CentAUR
Central Archive at the University of Reading
Reading's research outputs online
The Relationship between Competence and Patient Outcome with Low-Intensity Cognitive Behavioural Interventions

Running Head: PWP competence and Patient Outcome

Amanda Branson** PhD. Pamela Myles, MSca, Mishka Mahdi, MSc and Roz Shafranb PhD.

School of Psychology and CLS, University of Reading, Berkshire, United Kingdom.

*Corresponding Author

a. Address: School of Psychology and CLS, University of Reading, Earley Gate, Reading, Berkshire, RG14 6AL. Tel: +44 118 378 7531. email address: a.branson@reading.ac.uk

b. Address: University College London Great Ormond Street Institute of Child Health, Gower Street, London, WC1E 6BT, United Kingdom. Email address: r.shafran@ucl.ac.uk

Keywords: IAPT, training, depression, anxiety, low-intensity, PWP, psychological wellbeing practitioner
Abstract

**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 British Government’s Improving Access to Psychological Therapies (IAPT) initiative.

**Methods:** Forty-seven PWPs treating 3688 patients participated. Relationships between PWP scores on three Observed Standardised Clinical Examinations (OSCEs) and reliable change in patients’ symptoms of anxiety and depression were explored at two time points: during the yearlong training phase, and over a twelve 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 to those treated by the least 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.


**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 in the form of guided self-help (Cuijpers & Schurmans, 2007) and Computerised Cognitive Behavioural Therapy (CCBT: see Ferriter, Kaltenthaler, Parry, & Beverley, 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 Cognitive Behavioural Therapy (CBT: Turpin, 2008), and are considered less burdensome for patients (Bennett-Levy, 2010).

In order to meet demand, 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 Wellbeing 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, 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 relevant, but 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 & Pilling, 2008) whilst maintaining a high patient caseload. The clinical competence of PWPs during the study period was measured via Observed Standardised 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, Beitman, & Dewan, 2003) and it is only routinely used within the IAPT framework. Whilst 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 & Cooper, 2011; Ginzburg, Bohn, Höfling, & Weck, 2012; Muse & 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, Mathieson, & Beaumont, 2007; Keen & Freeston, 2008; McManus, Westbrook, Vazquez-Montes, Fennell, & Kennerley, 2010), and of a positive effect of therapist competence on patient outcomes (e.g. Green, Barkham, Kellett, & Saxon, 2014; Okiishi, Lambert, Nielsen, & Olgles, 2003; Saxon & Barkham, 2012), however relationships appear to be complex (Branson, Shafran, & Myles, 2015; Muse & 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 is 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\(^1\) comprising four modules which 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 trainees have the core knowledge, skills and attitudes necessary to competently assess and treat patients. Trainees were required to pass module one and two assessments prior to being deemed competent to treat patients. Modules three and four were

---
\(^1\) 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.
shorter in duration and were merged into one academic term at the University of Reading. Following successful completion of modules one and two trainees are 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.

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, Shafran, Layard, & Clark, 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)\(^3\).

---

\(^2\) Trainees typically manage 80% of a full caseload by month 10 of training.

\(^3\) Due to the timing of data collection, follow-up patient data (post) were not available for the last two cohorts of trainees (n=22)
Practitioners were PWPs trained on the University of Reading’s Postgraduate Certificate in Evidence Based Psychological Therapies, between academic years 2008/09 and 2011/12. 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 (SD=7.69), 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, 4 (8.7%) studied another Social Sciences, and 5 (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), 6 (12.8%) a Lower Second (2:2) and 1 trainee received a Third Class Pass (2.1%).

Chi-Square test 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, Spitzer, & Williams, 2001) and the Generalized-Anxiety-Disorder questionnaire (GAD-7: Spitzer, Kroenke, Williams, & Löwe, 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 seven 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 Reach Out, the national curriculum for PWP (Richards & Whyte, 2009). OSCEs took place on completion of each taught module, roughly three, six and nine months into training respectively. Reach-Out 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, prioritisation 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 OSCE1 (25.5%), OSCE2 (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 scrutinised 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 OSCE1 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 OSCE1 scores (M=82.38, SD=10.36) than did service 5 (M=60.8, SD=5.76). Scores on OSCEs2 and 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 OSCEs2 and 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 & Fleiss, 1979) argue for the use of a 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 OSCE3 (ICC=.84, p=.004), followed by the OSCE2 (ICC=.71, p=.019); reliability of OSCE1 (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 OSCE1-3 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, however 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: M=11.81, threshold=9, SD=2.70; GAD-7: M=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 one 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 & Freeston, 2008; McManus, et al., 2010) and patient outcome (e.g. Westbrook, Sedgwick-Taylor, Bennett-Levy, Butler, & McManus, 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) \(M=0.68\) than it was for patients treated by trainees (during training) \(M=0.72\), \(U=1597500.00, p=.004, r=.05\), whilst the proportion of patients who reliably deteriorated was greater for patients treated by trainees (during training) \(M=0.10\) than it was for patients treated by qualified PWPs (post-training) \(M=0.06\), \(U=1621783.50, p=.002, r=.05\).

