Why Are Some Therapists More Effective Than Others? Exploring the Role of Training, Self-reflection and Interpersonal Skills

Ryan Lawty

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Clinical Psychology

The University of Sheffield
Clinical Psychology Unit
Department of Psychology

June 2020

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Word Count

Section One: Literature Review

Excluding references and tables: 7953

Including references and tables: 11864

Section Two: Empirical Study

Excluding references and tables: 6808

Including references and tables: 8829

Total word count

Excluding references and tables: 14473

Including references and tables: 20670

Overall Summary

There are differences in the effectiveness of therapists reflected in the outcomes their patients achieve; this variability can be substantial and is called the 'therapist effect'. For example, the most effective therapists have twice the recovery rate of the least effective therapists. How and why some therapists are more effective than others, is therefore important to understand. To explore this further this thesis has two sections. The first section systematically reviews the literature to assess whether training is effective in improving the competence of therapists. The second section reports an empirical study exploring whether therapist interpersonal skills and self-reflection are associated with the outcomes their patients achieve.

Section One: Previous reviews have been conducted about the effectiveness of training. The most recent of these was published in 2013, therefore, the focus of the current review was on studies published from 2013 to 2020. Twelve new studies were identified. The findings were indicative that training does increase competence of therapists and this seemed to be especially the case when training was of longer duration and involved components of practice with real world clients. However, the general quality of the research was low, so conclusions are made cautiously. It is recommended that more high-quality research is required to better understand the effects of training interventions on therapists' competence.

Section Two: Therapists (N=61) completed surveys about their self-reflective abilities, social skills and interpersonal problems. Supervisors (N=19) of these therapists were also asked to rate the therapists' self-reflective abilities. Patient depression and anxiety outcomes for these therapists were collected from electronic clinical records (anonymously). Patient outcome data (N=3112) was available for N=42 therapists; corresponding supervisor reported data was available for N=18 therapists. Statistical analysis using multi-level modelling indicated that

therapists' higher self-reported social skills were associated with better treatment outcomes for depression, but not for anxiety. Therapist-reported self-reflection and interpersonal problems were not related to treatment outcomes. A sensitivity analysis suggested that important outlying therapists' responses might undermine relationships between therapist-reported self-reflection and treatment outcomes. There was insufficient supervisor data to make meaningful conclusions. These findings contribute to a growing literature concerning why therapist effects occur and what steps may be taken to increase the effectiveness of therapists. Future research should continue to strive for understanding about the characteristics/ abilities that make therapists effective and what can be done to instil or enhance effectiveness.

Acknowledgements

I would like to say a huge thank you to both of my supervisors, Dr Jaime Delgadillo and Dr Stephen Kellett for their guidance and support throughout this process. I would particularly like to acknowledge them for finding the time in their busy schedules to respond to questions, offer advice and read drafts so responsively which has been very containing at times of stress. Additionally, I would like to thank Dave Saxon, for his time, patience and knowledge in helping me to learn and understand multilevel modelling analysis.

I would also like to acknowledge the therapists and supervisors who took time to participate in this research and thank the NHS data managers, Amy Southgate, Laura Parkhouse, Rebecca Parr and Naida Khan, for working so hard to share patient outcome data as quickly as they could.

Finally, to my friends and family, thank you for all the support, love and patience you have given me. Especially thank you to my partner, Em, for your relentless encouragement, love and belief in me; you have been a rock and I could not have done this without you.

Table of Contents

Section One: Systematic Literature Review

Does Training Improve the Clinical Skills of Trainee and Qualified Therapists? A Systematic Literature Review.

Abstract

Objectives: Previous reviews have suggested that therapist training is effective at improving skills, knowledge and competence. However, the reviews highlighted methodological limitations in the literature and therefore conclusions are made cautiously. This review therefore aimed to revisit the question of the effectiveness of training therapists. The primary aim was to assess the evidence for effectiveness of training and also to establish whether methodological problems highlighted in previous reviews have been addressed. **Method:** A systematic literature review was conducted utilising Scopus, Medline and PsychINFO databases with subsequent forward and backward searches and searches of the grey literature. Various search terms were used relating to development of therapist competence during training. Methodological appraisal of the identified studies was conducted, and the findings synthesised in a narrative review. **Results:** Twelve studies were identified for inclusion. Generally, findings indicated that training is effective at increasing skills, knowledge and competence of trainees. This was most clear for studies in which interventions were assessed over longer periods and had components of practice with real patients. However, the studies reviewed generally were of poor methodological quality due to lack of comparison groups, poor sample sizes and the limited focus on high level assessment of global competencies. **Conclusions:** The findings again suggest that training is effective at increasing the competence of trainees. Recommendations for improving study designs do not appear to have been heeded. Therefore, there remains significant limitations in the literature, highlighting an urgent need for more high-quality research to understand the impact of training on trainee competence.

Keywords: “training” “therapist” “psychotherapist” “cognitive behaviour therapist” **“**counsellor” “competence” “ability” “skill”

Practitioner Points

- There is some evidence that training programmes do increase the competence of trainee and qualified therapists.
- The methodology in the area is generally weak, therefore, conclusions are made cautiously.
- More high quality and routine research should be conducted into the effectiveness of training interventions for therapists.

Introduction

Psychotherapy is an effective and recommended treatment for a wide range of mental health difficulties (British Psychological Society & Royal College of Psychiatrists, 2011). There is a recognised demand for psychological therapies reflected in the development of the Improving Access to Psychological Therapies (IAPT) initiative launched in 2008 in the UK (National Collaborating Centre for Mental Health, 2019). This is reflected internationally with similar implementation programmes following this model in countries such as Canada, Australia, Norway and Sweden (Clark, 2019). This means there is a huge demand for training of psychological therapists and for associated training programmes to ensure the competency and safety of the training therapists. Training courses for therapists are both costly and time intensive (Fairburn & Cooper, 2011), so it is imperative to index programme effectiveness and to also understand how to improve programmes to train therapists to deliver these technically and interpersonally complex interventions. Evidence of the effectiveness of the training of psychotherapists is therefore needed to reassure educational and service commissioners and the patients receiving the interventions. This review centres on evaluating the evidence that the training process develops the competency of the trainee psychotherapists.

Defining Therapist Competency

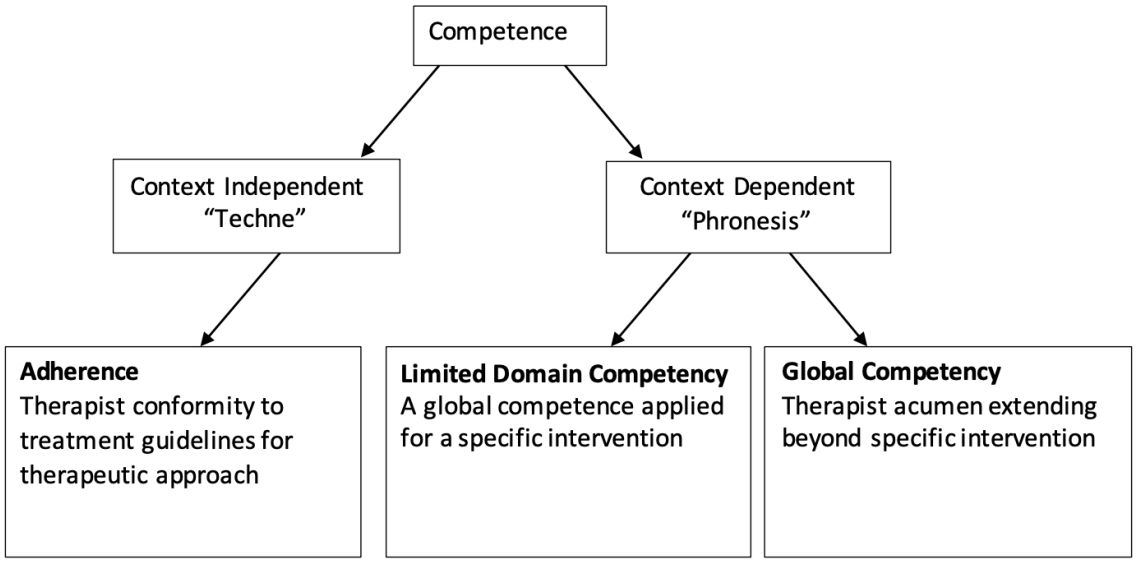
Achieving competency is central to therapist training. Indeed, ensuring that therapists are competent is seen as an ethical imperative (American Psychological Association, 2012; British Psychological Society, 2018). There has been a shift to competency-based frameworks for training (Kaslow, 2004), with many professional trainings now setting competency benchmarks, which trainees are required to meet (British Psychological Society, 2019; Roth & Pilling, 2007).

However, the bar of what constitutes competence is ambiguous and it is questionable where the line of competence sits on the spectrum between dangerously incompetent to highly skilled (Sharpless & Barber, 2009). Additionally, having a level at which competence is attained is suggestive that competence has an end point, whereas competence might be better thought of as an ongoing process within an individual over a lifetime of therapeutic learning (Nagy, 2005).

This may reflect the difficulty in defining and measuring therapist competence. A commonly used and broad definition of competency is “the extent to which a therapist has the knowledge and skill required to deliver a treatment to the standard needed for it to achieve its expected effects” (Fairburn & Cooper, 2011). To add nuance, competence can be separated into context dependent versus independent, and within context dependent this can be further separated into limited domain competencies and global competencies (Barber et al., 2007). Context independent competence refers to the knowledge of treatment techniques and application of these (e.g. adherence to a treatment model across sessions), limited domain competence refers to more global skills but still related to application of skills within treatment modalities, whereas broader global competencies reflect the skills that transcend specific therapies (see Figure 1) . Roth and Pilling (2007) termed these meta-competencies and particularly emphasised how meta competencies should influence therapist decision making. Sharpless and Barber (2009) extend this distinction and defined competence in terms drawn from Aristotle, (1984) ways of knowing and intellectual virtues. They use the term “techne” to refer to the “what” of knowledge, this includes the technical or declarative knowledge of therapeutic models (e.g. knowing what graded exposure is) and is knowledge; which is rule governed, has an attainable

Figure 1

An Illustration of the Different Types of Competency



end point and is independent of the context. They borrow the term "phronesis" to reflect competence which is driven by values which regulate interactions allowing flexibility and adaptability which is therefore a more context dependent application of skills (e.g. knowing when to be assertive in terms of ensuring the conditions of effective exposure). As opposed to being competence of "what", they suggest "phronesis" is the competence of "when" and "where".

