationship between practitioner competence and patient outcome is consistent with findings from high=intensity interventions (e.g. Webb et al., 2010). Several studies have observed a similar pattern of results. Whilst most therapists have overlapping outcomes, outliers can be found at both tails; at one pole therapists of low competence experience much higher rates of deterioration than others, whilst at the other, 'elite' therapists experience a more consistent pattern of improvement (Green et al., 2014; Okiishi et al., 2003; Saxon and Barkham, 2012). It has been suggested that the relationship may be more evident in some contexts but not others (e.g. only with certain therapeutic modalities) (Webb et al., 2010). This study indicates that such a relationship exists within the lower intensity interventions. It may also be the case that for some patients, practitioner competence is less important than for others and it is the interplay between practitioner and patient characteristics that moderates patient outcome.

A major strength of the current study is the large patient sample, spanning the first three years of IAPT. However, a number of limitations must be highlighted. Whilst the use of OSCE has been supported and is widespread in medical training (Fairburn and Cooper, 2011), questions remain over the reliability of the measures. Few trainees scored below the threshold of competence (50) on OSCEs 1–3, and this restriction of range may have made it more challenging to detect the impact of competence. Furthermore, whilst module-specific OSCEs allow for targeted examination of a specific set of competences, the necessary differences in items incorporated in each rating scale prevents evaluation of change in competence over time, and prohibits the pooling of data across multiple levels of measurement to glean a more reliable estimate of competence (see Keen and Freeston, 2008). It should also be noted that subsequent to the study period, the revised curriculum for PWPs (UCL, 2015) has reduced the frequency of OSCE from three to one, in favour of scrutinizing clinical practice. Future research exploring the inter-relationships between training, competence and patient outcome incorporating clinical practice as a measure of competence would be a valuable addition to the literature.

The naturalistic design enables exploration of data from routine clinical practice and may therefore offer a representative assessment of the effectiveness of treatment. Nevertheless, the uncontrolled retrospective nature of the study is a limitation. Patient data were drawn from archives and therefore it is not possible to ascertain whether there were any systematic differences in the allocation process, how accurately data were entered, whether missing data

were missing at random, or if certain PWPs failed to input session-by-session data routinely or strategically. Furthermore, whilst IAPT has been rolled out across England, it remains to be seen whether this framework is generalizable more broadly. Importantly, it must also be acknowledged that the current study does not account for the fact that patient outcomes may be affected by factors other than competence (Trepka et al., 2004; Webb et al., 2010). Patient demography and case history for example may moderate the relationship between competence and outcome, although these data were not available via the IAPT computer system, and self-report patient symptomology (PHQ-9 and GAD-7) was used as a proxy for diagnosis and complexity.

In conclusion, this study indicates that training improves PWP competence, and further that patient outcomes are affected by practitioner competence using low-intensity interventions in a way similar to that found for more traditional, higher intensity interventions (Branson et al., 2015). Further research with a prospective, experimental or randomized controlled design would further inform the nature of the relationship between competence and patient outcome for CBT interventions delivered across the full range of intensities.

References

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders IV*. Washington DC.

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

Bennett-Levy, J. (2010). *Oxford Guide to Low Intensity CBT Interventions*. Oxford: Oxford University Press.

Branson, A., Shafran, R. and Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy*, *68*, 19–26. doi: https://doi.org/10.1016/j.brat.2015.03.002

Cuijpers, P. and Schuurmans, J. (2007). Self-help interventions for anxiety disorders: an overview. *Current Psychiatry Reports*, *9*, 284–290.

Department of Health (2010). *Realising the Benefits: IAPT at Full Roll Out*. London: Department of Health. Available at: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_112982 (accessed 2 August 2017).

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

Ferriter, M., Kaltenthaler, E., Parry, G. and Beverley, C. (2008). Computerised CBT: a review. *Mental Health Today*, *193*, 30–31.