**Effect of competence on patient outcome**

Spearman Rho correlations between OSCEs1-3, and reliable change in patients symptoms (see Table 2) were inconsistent, and predominantly in the modest range. A small number of associations reached significance: OSCE1 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 ($r=.405$, $p=.01$). No associations reached significance for OSCE2 or 3. When all patients were included (All times), significant negative associations emerged only between OSCE3 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 OSCE1 ($R^2=0.012$, $F(2,44)=0.27$, $p=.85$; $R^2=0.018$, $F(2,44)=0.393$, $p=.68$), OSCE2 ($R^2=1.3$, $F(2,44)=3.149$, $p=.05$; $R^2=0.085$, $F(2,44)=2.054$, $p=.14$) and OSCE3 ($R^2=0.067$, $F(2,44)=1.586$, $p=.22$; $R^2=0.026$, $F(2,44)=0.567$, $p=.57$) and PHQ-9 reliable change and GAD-7 reliable change respectively.

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

---

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.
**OSCE1:** 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 \( (f_o=124, f_e=141.3, Z=-2.8) \).

**OSCE2:** Differences in reliable change in patients’ symptoms of depression by OSCE competence quartiles did not approach significance and no standardised 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 \( (f_o=38, f_e=26.7, Z=2.4) \).

**OSCE3:** 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 PWPss 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 PWPss 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 PWPss (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, Wolitzky-Taylor, Craske, & Rose, 2017; Rakovshik, McManus, Westbrook, & Kholmogorova, 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, DeRubeis, & Barber, 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 & 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 & Cooper, 2011), questions remain over the reliability of the measures. Few trainees scored below the threshold of competence (50) on OSCEs 1-3, 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 & 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 scrutinising 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, Rees, Shapiro, Hardy, & Barkham, 2004; Webb, et al., 2010). Patient demography and case history for example may moderate the relationship between competence and outcome, however 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, 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


Roth, A. D., & Pilling, S. (2008). Using an Evidence-Based Methodology to Identify the Competences Required to Deliver Effective Cognitive and Behavioural Therapy for
Depression and Anxiety Disorders. *Behavioural and Cognitive Psychotherapy, 36*(2). doi: 10.1017/s1352465808004141


References


**References**


**References**


theory, assessment and formulation to delayed-training control. *Behaviour Research and Therapy, 51*(6), 231-239. doi: 10.1016/j.brat.2013.01.009


References


methods analysis. *Behaviour Research and Therapy*, 63, 43-54. doi: https://doi.org/10.1016/j.brat.2014.08.009


References


References


Funding
This study was conducted with the support of a grant awarded by the British Association for Behavioural and Cognitive Psychotherapies (BABCP). All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Acknowledgements
We thank the participating IAPT PWPs and South Central Strategic Health Authority IAPT service sites for their support.

Ethics
We have abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the APA http://www.apa.org/ethics/code/. Ethics approval was granted for this project by NHS National Research Ethics Committee, reference number: 09/H0505/128.

Conflict of interests: The authors have no conflict of interest with respect to this publication.
References


Roth, A. D., & Pilling, S. (2008). Using an Evidence-Based Methodology to Identify the Competences Required to Deliver Effective Cognitive and Behavioural Therapy for
Depression and Anxiety Disorders. *Behavioural and Cognitive Psychotherapy*, 36(2). doi: 10.1017/s1352465808004141


<table>
<thead>
<tr>
<th>total patients</th>
<th>Reliable improvement</th>
<th>Reliable deterioration</th>
<th>Combined reliable improvement$^1$</th>
<th>Combined reliable deterioration$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PHQ-9</td>
<td>GAD-7</td>
<td>PHQ-9 &amp;/or GAD-7</td>
<td>PHQ-9 &amp;/or GAD-7</td>
</tr>
<tr>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>During training</td>
<td>2009</td>
<td>1206 60.0%</td>
<td>1051 52.3%</td>
<td>166 8.3%</td>
</tr>
<tr>
<td>Post training</td>
<td>1679</td>
<td>1098 65.4%</td>
<td>967 57.6%</td>
<td>113 6.7%</td>
</tr>
<tr>
<td>All time points</td>
<td>3688</td>
<td>2304 62.5%</td>
<td>2018 54.7%</td>
<td>279 7.6%</td>
</tr>
</tbody>
</table>

$^1$ 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.