Models of Competency Development

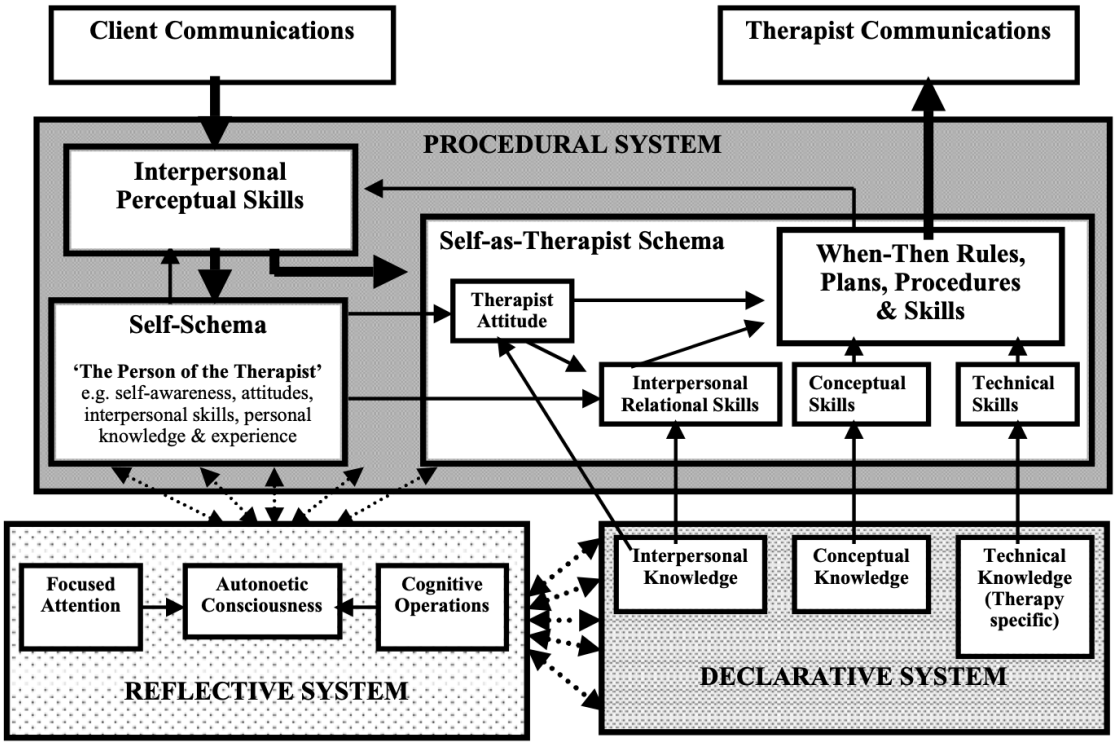
These different types of competence can be fitted to theories of competence development. For example, Sharpless and Barber (2009) apply Dreyfus & Dreyfus', (1986) theory of competence development. This model contains five stages, at stages one and two (Beginner and Advanced Beginner) learning of specific techniques and knowledge is emphasised, for example rule based learning (e.g. "if the client is [depressed], then I should deliver [behavioural activation"]). Reflecting the "techne" or technique/ knowledge specific competencies. Whereas stages three, four and five (Competence, Proficiency

and Expertise), increasingly emphasise the ability to respond independently, flexibly and intuitively; these stages capture the “phronesis” (e.g. “this client is [depressed], and I should deliver [behavioural activation] but I need to acknowledge the role that pain plays and ensure sufficient pacing”). This does not mean the absence of rule-based adherence, rather the skilful judgement and application of these skills in a meta-competent manner. In this framework, it could be argued that knowledge and adherence to a model is required for level 1 and 2, domain limited competency required for level 3 and global competency required to reach level 4 and 5.

This understanding of competency also fits well with a model of competency development proposed by Bennett-Levy (2006), containing interactive aspects of declarative, procedural and reflective (DPR) abilities (see Figure 2). The declarative system hosts knowledge related to interpersonal knowledge (e.g. importance of reflecting back/ summarising), conceptual knowledge (e.g. knowledge of a model theory) and technical knowledge (e.g. thought challenging technique). However, this alone can be of limited use if it is not translated in the procedural system (Binder, 1999). The procedural system relates to skills in the application of knowledge (phronesis; e.g. interpersonal skills), whilst the reflective system is a metacognitive position which requires the skills to attend to, evaluate and interpret thoughts, feelings behaviours and subsequent outcomes during sessions. The DPR model is argued to be key to the development of expertise as a therapist and particularly emphasises the role of clinical supervision and self-practice/self-reflection.

Figure 2

Bennett-Levy's (2006) DPR Model of Competency Development



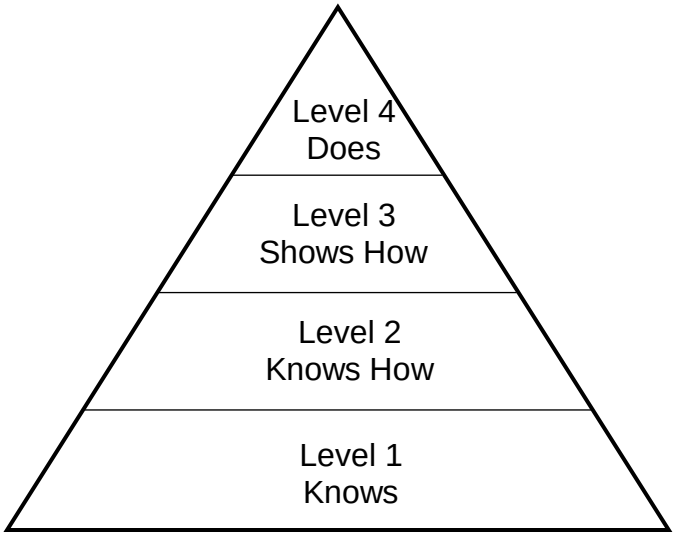
Assessment of Competency

There are various ways proposed and utilised to assess competence. Muse and McManus (2013) conducted a review of methods for assessing competence, specifically for CBT therapists. Their review is structured utilising Miller's (1990) framework for assessing clinical skills (see Figure 3). This framework defines skills assessments as being related to four hierarchical levels: Level 1, Knows; Level 2, Knows How, Level 3, Shows How, and Level 4 Does. Muse and McManus' (2013) review highlighted the types of assessment at different levels. At level 1, assessments such as essays and multiple-choice questions. At level 2, demonstration of practical understanding such as short answers, response to clinical vignettes and case reports. At level 3, assessments in which therapists demonstrate the skill such as in standardised role play and at Level 4 direct assessment of direct clinical work with clients demonstrating

competencies in real-world clinical practice. At level 4, observational measures are used to assess competency (e.g. Cognitive Therapies Scale; Blackburn et al., 2001), in which an expert observer watches or listens to a real session and evaluates competence based on pre-set criteria. This review recognised the distinction between domain-limited and global competencies and chose to exclude measures of global competence from the review. This is consistent with other authors who have highlighted that the literature about the assessment of competence in training is largely on adherence and domain limited competencies (Barber et al., 2007; Sharpless & Barber, 2009).

Figure 3

Representation of Miller's (1990) Hierarchical Framework of Competency Assessment



Competency and Outcome

In terms of the difference between competency and outcome, clinical competence is an attribute of a therapist measured on a valid and reliable observational measure, whereas patient outcome is the change scores achieved on a variety of nomothetic outcome measures (Bohart & Wade, 2013; Lambert, 2013). The association between competency and patient outcome is mixed

(Barber et al., 2007). When psychotherapies are combined there appears to be little association (Webb et al., 2010), but when meta-analysis is limited to CBT studies, a small significant total effect is apparent (r=0.24; Zarafonitis-Müller et al., 2014). When disorder-specific protocols drive interventions, a clearer association between competency and outcome emerges (Ginzburg et al., 2012). However, a more recent CBT study (Branson et al., 2015) found little support of an association between CBT competence and patient outcomes; but did find significant effects when comparing the best and the worst therapists (Branson et al., 2015).

Evidence suggests that different treatment modalities achieve similar outcomes in routine practice, known as the equivalence paradox (Hardy et al., 1998). This suggests that something other than domain-specific change methods enable change, known therefore as 'common factors' (Lambert, 2013). It is possible therefore, that issues with the focus of assessment (e.g. CBT competence) fail to capture common, more global factors of therapist competence. This may be due to global scores on CBT competency measures being reported, which tend to occlude sub-scale scores on interpersonal abilities of the therapist that create the core conditions. Whilst some attention has been paid to patient outcomes, these are generally in terms of positive outcomes, little attention has been paid to whether therapist competence relates to therapeutic deterioration or harm.

Previous Reviews

Four reviews have been conducted in this area. Herschell et al., (2010) conducted a systematic review of therapist training interventions. Results were presented by type of training intervention; reading of written materials or

treatment manuals, training workshops, training workshops with follow-up, self-directed training, pyramid training and multi-component training. The review found mixed evidence for reading of treatment manuals or written materials self-directed training and workshops. Findings were weakly indicative that these styles of training lead to improvement in knowledge (techne), but this was potentially short-lived and there was little evidence that these led to changes in clinical competency (procedural/ phronesis). There were too few studies about pyramid training to make conclusions. When workshops were followed up with supplemental intervention (e.g. feedback, consultation), there was some evidence that this resulted in translation of skills into clinical practice. The majority of studies were of multi-component training (e.g. including reading manuals, workshop training, consultation, booster training, training cases), and these were indicative of improvements in skills and/or knowledge. However, often the original studies in the review were evaluations of single training programmes, thus making it difficult to generalise and that these studies also lacked control groups, had small samples and were therefore vulnerable to maturation effects/bias.

Beidas and Kendall (2010) conducted a review structured from a systems-contextual perspective. As such, results were structured related to context with varying combinations of the following variables from training only (included in all), to training, organisational support, therapist and client variables. Similarly, to Herchsell et al., (2010), the findings were indicative that there is evidence of increases of self-reported knowledge, though for multiple studies identified, competence remained below the level required to be considered proficiently knowledgeable in the model. There was also lack of evidence that increases in knowledge translate into observer-rated competency. The strongest evidence for improvements following training in skill competence, as well as knowledge,

occurred when each element of the systems-contextual model was addressed. The lack of methodological control in the evidence-base was again noted and particularly highlighted issues with measurement relying heavily on self-report (i.e. which does not necessarily equate to actual improvements in clinical competency when independently observed).

Rakovshik and McManus (2010) specifically reviewed training interventions for CBT. Within the review's definition of competence, study inclusion criteria were studies of limited domain competence and adherence. Results were separated into three categories: 1) studies demonstrating therapists reached competence (meeting pre-existing guidelines on outcome measure), 2) therapists showing significant improvements but not meeting competence criteria and 3) studies not showing significant improvement. Findings indicated that where more training was provided studies were more likely to be in category 1 or 2 and where less training was provided the studies were more likely to be in category 3. In exploration of the active elements of training, the review noted that instruction (e.g. didactic teaching to improve declarative knowledge) may be necessary, but is insufficient to improve competency, whilst experiential learning (e.g. work with practice cases with feedback) seemed to be more beneficial. This review acknowledged that the majority of the original studies were not intended as training studies and highlighted the need for methodological improvements in the literature such as highlighting a need for clearer definitions and adopting a more controlled approach to researching training interventions.

The most recent review was conducted by Hill and Knox (2013). This was a review with an expansive scope considering the impact of training at both undergraduate, graduate and qualified levels as well as assessing the impact of

clinical supervision. The review found that much of the literature investigating undergraduate training is based on the Helping Skills model (Hill et al., 2007); designed to develop basic helping skills (e.g. empathy and reflecting), The review tentatively suggested that training for graduates and qualified therapists is effective at improving competence in empathy and reflection skills. However, in particular with graduate training they highlight methodological limitations relating to lack of control, particularly again over maturation variables and limited sample sizes.

Despite the methodological concerns repeated across the reviews, all four reviews are generally indicative that training appears somewhat effective at improving therapist competence. However, there are important caveats to this conclusion. It seems to be the case that the evidence is only consistent where training interventions are multifaceted and seems to be related to quantity (longer training more likely to increase competence). Also, they identify consistent and significant methodological weakness with the evidence base, primarily these relate to; lack of control over extraneous variables, small sample sizes, and reliance on self-report measures. Additionally, conspicuously, there is a lack of evidence about the developmental of global competencies by either intention (Rakovshik & McManus, 2010) or simply by omission of the distinction (Beidas & Kendall, 2010; Herschell et al., 2010; Hill & Knox, 2013).

Given these methodological flaws in the literature highlighted in the reviews, the current review returns to the question of the effectiveness of training therapists. Importantly, this is to understand whether this question can now be answered with fewer caveats than previously identified. The focus of this current review is the training of therapists of at least graduate level (training or qualified)

and informed by previous reviews, the focus is on training, provided face-to-face, and of at least 2-weeks in duration. This is to focus the current review on more substantial training interventions that are more likely to reflect the training programmes of professional therapists.

Aims

- 1) To assess the evidence for the effectiveness of psychotherapy training interventions.
- 2) To explore the extent to which the evidence base has responded to the methodological limitations highlighted previously.
- 3) To explore to what extent there is support for an increase across all types of competency.
- 4) To identify future methodological directions for training research.