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

Green, H., Barkham, M., Kellett, S. and Saxon, D. (2014). Therapist effects and IAPT Psychological Well-being Practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy*, *63*, 43–54. doi: https://doi.org/10.1016/j.brat.2014.08.009

Gyani, A., Shafran, R., Layard, R. and Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy*, *51*, 597–606. doi: https://doi.org/10.1016/j.brat.2013.06.004

IAPT (2011). *The IAPT Data Handbook: Guidance on recording and monitoring outcomes to support local evidence-based practice*. London. Available at: http://webarchive.nationalarchives.gov.uk/20160302160058/http://www.iapt.nhs.uk/silo/files/iapt-data-handbook-v2.pdf

Jacobson, N. S. and Truax, P. A. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

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

Kobak, K. A., Wolitzky-Taylor, K., Craske, M. G. and Rose, R. D. (2017). Therapist training on cognitive behavior therapy for anxiety disorders using internet-based technologies. *Cognitive Therapy and Research*, *41*, 252–265. doi: 10.1007/s10608-016-9819-4

Kroenke, K., Spitzer, R. L. and Williams, J. B. W. (2001). The PHQ-9. Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613.

Manring, J., Beitman, B. D. and Dewan, M. J. (2003). Evaluating competence in psychotherapy. *Academic Psychiatry*, *27*, 136–144. doi: 10.1176/appi.ap.27.3.136

McManus, F., Westbrook, D., Vazquez-Montes, M., Fennell, M. and Kennerley, H. (2010). An evaluation of the effectiveness of diploma-level training in cognitive behaviour therapy. *Behaviour Research and Therapy*, *48*, 1123–1132.

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

NHS Digital (2017). Improving access to psychological therapies (IAPT). Available at: http://content.digital.nhs.uk/catalogue/PUB23186/IAPT-month-oct-2016-exec-sum.pdf (accessed 2 August 2017).

NICE (2011). *Common Mental Health Disorders: Identification and Pathways to Care. Clinical Guideline 123*. London: The British Psychological Society and The Royal College of Psychiatrists. Available at: https://www.nice.org.uk/guidance/cg123

Okiishi, J. C., Lambert, M. J., Nielsen, S. and Olgles, B. (2003). Waiting for Supershrink: an empirical analysis of therapist effects. *Clinical Psychology and Psychotherapy*, *10*, 361–373.

Rakovshik, S. G., McManus, F., Westbrook, D. and Kholmogorova, A. B. (2013). Randomized trial comparing Internet-based training in cognitive behavioural therapy theory, assessment and

formulation to delayed-training control. *Behaviour Research and Therapy*, *51*, 231–239. doi: 10.1016/j.brat.2013.01.009

Richards, D. A. and Whyte, M. (2009). *ReachOut. National Programme Educator Materials to Support the Delivery of Training for Psychological Wellbeing Practitioners Delivering Low Intensity Interventions (2nd edn)*. London: Rethink.

Roth, A. D. and Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, *36*, 129–147. doi: 10.1017/s1352465808004141

Saxon, D. and Barkham, M. (2012). Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, *80*, 535–546.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F. and 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. doi: 10.1037/0022-006x.73.1.106

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

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Spitzer, R. L., Kroenke, K., Williams, J. B. and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097.

Trepka, C., Rees, A., Shapiro, D. A., Hardy, G. E. and Barkham, M. (2004). Therapist competence and outcome of cognitive therapy for depression. *Cognitive Therapy and Research*, *28*, 143–157.

Turpin, G. (2008). *Improving Access to Psychological Therapies (IAPT)*. Paper presented at the DCP Managers' Conference, Doncaster.

UCL (2015). National Curriculum for the Education of Psychological Well-being Practitioners, 3rd edn. Available at: https://www.ucl.ac.uk/pals/research/cehp/research-groups/core/pwp-review/docs/PWPREVIE_-curriculum

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

Westbrook, D., Sedgwick-Taylor, A., Bennett-Levy, J., Butler, G. and 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. doi: 10.1017/S1352465808004608

Williams, C. H. J. (2015). Improving Access to Psychological Therapies (IAPT) and treatment outcomes: epistemological assumptions and controversies. *Journal of Psychiatric and Mental Health Nursing*, *22*, 344–351. doi: 10.1111/jpm.12181

Ye, F. (2010). Change scores. In N. J. Salkind (ed), *Encyclopedia of Research Design*. Thousand Oaks, CA: SAGE Reference Online.