Table 1 Frequency and percentage of patients reliably improving and reliably deteriorating in their symptoms of depression (PHQ-9), anxiety (GAD-7) by timeframe
<table>
<thead>
<tr>
<th></th>
<th>DURING</th>
<th>OSCE1</th>
<th>OSCE2</th>
<th>OSCE3</th>
<th>PHQ-9 change¹</th>
<th>GAD-7 change¹</th>
<th>Reliable improv²</th>
<th>Reliable deter²</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSCE1</td>
<td>1</td>
<td>.341 *</td>
<td>.353 *</td>
<td>-220</td>
<td>-3.41 *</td>
<td>-290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSCE2</td>
<td>1</td>
<td>.514 **</td>
<td>.301</td>
<td>.295</td>
<td>.002</td>
<td>.205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSCE3</td>
<td>1</td>
<td>.018</td>
<td>-0.95</td>
<td>.243</td>
<td>-270</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9 change</td>
<td>1</td>
<td>.826 **</td>
<td>-6.87 **</td>
<td>486 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD-7 change</td>
<td>1</td>
<td>-6.95 **</td>
<td>.538 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliable improv’</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliable deter’</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>POST</th>
<th>OSCE1</th>
<th>OSCE2</th>
<th>OSCE3</th>
<th>PHQ-9 change¹</th>
<th>GAD-7 change¹</th>
<th>Reliable improv²</th>
<th>Reliable deter²</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSCE1</td>
<td>1</td>
<td>.123</td>
<td>-.292</td>
<td>.276</td>
<td>.309</td>
<td>-.218</td>
<td>.323</td>
<td></td>
</tr>
<tr>
<td>OSCE2</td>
<td>1</td>
<td>.078</td>
<td>-.104</td>
<td>-.134</td>
<td>.144</td>
<td>-.222</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSCE3</td>
<td>1</td>
<td>-.021</td>
<td>-.243</td>
<td>-.052</td>
<td>-.227</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9 change</td>
<td>1</td>
<td>.632 **</td>
<td>-7.82 **</td>
<td>.516 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD-7 change</td>
<td>1</td>
<td>-.711 **</td>
<td>.346 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliable improv’</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.449 **</td>
</tr>
<tr>
<td>Reliable deter’</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ALL TIME POINTS</th>
<th>OSCE1</th>
<th>OSCE2</th>
<th>OSCE3</th>
<th>PHQ-9 change¹</th>
<th>GAD-7 change¹</th>
<th>Reliable improv²</th>
<th>Reliable deter²</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSCE1</td>
<td>1</td>
<td>.334 *</td>
<td>.307 *</td>
<td>-.098</td>
<td>-.180</td>
<td>.169</td>
<td>-.112</td>
<td></td>
</tr>
<tr>
<td>OSCE2</td>
<td>1</td>
<td>.350 *</td>
<td>.019</td>
<td>.018</td>
<td>.069</td>
<td>.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSCE3</td>
<td>1</td>
<td>-.147</td>
<td>-.358 *</td>
<td>.256</td>
<td>-.441 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9 change</td>
<td>1</td>
<td>.725 **</td>
<td>-.672 **</td>
<td>.546 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD-7 change</td>
<td>1</td>
<td>-.737 **</td>
<td>.516 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliable improv’</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.427 **</td>
</tr>
<tr>
<td>Reliable deter’</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p<0.01 (2tailed)  ** p<0.001 (2tailed)

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.

**Table 2** Correlations (Spearman’s Rho) between OSCE scores and patient change variables across timeframes
<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th></th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th></th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean OSCE score (%)</td>
<td>58.0</td>
<td>74.7</td>
<td>95.0</td>
<td>Mean OSCE score (%)</td>
<td>54.6</td>
<td>73.1</td>
<td>94.8</td>
<td>Mean OSCE score (%)</td>
<td>65.1</td>
<td>80.6</td>
<td>97.8</td>
<td></td>
</tr>
<tr>
<td>Minimum/Maximum (%)</td>
<td>52-63</td>
<td>64-90</td>
<td>93-98</td>
<td>Minimum/Maximum (%)</td>
<td>50-59</td>
<td>60-8</td>
<td>90-100</td>
<td>Minimum/Maximum (%)</td>
<td>59-71</td>
<td>72-96</td>
<td>97-100</td>
<td></td>
</tr>
<tr>
<td>PWP n</td>
<td>10</td>
<td>32</td>
<td>5</td>
<td>PWP n</td>
<td>11</td>
<td>32</td>
<td>4</td>
<td>PWP n</td>
<td>10</td>
<td>32</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>patient n</td>
<td>406</td>
<td>2659</td>
<td>623</td>
<td>patient n</td>
<td>502</td>
<td>2437</td>
<td>749</td>
<td>patient n</td>
<td>634</td>
<td>2273</td>
<td>781</td>
<td></td>
</tr>
<tr>
<td>Reliable change (PHQ-9)</td>
<td>$\chi^2 (4)=2.73, \ p=.60$</td>
<td>$\chi^2 (4)=4.48, \ p=.35$</td>
<td>$\chi^2 (4)=14.60, \ p=.006$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliable change (GAD-7)</td>
<td>$\chi^2 (4)=14.29, \ p=.006$</td>
<td>$\chi^2 (4)=9.27, \ p=.06$</td>
<td>$\chi^2 (4)=18.31, \ p=.001$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3** Distribution of OSCE scores by performance group, sample sizes and $\chi^2$ statistics
Figures

**Figure 1** Graph showing mean percentage of patients reliably improving and deteriorating by timeframe, with upper and lower bound confidence intervals.