Methods

The search strategy was defined a priori, and an initial protocol was registered online on 8th February 2020[1]. However, following subsequent scoping and the literature identified, it became evident that this protocol was not satisfactory to answer the research question. Therefore, before proceeding with systematic searches, this information was used to create a more refined protocol registered on 22nd February 2020[2]. See footnote for DOIs of these online registrations.

Search Strategy

In order to refine and define the research question and to aid the assistance of the development of search terms the PICO framework was utilised (Schardt et al., 2007; see Table 1).

[1] https://doi.org/10.17605/OSF.IO/PZCBQ
[2] https://doi.org/10.17605/OSF.IO/VGMZD

Table 1

PICO Framework utilised to refine search strategy

Population	Therapists delivering psychotherapy who are either in or have completed postgraduate professional training
Intervention	Any training intervention which is aimed at improving the clinical skill, knowledge or competence of participants
Comparison	At minimum, those studies which have included pre and post measurements.
Outcome	Outcomes must assess clinical skill, knowledge or competence of participants.

Studies were identified by use of systematic searches of PsychINFO, Medline and SCOPUS. Initial searches were conducted on 23rd February 2020, each search was run again on 7th March 2020 to confirm that no new literature had been published prior to data extraction and analysis. Keyword searches were utilised and applied differentially to suit the search mechanics of each database. For an example of the full search terms in SCOPUS see Appendix 1. The search terms used were:

training

AND therapist OR psychologist OR psychotherapist OR "cognitive behavioral therapist" OR "cognitive behavioural therapist" OR cbt OR counselor OR counsellor

AND abilit OR skill OR competence*

References were managed using Mendeley reference manager. Articles identified from all three database were combined and duplicates were removed using Mendeley's "Check for Duplicates" tool. Subsequently, articles identified were screened iteratively. Screening was completed by the first researcher, where the first researcher was not confident in the decision, this was discussed with a secondary researcher and a shared decision about inclusion was made.

Screening was conducted against the study inclusion and exclusions criteria illustrated in Table 2. Screening was conducted by firstly considering title and abstract and then by full-text. As large amounts of irrelevant studies were identified in title and abstract screening reasons for exclusion were not recorded. Reasons for exclusion at full-text screen were recorded. Of the papers identified for inclusion, citation chaining analysis was conducted. This included screening reference lists for relevant studies and utilising the “Cited by” function in Google Scholar; this identified no further studies. The grey literature was searched using Google Scholar and the British Library of Online Theses, this identified one additional dissertation eligible for inclusion. The process of study selection is represented in by a PRISMA diagram below (see Figure 4;Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher et al., 2009).

Data Extraction

The studies identified investigated a range of outcomes with varying methodologies. To retain focus on the research question only data and outcomes relevant to the sample of interest were extracted. Therefore, data related to patient outcomes and patient measures as a proxy of therapist skills were not extracted. Additionally, only the quantitative components of the studies were extracted. A bespoke data extraction tool was created containing the following information: authors, year of publication, design, measurement time-points, outcome measurements (including type of assessment, type of competency), location, participant information (including number, gender, age, profession), training details (including length, components), study outcomes (including p-values additional reported statistics). This was first piloted with three studies and found to be fit for purpose.

Table 2

Inclusion and Exclusion Criteria

Inclusion	Exclusion
Therapists either in or after professional training	Not published in English
Papers published 2013-present	Training interventions with no face to face component
Full text available	Studies with duration<2 weeks
	Qualitative studies
	Preliminary/ pilot/ feasibility studies

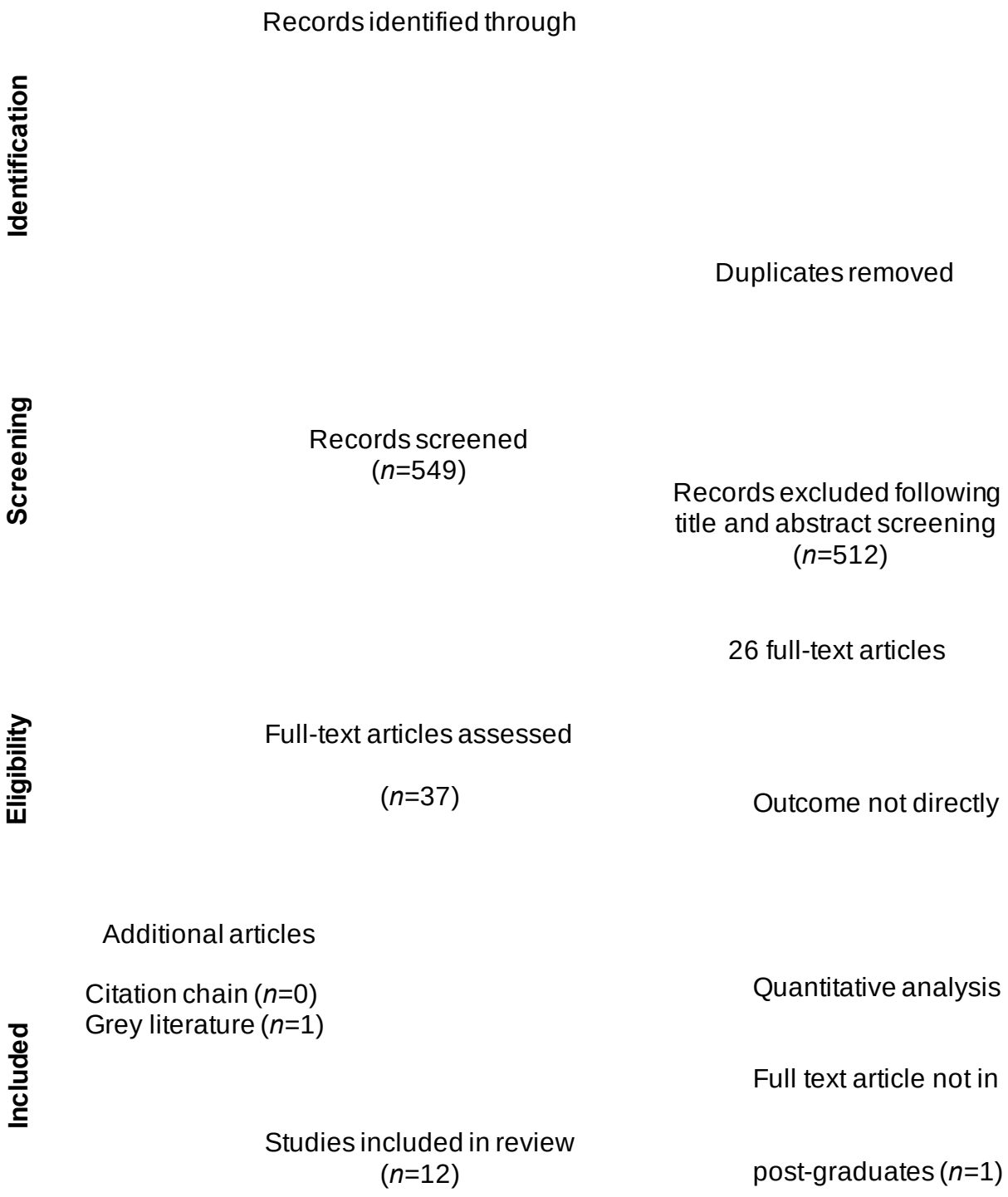
Quality Appraisal

To inform interpretation of the identified studies structured quality assessment of the papers identified was conducted. Due to the majority of studies being case series (also known as time series/ pre-post; National Institute for Clinical Excellence, 2012), a case series methodological quality design tool was used (Moga et al., 2012; Appendix 2). This checklist covers a range of areas related to study population, the intervention outcome measures and results. There are 18 items on the checklist however, item 6 was not utilised in this analysis as it related to a specific question about disease progression, not relevant to this review. The items were scored 0-1 therefore, it was possible for papers to score between 0-17.

Quality analysis was conducted by the primary researcher and a secondary researcher with specialist training in children's nursing, education, research and coaching. The secondary researcher was selected, as they were not familiar with the literature and therefore were less likely to have preconceived ideas. All papers were assessed independently and compared to assess inter-rater reliability (breakdown of appraisal scores in Appendix C). According to Landis and Koch, (1977) Cohen's Kappa Coefficient suggested "substantial agreement" ($\kappa = 0.66$, $p < 0.001$, 95% CI= 0.55-0.76; Appendix D). Data quality

Figure 4

PRISMA Diagram Illustrating Study Selection Process (Moher et al., 2009)



assessment was not utilised to define inclusion/ exclusion of studies, rather it was utilised to understand the quality of the existing literature and to give a sense of relative quality between studies. Whilst these scores give a useful indication of relative study quality, high scores should not be interpreted as universally strong methodology; use of this tool in itself (case-series) reflects a lack of randomised and controlled studies and is indicative of methodological limitations of the studies in the review.

Results

From systematic literature searches a total of 696 studies were identified. After removing duplicates across databases, this was reduced to 549 which were screened. Following title and abstract screening this was reduced further to 37 for full-text screening. Of the studies removed at full-text screening, studies were excluded due to: the intervention lasting less than two weeks (n=12), pre-post measures not taken (n=5), outcome measure not of therapist skill, ability or competence (n=3), pilot, preliminary or feasibility (n=2), quantitative data not reported for all participants (n=2). participants not post-graduate therapist (n=1) and full-text unavailable in English (n=1). This left 11 studies for inclusion. Forward and backward searches did not identify any further studies meeting inclusion/ exclusion criteria, searches of grey literature identified one additional thesis which met inclusion criteria. This meant there was a final sample of 12 studies included. See Table 3 for a summary of study characteristics, see Table 4 for summary of results and Table 5 for quality appraisal and comment on methodology of studies.

Full Training Course

Description of Studies

Five studies identified assessed the effectiveness of training courses at improving competencies of trainees (Beale et al., 2020; Hill et al., 2015; Liness et al., 2019a; Liness et al., 2019b; Pearson & Weinberg, 2017). The quality scores for these studies ranged from 9-15 with an average of 12.4. Four studies utilised a pre-post design and one study utilised a quasi-experimental design (Pearson & Weinberg, 2017). However, as the comparison group in the quasi-experimental design study was undergraduate trainees this data was not extracted, and therefore for the purposes of this review this paper is treated as pre-post also. Only two of the five studies collected follow-up data (Beale et al., 2020; Liness et al., 2019b). The sample sizes ranged from N=45-252 for four studies. One study had a smaller sample (N= 23), of which not all were included in all statistical analysis (Hill et al., 2015).

Training Implemented

Three of the five studies studied a single IAPT CBT postgraduate training course. This training lasted 12-months, the length of training in the other studies is less clear; Pearson and Weinberg (2017) indicate participants had teaching for 7-months but this was not clear if this was the length of the course. Whilst Hill et al., (2015) indicated participants could be in the training clinic for between 12-42 months it was difficult to know how long each trainee received training for. All of the training courses had large components of practice with real clients and received clinical supervision either individually, in groups, or both. Hill et al., (2015) provided minimal additional information about the training course. The other training courses indicated that training also included university study days, didactic teaching and skills training reflecting multiple components implemented in these training programmes.

Assessment Methods Used

All three studies assessing the CBT training course employed the Revised Cognitive Therapies Scale (CTS-R) to measure competence. This measure was used as an observer report in two of the studies (Liness et al., 2019a; Liness et al., 2019b) and both observer and self-report in the third (Beale et al., 2020). Assessment was made of trainee therapists recorded sessions, reflecting assessment at level 4. One study utilised the Emotional Intelligence Self-Assessment Questionnaire (EISAQ), a global competence assessed at level 2 (Pearson & Weinberg, 2017). The final study utilised two measures, the Post-Training Questionnaire (PTQ), this was a self-report assessment of abilities designed specifically for this study, and the Comparative Psychotherapy Process Scale (CPPS) a measure of therapist use of cognitive behavioural and psychodynamic interpersonal skills (Hill et al., 2015).

Findings

Each of the five studies demonstrated significant increases in competencies. Three studies assessing CBT competency showed significant increases in CBT competence. Competence was also maintained at follow-up, although follow-up was highly variable in one study (Liness et al., 2019b), and there was a smaller sample size in the follow-up in the Beale et al., (2020) study. Additionally, Beale et al., (2020) found agreement between self and observer reported competence increased between pre and mid ratings, but then remained stable at follow-up measurements. When competent versus incompetent (defined by observer ratings of competence) comparisons were made, it was found that competent trainees self-rated scores correlated with observer rating, but non-competent trainees did not, at any time point. Additionally, non-competent trainees had a greater tendency to overestimate their competence, whilst

competent trainees were more likely to underestimate competence suggesting differential ability to recognise own level of competence. Scores were found to be significantly higher at the end of training on the EISAQ, measure of emotional intelligence (Pearson & Weinberg, 2017). The final study (Hill et al., 2015) was less clear in its findings; of 24 items on the PTQ, participants self-reported significant improvements on 11 items (e.g. use of challenge and open questions) leaving non-significant changes on the remaining 13 items, this was not a standardised measure of competence rather it was designed for the purpose of this study. On the CPPS, significant increases in cognitive behavioural techniques occurred when assessed in session 3 but not at session 9 and there were no changes in use of psychodynamic techniques at any point. Only a small sample of the participants were included in this analysis, so it is difficult to say whether this was reflective of lack of change or whether this reflects a lack of power. Overall these findings are indicative that training courses do increase competence.

Limitations

Three of the five studies were conducted on the same training course and followed the same cohorts of students. The overlap between participants in these studies is not clear. These findings should be interpreted with due caution, as overlap could inflate the findings. Also, these studies were of individual training courses which means that it is difficult to know whether the findings generalise to other training contexts. Additionally, given the multiple components that were used as part of the training, there is no way of ascertaining from these studies which components of training were important in changing competence of trainees, making it difficult to apply the findings to inform future training. Also, a lack of control groups in these studies make it more difficult to conclude about

the effectiveness of training, as other factors such as maturational effects could be responsible for changes observed.

Table 3

Summary of Study Characteristics

Authors	Design	Sample/ Setting	Intervention
Beale et al., (2020)	Pre-post	**Description:** CBT post-graduate trainees. **Location:** UK. **Number:** 150**. Female/Male:** 116/34. **Age:** Median=32. **Profession/Background:** Psychological Wellbeing Practitioner=50, Clinical Psychologist=49, Counselling Psychologist=18,Counsellor/Psychotherapist=17, Mental Health Nurse=8, Occupational Therapist=5, Other=3	**Training Details:** IAPT Cognitive Behavioural Therapy training course. Including, practice with real clients, supervision, didactic teaching, academic study days, skills training. **Duration:** 1 year.
Chaddock et al., (2014)	Small n	**Description:** CBT post-graduate trainees**. Location:** UK. **Number:** 4**. Female/Male:** Not stated. **Age:** Not stated. **Profession/Background:** Nurses	**Training Details:** Self-reflection self-training workbook. Completed within an IAPT Cognitive Behavioural Therapy training course. **Duration:** 3 months.
Chao et al., (2017)	RCT	**Description:** White counsellor psychology trainees. **Location:** China. **Intervention Group**, **Number:** 52, **Female/Male:** 34/18. **Age:** Range=23-42. **Profession:** not stated. **Control Group, Number:** 22. Female/ Male: 15/7 **Age:** range=23-42	**Training Details:** Multicultural training course including didactic learning, experiential exercises, role plays, community service, group discussion and written essays. Completed within a counselling psychology training course. **Duration:** 3.75 months.
Hill et al., (2015)	Pre-post	**Description:** Doctoral Counselling Trainees. **Location:** USA. **Number:** 23. **Female/Male:** 15/8. **Age:** Mean=29.65	**Training Details:** Counselling Psychology Doctoral Program with focus on learning and implementing psychodynamic and interpersonal constructs. Including practice with real clients. **Duration:** Between 12 to 42 months. Unclear when end-of-training measures administered
Hume (2017)	Small n	**Description:** CBT post-graduate trainees**. Location:** New Zealand. **Number:** 7. **Female/Male:** 7/0**. Age:** Mean=47.5 range=33-57**. Profession/Background:** Counselling=2, Social Work=2, Nursing=2, Unknown=1	**Training Details:** Self-reflection self-training workbook. Completed within a training course for postgraduate diploma in CBT. **Duration:** 3.5 months.

Authors	Design	Sample/ Setting	Intervention
Keng et al., (2017)	RCT	**Description:** Masters Clinical Psychology Trainees. **Location:** Singapore. **Number:** 25. **Female/Male:** 18/7. **Age:** M=28.08, range=24-59	**Training Details:** Functional analytic psychotherapy training including didactic teaching, experiential exercises, reflective logs, group discussion. Completed within a masters clinical psychology training course. **Duration:** 2 months.
Liness et al., (2019a)	Pre-post	**Description:** CBT post-graduate trainees**. Location:** UK. **Number:** 252. **Female/Male:** 201/51. **Age:** Mean=34.23. **Profession/ Background:** Clinical Psychologist=85, Counselling Psychologist=37, Mental Health Nurse=26, Occupational Therapist=9, Counsellor=7, Forensic Psychologist=2, Psychiatrist=1 Social Worker=1, Psychological wellbeing Practitioner= 69, non-accredited counsellor/psychotherapist=15	**Training Details:** IAPT Cognitive Behavioural Therapy training course. Including, practice with real clients, supervision, didactic teaching, academic study days, skills training. **Duration:**1 year.
Liness et al., (2019b)	Pre-post	**Description:** CBT post-graduate trainees. **Location:** UK. **Number:** 45**. Female/Male:** 33/12. **Age:** Median=33, IQR= 9. **Profession/ Background:** Clinical Psychologist=16, Psychological Wellbeing Practitioner=14, Counselling Psychologist=3, Other =3.	**Training Details:** IAPT Cognitive Behavioural Therapy training course. Including, practice with real clients, supervision, didactic teaching, academic study days, skills training. **Duration:** 1 year.
Murray (2017)	Pre-post	**Description:** IAPT therapists with CBT training**. Location:** UK. **Number:** 20. **Female/Male**: 19/1. **Age:** Not provided. **Profession/Background: "**Varied" including: clinical and counselling psychology, psychiatric nursing,	**Training Details:** Post-qualification training for treatment of PTSD including four day workshops and webinars, including roleplays, therapy videos and experiential exercises plus additional consultation with trainer**. Duration:** 6 months.
Pakenham (2015)	Pre- post	**Description:** Clinical Psychology Trainees. **Location:** Australia. **Number:** 32**. Female/Male:** 28/4 **Age:** Mean=27.66, Range=21-50	**Training Details:** ACT training including practice with real clients, didactic teaching, case conceptualisation focus on self-care. Completed within a clinical psychology training course. **Duration:** 3 months.

Authors	Design	Sample/ Setting	Intervention
Pearson and Weinberg (2017)	Quasi-experimental pre-post (only one group reported on in this review)	**Description:** Undergraduate students (not reported here) and Postgraduate Counselling Trainees**. Location:** UK. **Number:** 58. **Female/Male:** 44/14**. Age:** Mean=35.4years, SD=10.4.	**Training details:** Counselling training course including practice with real clients, one study day per week, personal development groups, skills training lectures and supervision. **Duration:** Teaching lasted 7 months but overall training unclear.
Rantanen and Soini (2018)	Pre-post	**Description:** Masters Educational Psychologists. **Location:** Finland. **Number:** 16. **Female/Male:** 14/2**. Age:** not stated**. Profession/Background:** not stated	**Training details:** Peer group consultation training model including, practice with simulated clients, skills training, group discussion, written reflection reports. **Duration:** 1 month.

Note: IAPT= Improving Access to Psychological Therapies, PTSD= Post-traumatic Stress Disorder

Specific Training and Training Course

Description of Studies

Six studies assessed the effectiveness of specific training conducted in addition to broader training courses (Chaddock et al., 2014; Chao et al., 2017; Hume, 2017; Keng et al., 2017; Pakenham, 2015; Rantanen & Soini, 2018). Two studies utilised pre-post design (Pakenham, 2015; Rantanen & Soini, 2018), two studies controlled designs (Chao et al., 2017; Keng et al., 2017) and two studies utilised a small n design (Chaddock et al., 2014; Hume, 2017). Four of the six studies collected follow-up data. The quality scores for these studies ranged from 8 to 13 with an average of 10. Only Keng et al., (2017) scored above the average for all papers assessed. Sample sizes were generally small ranging from 4-52 participants including two studies with very small samples (N=4 in Chaddock et al., 2014; N=7 in Hume, 2017).

Training Implemented

The training length of these programmes ranged from 1-4 months. As with the full training courses only these training programmes also generally had multiple components, apart from two studies evaluating the use of self-reflection and self-practice workbook (Chaddock et al., 2014; Hume, 2017). There was a common theme of these training interventions focussing on the personal aspect of therapist development (Chaddock et al., 2014; Hume, 2017; Keng et al., 2017; Pakenham, 2015), and this was explored through activities such as reflective logs, experiential exercises and group discussions. Only Pakenham, (2015) reported training application of skills with real clients, whilst one study utilised role plays to develop and assess skill (Rantanen & Soini, 2018). Many of the studies also included some aspect of didactic teaching. (Chao et al., 2017; Keng et al., 2017; Pakenham, 2015; Rantanen & Soini, 2018). Although the training

interventions are typically well described in these studies, the broader training programmes within which these courses were run were not well described.

Assessment Methods Used

A range of assessment methods were utilised, primarily these were self-reported representing assessment at level 2. Chaddock et al., (2014) and Hume (2017) both used self-reported measures, the Cognitive Therapist Self-Monitoring Scale (CTSMS) and the Cognitive Therapy Empathy Scale (CTES). Chaddock et al., (2014) did not apply the full scale and asked participants to select only items pertinent to them. Additionally, Hume (2017) also utilised the Cognitive Flexibility Inventory (CFI) and the Self-reflection and Insight Scale (SRIS) whilst Pakenham (2015) utilised the Counsellor Activity Self-Efficacy Scale (CASES) all reflecting level 2 assessments of global competencies. Chao et al., (2017) used the Multicultural Counselling Knowledge and Awareness Scale relating to knowledge level of competence. Keng et al., (2017) also utilised self-report measures, the Interpersonal Reactivity Index (IRI) and the Facilitative Analytic Psychotherapy Scale Impact Scale (FAPIS; measure of competencies related to FAP). Additionally, Keng et al., (2017) also employed an observer rated measure of quality of response to a case vignette (FAP-Vig). Similarly, Rantanen & Soini (2018) utilised the observer rated Counsellor Response Coding System (CRCS), a level 3 assessment as this was a measure of trainees' competence with a simulated patient. The CRSC measured use of skills rather than a judgement of the skill in application of the skill, and therefore reflects an assessment of knowledge (to apply the skill).

Findings

Whilst positive outcomes were reported across all studies there were

significant methodological limitations in some that makes it difficult to draw meaningful conclusions. The strongest evidence was found by Keng et al., (2017) who had a control group. Keng et al., (2017) found significant between group effects on observer rated FAP competence and the IRI but not self-rated FAP competence. However, within group analysis did indicate significant improvements both at post and follow up in observer rated FAP and at follow-up in self-rated FAP competence. The post self-rated score did not reach significance, despite having medium effect size (d=0.53) this may reflect the study being underpowered to detect this difference at 5% significance level. There was some evidence that training could lead to improvements in the self-efficacy of trainees (Pakenham, 2015). The findings in the remaining studies should be interpreted with additional caution. Rantanen & Soini, (2018) analysed the questionnaire measure implemented at the individual level and found significant increase in use of reflection and significant decrease in self-disclosure but no significant differences in other items. However, as items of this scale were assessed on the individual level rather than global score, the p value should have been adjusted to take into account multiple comparisons (e.g. 0.05 divided by 10 comparisons= p-value of 0.005) which would have meant that reflection would not have been a statistically significant change. Therefore, of ten comparisons, only one improvement could be considered statistically significant, which is not very strong evidence for the efficacy of this training intervention. It could be due to small sample size or reflect the shortness of this intervention leaving too little time for change to occur. Two studies indicated some improvements in self-rated skills such as self-reflection and empathy, however these studies did not utilise inferential statistics, preferring in depth analysis of a small number of participants making it difficult to draw conclusions or make generalisations from these

Table 4

Summary of the Results of Studies

Authors	Measures([a],[b])	Description of Outcome
	Qualitative Description	
Beale et al., (2020)	**Self:** CTS-R (4, B). **Observer:** CTS-R (4, B). Also, C by comparison between measures.	Significant correlations between self- and observer-reported competence (Baseline: p<0.001, r=0.29; Mid: p<0.001, r=0.50; End p<0.001, r=0.27). This was only the case for competent therapists (Baseline: p<0.001, r=0.39; Mid: p<0.001, r=0.60; End p=0.004, r=0.27). Therapists rated as non-competent by observers scores did not correlate, non-competent therapists typically overestimated their competence, particularly at end of training (Baseline: p=0.56, r=0.10; Mid: p=0.13, r=0.25; End p=0.14, r=0.24).
	Changes in ability to accurately assess one's cognitive therapy skills (CTS-R).	
Chaddock et al., (2014)	**Self:** selection of items from CTSMS (2, C) and CTES (2, C)	Visual analysis indicated differential responses to training, some participants scores increased whilst others decreased or stayed the same.
	Changes in items in self-monitoring (CTSMS) and empathy (CTES)	
Chao et al., (2017)	**Self:** MCKAS (1, A)	Reliable and statistically significant change reported in MCKAS scores for experimental group but not in comparison group (p=not reported, reliable change index=3.05). However, means provided in the paper do not seem to support this. A t-test comparing scores using data provided in the paper suggests there were no significant differences between groups at baseline (p=0.83), but scores in the comparison group were significantly higher (desirable) in the comparison group than experimental group at end-of-intervention (p=0.02) and follow up (p=0.01).
	Changes in multicultural awareness (MCKAS).	
Hill et al., (2015)	**Self:** PTQ (2, B/C) **Observer:** CPPS (4, A).	Significant changes observed in 11/24 self-rated abilities (p<0.001). Significant observer rated increase in frequency of use CB skills at session 3 (p=0.03, t=2.58) but not at session 9 (p>0.05, t=-0.55) and no increase in use of PI skills at either time (session 3: p>0.05, t=0.02; session 9: p=0.05, t=-0.66).
	Changes in self-rated abilities (PQ) and in observer rated use of cognitive behavioural (CB) and psychodynamic interpersonal skills (PI; CPPS)	

Authors	Measures([a],[b]) Qualitative Description	Description of Outcome
Hume, (2017)	**Self:** CTES, CTSMS, CFI, SRIS (2, C) Changes in self-rated empathy (CTES), self-monitoring (CTSMS), cognitive flexibility (CFI) and self-reflection (SRIS).	Descriptive statistics and visual analysis suggestive of improvements on all measures, clearer increase pre-post CTSMS (Mean change=35.1%). Improvements but less clear on CFI (Mean change=7.4%), CTES (Mean change=4.1%) and SRIS (Mean change=2.9%).
Keng et al., (2017)	**Self:** IRI (2, C), FAPIS (2, B) **Observer:** FAP-Vig (3, B) Between and within group changes in self-rated emotional and cognitive empathy (IRI) and competencies related to functional analytic psychotherapy (FAPIS) and in observer rated quality of response to a case vignette (FAP-Vig).	Significant between group effect pre-post on IRI ($p=0.02$, $f^2=0.32$) and FAP-Vig ($p=0.04$, $f^2=0.23$) but not FAPIS ($p=0.20$, $f^2=0.09$). Significant pre-follow up on FAP-Vig ($p=0.02$, $f^2=0.44$) but not IRI ($p=0.15$, $f^2=0.11$) or FAPIS ($p=0.37$, $f^2=0.04$). Significant within group effect pre-post IRI ($p=0.02$, $d=0.80$) and FAP-Vig ($p=0.02$, $d=0.76$) but not FAPIS ($p=0.09$, $d=0.53$) .Significant within group effect pre-follow up on IRI ($p=0.04$, $d=0.66$), FAP Vig ($p=0.03$, $d=0.73$) and FAPIS ($p=0.03$, $d=0.71$). No significant changes were identified in control group pre-post (IRI: $p=0.34$; FAP Vig: $p= 0.89$; FAPIS: $p= 0.61$) or pre follow-up (IRI: $p=0.69$; FAP Vig: $p= 0.97$; FAPIS: $p= 0.18$).
Liness et al., (2019a)	**Observer:** CTS-R (4, B) Changes in observer rated cognitive therapy skills (CTS-R)	Significant increases on CTS-R and CTS-R subscales between each time points in treatment of both depression (Total: $p<0.001$, partial $\eta^2=0.50$; Generic: $p<0.001$, partial $\eta^2=0.36$; Specific: $p<0.001$, partial $\eta^2=0.76$) and anxiety (Total: $p<0.001$, partial $\eta^2=0.49$; Generic: $p<0.001$, partial $\eta^2=0.33$; Specific: $p<0.001$, partial $\eta^2=0.72$). Post-hoc analyses indicate significant differences between pre-mid and mid-end on both depression and anxiety competence ($p<0.001$).
Liness et al., (2019b)	**Observer:** CTS-R (4, B) Changes in observer rated cognitive therapy skills (CTS-R)	Significant increases on CTS-R and CTS-R subscales between time points (Total: $p<0.001$, partial $\eta^2=0.27$; Generic: $p=0.004$, partial $\eta^2=0.17$; Specific: $p=0.003$, partial $\eta^2=0.0.19$). Post-hoc analysis indicated significant difference between pre and post scores ($p<0.001$) but not post and follow-up scores ($p=0.10$) indicating maintenance.

Authors	Measures([a],[b]) Qualitative Description	Description of Outcome
Murray (2017)	**Self:** PTSD Checklist (PTSD C) (1, A). **Observer:** CTS-PTSD (4, B) Changes self-rated in PTSD specific competencies (PTSD C) and observe rated changes in overall CBT, PTSD specific, CBT generic competencies (CTS-PTSD).	Significant increase in self-rated PTSD C (pre-post: $p<0.01$, $z=2.45$; pre to follow-up: $p<0.01$, $z=2.42$) maintained at follow up (post to follow-up: $p>0.05$, $z=0.41$). Significant increase in CTS-PTSD for PTSD specific skills pre-post ($p<0.05$, $z=2.09$), pre-follow up ($p<0.01$, $z=2.35$), maintained post-follow up ($p>0.05$, $z=1.02$). Supervisor rating of overall competence increased between pre and post ($p<0.05$, $z=1.92$), but not pre-follow up ($p>0.05$, $z=0.98$) or post to follow up ($p>0.05$, $z=0.33$) and there no significant changes in the generic subscale at all time points ($p>0.05$)
Pakenham (2015)	**Self:** CASES (2, C). Changes in self-rated self-efficacy in helping skills (CASES).	Significant increases between pre and post scores on CASES ($p<0.05$, $t=-2.73$)
Pearson and Weinberg (2017)	**Self:** EISAQ (2, C) Changes in self-rated emotional intelligence (EISAQ)	Significant increases between pre and post scores on EISAQ ($p<0.001$, $z=4.45$)
Rantanen and Soini (2018)	**Observer:** CRCS (3, A) Changes in observer rated frequency and types of responses	Significant increase in use of reflection ($p=0.04$, $t=-2.21$) and reduction in self- disclosure ($p=0.001$, $t=4.22$) but not on other indices (questioning, conclusion, suggestion, feelings, explanations, action, context, strengths)

Note: CTS-R= Cognitive Therapy Scale Revised, CTSMS= Cognitive Therapist Self-Monitoring Scale, CTES= Cognitive Therapy Empathy Scale, MCKAS= Multicultural Counselling Knowledge and Awareness Scale, PTQ= Post-training Questionnaire. CPPS= Comparative Psychotherapy Process Scale, CFI= Cognitive Flexibility Inventory, SRIS, Self-reflection and Insight Scale, IRI=Interpersonal Reactivity Index, FAPIS=FAP Impact Scale, FAP-Vig= FAP-Vignette. CTS-PTSD, Cognitive Therapy Scale- Post Traumatic-Stress Disorder, CASES= Counsellor Activity Self-Efficacy Scale, EISAQ= Emotional Intelligence Self-Assessment Questionnaire. CRCS= Counsellor Response Coding System.

[a]=Level of assessment from Miller (1990) 1-4. [b]=Level of competence assessed, A=Adherence/ Knowledge, B=Domain Limited, C= Global.

findings (Chaddock et al., 2014; Hume, 2017). Finally, Chao et al., (2017) demonstrated significant increases of multicultural awareness and knowledge, however, despite use of a control group and randomisation, this was identified as joint lowest quality study in the review (quality score=8). The results are difficult to understand, and it does not appear that their conclusions are substantiated by mean scores reported.

Limitations

Whilst the aim of these studies was to assess specific training programmes a lack of control groups makes it difficult to establish whether the training identified was responsible for changes observed in outcomes, compared with the broader training trainees received. Only Keng et al., (2017) was adequately designed to make these conclusions. This is particularly important as whilst the training interventions were generally well described, other training as part of the larger programmes were not, which likely contributed to the changes observed. Additionally, these studies generally had small sample sizes this effects the generalisability of the samples; additionally it might be that studies were underpowered to detect significant changes and hence made type-2 errors.

Qualified Therapists

Description of Study

Only one study identified assessed training for already qualified therapists (Murray, 2017). This study had 20 participants and utilised a pre-post design with a one-year follow up. Relative to other studies in the review this study quality was just above the average (quality score=13).

Training Implemented

This training course lasted 6-months designed to train therapists in the CBT treatment of PTSD. It comprised of four full day workshops, four half day webinars and monthly model specific consultation with the trainer. Whilst the training was taking place participants worked within an IAPT service whilst they continued to provide therapy to patients.

Assessment Methods

This study utilised an observer and self-report measure of competence using the CTS, adapted for PTSD and a self-reported checklist of PTSD competencies. The observer report was of assessment of clinical practice and therefore level 4 whilst the self-reported was categorised as a level 2 assessment. Both assessed competencies of knowledge/adherence and domain limited competency.

Findings and Limitations

The study reported significant self-rated improvements in PTSD checklist competencies and observer-rated specific PTSD competencies between pre-post and pre follow-up. However, no significant changes in observer rated overall and generic CBT skills were found. In this study, not finding changes in more generic CBT skills could be reflective of therapists already having been assessed as competent to pass their training course or may reflect the specificity of the training to PTSD related skills. It is not clear how representative the sample of this study was of the population, as participating services volunteered two participants each, which could have resulted in a selection bias.

Table 5

Methodological Critique of the Studies

Authors	Critique	Quality Score
Beale et al., (2020)	Much smaller number of participants at follow-up and generally participants at follow-up considered to be competent, not allowing for comparisons between competent and non-competence therapists, no control group.	15
Chaddock et al., (2014)	Very small number of participants, little information about participant characteristics, inferential statistics not conducted, outcome measures not used in a way consistent with standardised design, no meaningful control group.	8
Chao et al., (2017)	Method of data analysis presented in results section, results not presented in clear understandable way and conclusions within the paper do not appear to be supported by descriptive statistics included in the paper.	8
Hill et al., (2015)	The primary self-report measure was not standardised with psychometric qualities assessed, additionally items analysed on an individual item level (though adjustments for multiple comparisons were made), no control of participants receiving training as normal or no training.	9
Hume, (2017)	Very small number of participants, inferential statistics not conducted, difficult to generalise from.	12
Keng et al., (2017)	Small numbers in each group means it is possible not enough power resulted in some type-2 errors, unclear if this was the case as post-hoc analysis not conducted.	13
Liness et al., (2019a)	No follow up, no control group receiving no training.	13
Liness et al., (2019b)	A lot of variability in the length of time to follow-up, with little control on other factors in this time. No control group receiving no training.	15
Murray (2017)	Participants recruited by therapy services volunteering two therapists each, introduction risk of selection bias, no control group receiving no training.	13
Pakenham (2015)	No follow up, no control group receiving training as normal or no training.	9

Authors	Critique	Quality Score
Pearson and Weinberg (2017)	No follow up, no control group receiving no training.	10
Rantanen and Soini (2018)	Training intervention relatively short, small number of participants, limited information about participants, outcome measure assessed at individual item level without adjusting for multiple comparisons.	10

Discussion

This systematic literature review aimed to assess whether training interventions for post-graduate trainee therapists and qualified therapists actually improve the competence of therapists. It aimed to provide an update to reviews published in 2010 and 2013. Twelve new articles were identified and included. Five studies investigated full training programmes, six studies investigated specific training interventions as part of a full training programme and one study investigated the training of qualified therapists. Overall, the studies were generally positive indicating the effectiveness of training interventions. However, this is reported cautiously, given the significant methodological limitations within and across these studies. The quality assessment of the studies in general suggested a methodologically weak evidence-base and therefore conclusions drawn in this discussion are tentative.

The strongest and most consistent evidence in this review was found for full training programmes and a training programme for qualified therapists. This evidence was supportive of specific improvements in CBT competence (Beale et al., 2020; Liness et al., 2019a; Liness et al., 2019b; Murray 2017). There was some indication from these training programmes that training can also improve self-reported global competencies (e.g. emotional intelligence, dealing with difficult patients; Hill et al., 2015; Pearson & Weinberg, 2017). Additionally, a novel insight was provided by Beale et al., (2020), who captured global competence of self-awareness/ self-reflection by identifying improvements in trainee abilities to accurately assess their own competence. There are several components that these training programmes had in common. For example, these training programmes all included a large component of closely supervised

practice with real-world clients and these training programmes were relatively longer in duration.

There was some support of these findings from studies assessing specific training interventions. The specific training programme that did report practice with real world clients was also positive, reporting improvements in trainee's self-efficacy (Pakenham, 2015) suggestive that this may be an important component of training. Additionally, Keng et al., (2017) illustrated improvements in FAP which is supportive that training can improve specific modality competence. Contrary to the full training programmes Rantanen and Soini (2018) only reported a significant effect on two out of ten (one out of ten if interpretation in this review is taken) variables assessed which is seen as fairly weak evidence of effectiveness. This training programme was the shortest included in the review at only one month long; this is arguably consistent with a view that training requires longer duration to be effective. The remaining studies are considered especially weak evidence for the purpose of this review. One study provided support that training could increase self-monitoring skills and did have methodological strengths (reflected in appraisal score of 12). However, a small number of participants and lack of inferential statistics make it difficult to use this evidence to inform conclusions (Hume, 2017). The final two studies had weak methodology reflected in their appraisal score (both=8) and it was therefore difficult to draw any meaningful conclusions (Chaddock et al., 2014, Chao et al., 2017).

There was a range of methods of assessment of competency utilised in the studies, assessing different levels and types of competency. The highest level of assessment (according to Miller, 1990) was typically completed by observers reviewing video tapes of sessions and scoring trainees according to competence

within a model. This either reflected judgements about competency (e.g. Liness et al., 2019a; Liness et al., 2019b; Keng et al 2017; Murray, 2017) or frequency use of model specific skills (e.g. Rantenen & Soini, 2018). Many of the studies also employed self-reported measures which incorporated measures about knowledge (e.g. Murray, 2017, Chao et al., 2017) and global competencies (e.g. Pearson & Weinberg, 2017). The several studies utilising observer measurement represents some improvement in the evidence base which had previously been critiqued for a focus on self-report measures (Beidas & Kendall, 2010). However, even where higher level of assessment was employed, these were generally at the domain limited level of global competency, omitting assessment of broader global competencies. This could reflect the way in which competency is assessed and defined by training courses (e.g. Roth & Pilling, 2008) and may indicate a need for these to be more nuanced, including all the technical, personal and interpersonal competencies required to be an effective therapist (Koddebusch & Hermann, 2018). It could also reflect the difficulty of measuring these concepts by direct assessment in clinical practice (Bennett-Levy, 2019). However, there are possibilities, for example, Anderson and Patterson (2013) developed a direct assessment of interpersonal skills which could be applied to assessment of trainee competence. Similarly, Tweed et al., (2010) developed a measure of clinical skills which is appropriate regardless of treatment modality. This is an important oversight in the literature as if these competencies cannot be trained, it would have important implications for recruitment onto therapist training programmes.

Although all studies displayed at least some significant improvements, this was not the case for all of the outcomes measured. In some studies, where changes did not meet significance it is possible that this was related to limitations

in the sample size meaning the studies were potentially underpowered. Unfortunately, as power calculations were not conducted, it is difficult to say whether this was the case. This means there is risk that these studies made type-2 errors which would understate the effectiveness of the training interventions. This is a limitation which has been highlighted in previous reviews (e.g Herschell et al., 2010).

To consider potential mechanisms, Bennett-Levy's model of skills acquisition (2006) appears useful. Whilst declarative learning can occur quickly, procedural learning requires practice and takes longer for consolidation to occur (Ullman, 2016). This is consistent with the evidence base, as highlighted in this review and the previous reviews, short-term training seems to result in increased knowledge (declarative/ techne) but not improvements in the procedural system (Herschell et al., 2010; Beidas & Kendall, 2010). This is also consistent with the hierarchical model adopted by Sharpless and Barber (2009), where declarative knowledge is acquired in the beginner stages but suggests that both practice and time are required for development of procedural skills and therefore achievement of competence. This is theoretically coherent, but this conclusion is hindered by limitations in the studies. It is difficult to get a clear sense of what the training and practice included, and these elements of training have not been adequately isolated and controlled for to make causal conclusions. For example, given the length of the training in these studies, without control groups it is hard to say practice was the element responsible for change, as changes could arguably be related to simple maturation effects (Lewis-Beck et al., 2004).

The findings of this review are consistent with previous reviews in this area (Beidas & Kendall, 2010; Herschell et al., 2010; Hill & Knox, 2013; Rakovshik &

McManus, 2010). There is indication that longer and multifaceted training courses do increase competencies. However, important methodological concerns highlighted previously remain and continue to make it difficult to have a confident view of how to effectively train therapists, for how long, what the most important training factors are and who is most likely to benefit. As it stands, the current literature serves little to inform the design and development of training courses. This highlights an increased urgency for high quality studies to be conducted in this area.

Limitations

There are also some limitations in the current review. Any studies published in other languages were excluded from the review; this means training studies from other cultures were potentially overlooked. Additionally, studies with positive results are more likely to be published in international journals (i.e. also typically published in English), which may mean there is a language bias in this review (Higgins & Green, 2011). Data synthesis in this review was narrative rather than meta-analytic. On balance, this was appropriate for several reasons; where there is heterogeneity in studies (as in this review) it is appropriate not to pool data (Rao et al., 2017), the small number of studies would prevent moderator analysis (Steel & Kammeyer-Mueller, 2002) and meta-analyses are negatively impacted by poor study quality (Egger et al., 1997). More high-quality research is required before meta-analytic strategy would be appropriate. The aim of this study was to review therapist competence, skills and abilities, therefore only this information was extracted from the studies. In doing so it is possible that important information providing greater context to the findings may have been missed. This is potentially particularly the case where studies also conducted qualitative analysis providing more in-depth participants' perspectives.

Only one researcher conducted a systematic literature search and screening of studies, there is evidence that this process can identify on average 9% more papers if conducted by two researchers (Edwards et al., 2002), this highlights the possibility that some eligible papers may have been missed in the current review. Additionally, whilst some studies were discussed with a secondary researcher, not all decisions about inclusion and exclusion were made collaboratively, therefore there could have been bias related to researcher subjectivity in application of eligibility criteria (Drucker et al., 2016). Furthermore, although reasons for exclusion were recorded at full-text screening stage, the specific papers excluded were not documented which reduces transparency of the study selection process.

The pre-published protocol indicated that the Downs & Black, (1998) appraisal tool would be used. This tool was tested with the papers in this review; however, the two reviewers subjectively felt this tool not was not well suited to the papers included in the review (case-series rather than experimental), this was reflected in reliability analysis (kappa=0.55; Appendix D). Therefore, the decision was made to utilise the Moga et al., (2012) tool. This was a viable alternative; however, this reflects a divergence from the pre-published protocol.

Future Research and Training Directions

1. There is a clear need for higher quality research to be conducted. It is imperative that this research includes appropriate control and comparison groups, sufficient sample sizes, and more long-term follow-up.
2. Whilst trainee therapists can usefully reflect on competency by using validated scales, this needs to be complemented with observer-rated assessments to avoid perceptual bias.

3. As there is confusion currently about which components of training are effective (and which are possibly redundant) the field would particularly benefit from using randomised dismantling approaches (e.g. Resick et al., 2008)
4. Training courses should be assessing therapist competence routinely; therefore, this is likely a huge resource of existing competency assessment outcome data which training courses should endeavour to publish. The establishment of practice-research networks between training providers is therefore indicated (e.g. Lucock et al., 2017). This type of collaboration would allow useful comparisons between similar training courses.
5. It is imperative that future research includes higher level of assessment of clinical skills (e.g. with real patients). It is also very important that higher order competence beyond those required in specific treatment modalities are assessed. This would have important implications, as if skills such as self-reflection, interpersonal skills and emotional intelligence cannot be reliably improved by training, then recruitment would have to place its focus in this area.
6. Lacking from the research is any attempt to address economic and cost analysis of training programmes. Given that government funding is frequently utilised to fund training places it is important that this information is better understood. Economic evaluations of training courses therefore need to be integrated with evaluations of their effectiveness.
7. None of the studies report ‘adverse events’ such as the rate of trainees that fail a component of a course or the entirety of the course. Reporting

of these fail rates would also be useful in considering why training does not enable competence in some cases.

Conclusion

In summary, twelve new studies were identified in this review. The findings were indicative that training is effective at increasing trainee competence. This evidence was strongest for full training programmes as opposed to specific training packages within a larger training programme where the evidence was more mixed. It is proposed that this could be related to the full training programmes incorporating practice with real clients and the longer time frame in which trainee competence was assessed, which may relate to the process of learning and consolidating procedural skills. However, these conclusions are presented with caution as there remain significant limitations with the evidence base. Notably, there is a lack of adequately controlled research, there is a lack of high-level assessment of global competencies and lack of adequately powered research. This leaves a huge amount of unanswered questions about factors that contribute (or do not) to competence development. This is consistent with problems identified in previous reviews and highlights an urgent need for more high-quality research to be conducted that can better understand the specific impacts of training on trainee competence.

References

American Psychological Association. (2012). *Revised competency benchmarks in professional psychology*. http://www.apa.org/ed/graduate/benchmarks-evaluation-system.aspx

Anderson, T., & Patterson, C. (2013). *Facilitative interpersonal skill task and rating method*. http://www.safranlab.net/uploads/7/6/4/6/7646935/fis_task_and_coding_manual_2013_10__1_.pdf

Aristotle. (1984). Nichomachean ethics. In J. Barnes (Ed.), *The complete works of Aristotle (in 2 volumes).* Princeton University Press.

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

Beale, Sarah, Liness, S., & Hirsch, C. R. (2020). Trainee self-assessment of cognitive behaviour therapy competence during and after training. *Cognitive Behaviour Therapist*, *13*, 1–12. https://doi.org/10.1017/S1754470X19000357

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 and Practice*, *17*(1), 1–30. https://doi.org/10.1111/j.1468-2850.2009.01187

Bennett-Levy, J. (2006). Therapist skills: A cognitive model of their acquisition

and refinement. *Behavioural and Cognitive Psychotherapy*, *34*(1), 57–78. https://doi.org/10.1017/S1352465805002420

Bennett-Levy, J. (2019). Why therapists should walk the talk: The theoretical and empirical case for personal practice in therapist training and professional development. *Journal of Behavior Therapy and Experimental Psychiatry*, *62*, 133–145. https://doi.org/10.1016/j.jbtep.2018.08.004

Binder, J. L. (1999). Issues in teaching and learning time-limited psychodynamic psychotherapy. *Clinical Psychology Review*, *19*(6), 705–719. https://doi.org/10.1016/S0272-7358(98)00078-6

Blackburn, I., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The Revised Cognitive Therapy Scale (CTS-R): Psychometric Properties. *Behavioural and Cognitive Psychotherapy*, *29(4)*, 431–446. https://doi.org/10.1017/S1352465801004040

Bohart, A. C., & Wade, A. G. (2013). The client in psychotherapy. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (6th ed., pp. 219–257). John Wiley & Sons Inc.

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

British Psychological Society & The Royal College of Psychiatrists. (2011). *Common mental health disorders the NICE guideline on Identification and pathways to care*. The British Psychological Society. https://www.nice.org.uk/guidance/cg123/evidence/cg123-common-mental-

health-disorders-full-guideline3

British Psychological Society. (2018, July). *Code of ethics and conduct.* The British Psychological Society https://www.bps.org.uk/sites/beta.bps.org.uk/files/Policy%20-%20Files/Code%20of%20Ethics%20and%20Conduct%20(2009).pdf

British Psychological Society. (2019, January). *Criteria for the accreditation of postgraduate training programmes in clinical psychology*. https://www.bps.org.uk/sites/bps.org.uk/files/Accreditation/Clinical Accreditation Handbook 2019.pdf

Chaddock, A., Thwaites, R., Bennett-Levy, J., & Freeston, M. H. (2014). Understanding individual differences in response to self-practice and self-teflection (SP/SR) during CBT training. The *Cognitive Behaviour Therapist*, *7(e14)*. https://doi.org/10.1017/S1754470X14000142

Chao, R. C.L., Paiko, L., Zhang, Y. S. D., & Zhao, C. (2017). Service-learning: A training method to enhance multicultural competence toward international students. *Scholarship of Teaching and Learning in Psychology*, *3*(1), 28–42. https://doi.org/http://dx.doi.org/10.1037/stl0000078

Clark, D. M. (2019). *IAPT at 10: Achievements and challenges*. National Health Service. https://www.england.nhs.uk/blog/iapt-at-10-achievements-and-challenges/

Downs, S., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality of health care interventions. *Journal of Epidemiogyl Community Health*, *52(6)*, 377–384. https://doi.org/10.1136/jech.52.6.377

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the age of the computer*. Free Press.

Drucker, A. M., Fleming, P., & Chan, A. (2016). Research techniques made simple: Assessing risk of bias in systematic reviews. *Journal of Investigative Dermatology*, *136*(11), e109–e114. https://doi.org/doi.org/10.1016/j.jid.2016.08.021

Edwards, P., Clarke, M., Diguiseppi, C., Pratap, S., Roberts, I., & Wentz, R. (2002). Identification of randomized controlled trials in systematic reviews: Accuracy and reliability of screening records. *Statistics in Medicine*, *21 (11)*, 1635–1640. https://doi.org/10.1002/sim.1190

Egger, M., Smith, G. D., Schneider, M., Minder, C., Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

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

Ginzburg, D. M., Bohn, C., Hofling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy*, *50*(12), 747–752. https://doi.org/10.1016/j.brat.2012.09.001

Hardy, G. E., Shapiro, D. A., Stiles, W. B., & Barkham, M. (1998). When and why does cognitive-behavioural treatment appear more effective than psychodynamic-interpersonal treatment? Discussion of the findings from

the second Sheffield psychotherapy project. *Journal of Mental Health*, *7*(2), 179–190. https://doi.org/10.1080/09638239818229

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review*, *30*(4), 448–466. https://doi.org/10.1016/j.cpr.2010.02.005

Higgins, J. P., & Green, S. (2011). Addressing reporting biases. In J. A. . Sterne, M. Egger, & D. Moher (Eds.), *Cochrane handbook for systematic reviews of intervention. Version 5.1.0*. The Cochrane Collaboration. www.handbook.cochrane.org

Hill, C. E., Baumann, E., Shafran, N., Gupta, S., Morrison, A., Rojas, A. E. P., Spangler, P. T., Griffin, S., Pappa, L., & Gelso, C. J. (2015). Is training effective? A study of counseling psychology doctoral trainees in a psychodynamic/ interpersonal training clinic. *Journal of Counseling Psychology*, *62*(2), 184–201. https://doi.org/10.1037/cou0000053

Hill, C. E., & Knox, S. (2013). Training and Supervision in Psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (6th Ed., pp. 775–811). John Wiley & Sons Inc.

Hill, C. E., Stahl, J., & Roffman, M. (2007). Training novice psychotherapists: Helping skills and beyond. *Psychotherapy*, *44*(4), 364–370. https://doi.org/10.1037/0033-3204.44.4.364

Hume, F. A. (2017). *Evaluating the effects of self-practice/self-reflection on cognitive flexibility, empathy, insight, self-compassion, self-monitoring, and stress in postgraduate cognitive behaviour therapy trainees* [Dissertation].

Massey University. https://mro.massey.ac.nz/handle/10179/12492.

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

Keng, S. L., Waddington, E., Lin, X. B., Tan, M. S. Q., Henn-Haase, C., & Kanter, J. W. (2017). Effects of functional analytic psychotherapy therapist training on therapist factors among therapist trainees in Singapore: A randomized controlled trial. *Clinical Psychology and Psychotherapy*, *24*(4), 1014–1027. https://doi.org/10.1002/cpp.2064

Koddebusch, C., & Hermann, C. (2018). A proposed conceptualization of therapeutic competence: the three level model. *Research in Psychotherapy: Psychopathology, Process and Outcome*, *21(1)*, 12–23. https://doi.org/10.4081/ripppo.2018.286

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M J Lambert (Ed.), *Handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). John Wiley & Sons Inc.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lewis-Beck, M., Bryman, A., & Futing Liao, T. (2004). *The SAGE encyclopedia of social science research methods*. Routledge.

Liness, S., Beale, S., Lea, S., Byrne, S., Hirsch, C. R., & Clark, D. M. (2019a). Multi-professional IAPT CBT training: Clinical competence and patient outcomes. *Behavioural and Cognitive Psychotherapy*, *47*(6), 672–685. https://doi.org/10.1017/S1352465819000201

Liness, S., Beale, S., Lea, S., Byrne, S., Hirsch, C. R., & Clark, D. M. (2019b). The sustained effects of CBT training on therapist competence and patient outcomes. *Cognitive Therapy and Research*, *43*(3), 631–641. https://doi.org/10.1007/s10608-018-9987-5

Lucock, M., Barkham, M., Donohoe, G., Kellett, S., Mcmillan, D., Mullaney, S., Sainty, A., Saxon, D., Thwaites, R., & Delgadillo, J. (2017). The role of practice research networks (PRN) in the development and implementation of evidence: The northern improving access to psychological therapies PRN case study. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(6), 919–931. https://doi.org/10.1007/s10488-017-0810-5

Miller, G. E. (1990). The Assessment of Clinical Skills/ Competence/ Performance. *Academic Medicine*, *65*(9). 63-37. https://journals.lww.com/academicmedicine/Abstract/1990/09000/The_assessment_of_clinical.45.aspx

Moga, C., Guo, B., Schopflocher, D., & Harstall, C. (2012, March). *Development of a quality appraisal tool for case ceries studies using a modified delphi technique*. Insitute of Health Economics. https://www.ihe.ca/advanced-search/development-of-a-quality-appraisal-tool-for-case-series-studies-using-a-modified-delphi-technique

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal*, *339*(j7716), b2535–b2535. https://doi.org/10.1136/bmj.b2535

Murray, H. (2017). Evaluation of a trauma-focused CBT training programme for IAPT services. *Behavioural and Cognitive Psychotherapy*, *45*(5), 467–482. https://doi.org/10.1017/S1352465816000606

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, *33*(3), 484–499. https://doi.org/10.1016/j.cpr.2013.01.010

Nagy, T. F. (2005). Competence. *Journal of Aggression, Maltreatment & Trauma*, *11*(1), 27–49. https://doi.org/10.1300/J146v11n01_03

National Collaborating Centre for Mental Health. (2019). *The Improving Access to Psychological Therapies Manual.* NHS Digital. https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual.pdf

National Institute for Clinical Excellence. (2012). Algorithm for classifying quantitative (experimental and observational) study designs. Methods for the Development of NICE Public Health Guidance. https://www.nice.org.uk/process/pmg4/chapter/appendix-e-algorithm-for-classifying-quantitative-experimental-and-observational-study-designs

Pakenham, K. I. (2015). Effects of acceptance and commitment therapy (ACT) training on clinical psychology trainee stress, therapist skills and attributes, and ACT processes. *Clinical Psychology and Psychotherapy*, *22*(6), 647–655. https://doi.org/10.1002/cpp.1924

Pearson, A., & Weinberg, A. (2017). The impact of counsellor training on emotional intelligence. *British Journal of Guidance and Counselling*, *45*(5), 610–621. https://doi.org/10.1080/03069885.2016.1226496

Rakovshik, S. G., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. *Clinical Psychology Review*, *30*(5), 496–516. https://doi.org/10.1016/j.cpr.2010.03.004

Rantanen, A. P., & Soini, H. S. (2018). Changes in counsellor trainee responses to client's message after peer group consultation (PGC) training. *British Journal of Guidance and Counselling*, *46*(5), 531–542. https://doi.org/10.1080/03069885.2016.1277381

Rao, G., Lopez-Jiminex, F., Boyd, J., D'Amico, F., Durant, N. H., Hlatky, M. A., Howard, G., Kirley, K., Masi, C., Powell-Wiley, T. F., Solomonides, A. E., West, C. P., & Wessel, J. (2017). Methodological standards for meta-analyses and qualitative systematic reviews of cardiac prevention and treatment studies. *Circulation*, *136*(10), 172-194. https://doi.org/10.1161/CIR.0000000000000523

Resick, P. A., Galovski, T. E., Uhlmansiek, M. O. B., Scher, C. D., & Clum, G. A. (2008). A randomized clinical trial to dismantle components of cognitive processing therapy for posttraumatic stress disorder in female victims of interpersonal violence. *Journal of Consulting and Clinical Psychology. 76*(2), 243–258. https://doi.org/10.1037/0022-006X.76.2.243

Roth, A. D, & Pilling, S. (2007, January). *The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders: Best practise guidance.* Department of Health. https://www.uea.ac.uk/documents/246046/11919343/CBT_Competence_List.pdf/fcb5b875-67da-4f85-966c-a01506dacf60

Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, *36*(2), 129–147. https://doi.org/10.1017/S1352465808004141

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, *7(16)*. https://doi.org/10.1186/1472-6947-7-16

Sharpless, B. A., & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review*, *29*(1), 47–56. https://doi.org/10.1016/j.cpr.2008.09.008

Steel, P. D., & Kammeyer-mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Applied Psychology. 87*(1), 96–111. https://doi.org/10.1037//0021-9010.87.1.96

Tweed, A., Graber, R., & Wang, M. (2010). Assessing trainee clinical psychologists' clinical competence. *Psychology Learning and Teaching Teaching*, *9*(2), 50–60. https://doi.org/10.2304/plat.2010.9.2.63

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of langauge learning, knowledge and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language*. Elsevier Inc.

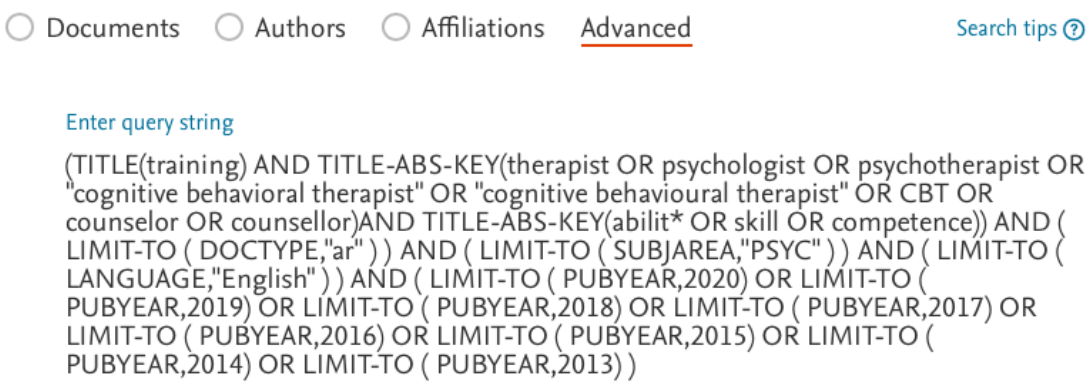
Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review.

Journal of Consulting and Clinical Psychology, *7*(2), 200–211. https://doi.org/10.1037/a0018912

Zarafonitis-Müller, S., Kuhr, K., & Bechdolf, A. (2014). The relationship between therapist's competence and adherence to outcome in cognitive-behavioural therapy: Results of a meta-analysis. *Fortschritte Der Neurologie Psychiatrie*, *82*(9), 502–510. https://doi.org/10.1055/s-0034-1366798

Appendices

Appendix A: Example Search Terms Used in SCOPUS



Appendix B: Quality Appraisal Tool

Moga et al., (2012)

Study objective

1. Is the hypothesis/aim/objective of the study clearly stated in the abstract, introduction, or methods section?

Yes: The hypothesis/aim/objective of the study is clearly stated in the abstract, introduction, or methods section.
No: The hypothesis/aim/objective is not provided in the abstract, introduction, or methods section.

Study population

2. Are the characteristics of the participants included in the study described?

Yes: The most relevant characteristics are presented. The authors should report the total number, age, and gender distribution of the participants. Ethnicity, severity of disease/condition, comorbidity, or etiology should also be included, if relevant.
No: The most relevant characteristics of the participants are not reported. If only the number of participants was reported or any of the relevant characteristics is missing, the question should be answered no.
Note: Assessor(s) should decide which aspects are important before using the checklist.

3. Were the cases collected in more than one centre?

Yes: Cases are collected in more than one centre (multicentre study).
No: Cases are collected from one centre, or it is unclear where patients came from.

4. Are the eligibility criteria (inclusion and exclusion criteria) to entry the study explicit and appropriate?

Yes: The eligibility criteria are clearly stated and replicable, and match the objective of the study.
No: The eligibility criteria are not clearly stated.
Note: Assessor(s) should decide which aspects are important before using the checklist.

5. Were participants recruited consecutively?

Yes: There is a clear statement that the participants are recruited consecutively.
No: The participants were recruited based on other criteria, such as access to intervention determined by the distance or availability of resources. The method used to recruit participants is not clearly stated.

6. Did participants enter the study at a similar point in the disease?

Yes: There is a clear description about the clinical status of participants, duration of condition (exposure) before the intervention, comorbidity, severity, or complications of all participants in the study.
No: There is no description about whether participants entered the study at a similar point in the disease. Participants did not enter the study at a similar point in the disease, as revealed by a wide range of disease duration before entering the study or different comorbidities or complications due to progression of their condition/disease.
Note: Assessor(s) should decide which aspects are important before using the checklist.

Intervention and co-intervention

7. Was the intervention clearly described in the study?

Yes: There is a detailed description about the characteristics of the intervention (e.g. dosage, frequency of administration, duration, permanent or temporary intervention, and technical parameters/characteristics of a device).
No: The intervention is only mentioned by name without any details, the information provided is unclear, or important parameters of the intervention are missing from the presentation.
Note: Assessor(s) should decide which aspects are important before using the checklist.

8. Were additional interventions (co-interventions) clearly reported in the study?

Yes: The name or type of co-intervention is acknowledged in the study. The question should be answered yes if it is obvious (based on study context) that co-interventions were unnecessary.

No: Co-intervention(s) are not reported, or the name(s) or type(s) of co-intervention(s) are unclear.
Note: Assessor(s) should decide which aspects are important before using the checklist.

Outcome measures

9. Are the outcome measures clearly defined in the introduction or methods section?

Yes: All relevant (primary and secondary) outcomes that match the objective(s) of the study are described in the introduction or methods section (e.g. accomplished, measurable improvements or effects, symptoms relieved, improved function, improved test scores, and quality of life measures).
No: The outcomes are reported for the first time in the results or conclusion section of the study.
The relevant outcomes are briefly mentioned without any details in the results, discussion, or conclusion section(s).
The outcomes reported are not relevant to study objective(s).
Note: Assessor(s) should decide which aspects are important before using the checklist.

10. Were relevant outcomes appropriately measured with objective and/or subjective methods?

Yes: Appropriate methods used to measure the outcomes are described in the methods section. These measures might be objective (e.g. gold standard tests or standardized clinical tests), and/or subjective (e.g. self-administered questionnaires, standardized forms, or patient symptoms interview forms).
No: No details are provided on the objective or subjective methods used to measure study's outcomes.

11. Were outcomes measured before and after intervention?

Yes: The relevant outcomes are measured before and after applying the intervention.
No: The outcomes are measured only after applying the intervention.

Statistical analysis

12. Were the statistical tests used to assess the relevant outcomes appropriate?

Yes: The statistical tests are clearly described in the methods section and are used appropriately (e.g. parametric test for normally distributed population vs. nonparametric test for non-Gaussian population).
No: The statistical tests used to assess the relevant outcomes are inappropriate. From the information available it is unclear the distribution of the population from which the participants at the study were selected.

Results and conclusions

13. Was the length of follow-up reported?

Yes: The length of follow-up is clearly reported.
No: The length of follow-up is not reported, or the duration of the study is unclear.

14. Was the loss to follow-up reported?

Yes: The number or proportion of patients lost to follow-up is reported.
No: The number or proportion of patients lost to follow-up is not reported.

15. Does the study provide estimates of the random variability in the data analysis of relevant outcomes?

Yes: The study reports estimates of the random variability (e.g. standard error, standard deviation, confidence intervals) for all relevant primary and secondary outcomes.
No: Estimates of the random variability are not reported for all relevant outcomes. The presentation of the random variability is unclear (e.g. measure of dispersion reported without indicating if it is standard deviation or standard error).

16. Are adverse events reported?

Yes: The undesirable or unwanted consequences of the intervention during the study period or within a prespecified time period are reported. Absence of any adverse event(s) is acknowledged in the study.
No: There is no statement about the presence or absence of adverse events.

17. Are the conclusions of the study supported by results?

Yes: The main conclusions of the study are supported by the evidence presented in the results section.

No: The conclusions are not supported by the evidence presented in the results section.

Competing interest and source of support

18. Are both competing interest and source of support for the study reported?

Yes: Both competing interest and source of support (financial or other) received for the study are reported, or the absence of any competing interest and source of support is acknowledged.
No: Either there is no information available about competing interests and sources of support, or only one of these elements is reported.

Appendix C: Tables of Study Quality Analysis

Primary Researcher

Authors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
Beale et al., (2020)	1	1	0	1	1	-	1	1	1	1	1	1	1	1	1	0	1	1	15
Chaddock et al., (2014)	1	0	0	1	1	-	1	0	1	0	1	0	0	0	0	0	1	1	8
Chao et al., (2017)	1	1	0	0	0	-	1	0	1	1	1	0	1	0	1	0	0	0	8
Hill et al., (2015)	1	1	0	0	1	-	0	0	1	1	1	1	0	0	1	0	1	0	9
Hume, (2017)	1	1	0	1	1	-	1	0	1	1	1	0	1	1	1	0	1	0	12
Keng et al., (2017)	1	1	0	1	1	-	1	0	1	1	1	1	1	1	1	0	1	0	13
Liness et al., (2019a)	1	1	0	1	1	-	1	1	1	1	1	1	0	0	1	0	1	1	13
Liness et al., (2019b)	1	1	0	1	1	-	1	1	1	1	1	1	1	1	1	0	1	1	15
Murray, (2017)	1	0	1	0	0	-	1	1	1	1	1	1	1	1	1	0	1	1	13
Pakenham, (2015)	1	1	0	0	0	-	1	0	1	1	1	1	0	0	1	0	1	0	9
Pearson and Weinberg (2017)	1	1	0	0	1	-	0	0	1	1	1	1	0	0	1	0	1	1	10
Rantanen and Soini (2018)	1	0	0	0	1	-	1	0	1	1	1	1	0	0	1	0	1	1	10

Secondary Researcher

Authors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
Beale et al., (2020)	1	1	0	1	0	-	1	0	1	1	1	1	0	1	1	0	1	1	12
Chaddock et al., (2014)	1	0	0	0	1	-	1	0	1	1	1	0	0	0	0	0	1	1	8
Chao et al., (2017)	1	1	0	0	0	-	1	0	1	1	1	1	0	0	0	0	1	0	8
Hill et al., (2015)	1	1	0	0	1	-	1	0	1	1	1	1	0	1	1	0	1	0	11
Hume, (2017)	1	1	0	1	1	-	1	0	1	1	1	0	1	0	0	0	1	0	10
Keng et al., (2017)	1	1	0	1	1	-	1	0	1	1	1	1	1	1	1	0	1	0	13
Liness et al., (2019a)	1	1	0	1	1	-	1	1	1	1	1	1	1	1	1	0	1	1	15
Liness et al., (2019b)	1	1	0	1	1	-	1	1	1	1	1	1	1	1	1	0	1	1	15
Murray, (2017)	1	0	1	0	1	-	1	0	1	0	1	1	1	1	1	0	1	0	11
Pakenham, (2015)	1	1	0	0	1	-	1	0	1	0	1	1	1	1	1	0	0	0	10
Pearson and Weinberg (2017)	1	1	0	0	1	-	1	0	1	1	1	1	1	1	0	0	1	1	12
Rantanen and Soini (2018)	1	0	0	0	1	-	1	0	1	0	1	1	1	0	1	0	0	1	9

Appendix D: Kappa Analysis

Moga et al., (2012)

IHER1 * IHER2 Crosstabulation

Count

		IHER2		
		.00	1.00	Total
IHER1	.00	54	15	69
	1.00	16	118	134
Total		70	133	203

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.661	.056	9.417	.000
N of Valid Cases		203			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Downs and Black (1998)

DABR1 * DABR2 Crosstabulation

Count

		DABR2		
		.00	1.00	Total
DABR1	.00	129	31	160
	1.00	42	120	162
Total		171	151	322

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.547	.047	9.834	.000
N of Valid Cases		322			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Section Two: Empirical Study

Therapist Interpersonal Functioning and Self-reflection Skills as Predictors of Patient Treatment Outcomes in IAPT Services

Abstract

Objectives: Therapist effects on patients' outcomes are well established. However, the understanding of how and why some therapists are more effective than others is less well developed. The current study aimed to explore whether therapist self-reported self-reflection and interpersonal functioning, and supervisor reported self-reflection was associated with patient outcomes. **Design:** Data was collected from therapists and their supervisors cross-sectionally utilising an electronic survey. Patient data was collected retrospectively from electronic record systems from two NHS Trust Improving Access to Psychological Therapies (IAPT) services. **Methods:** Therapists (N=61) and supervisors (N=19) were recruited. Therapists completed the Groningen Reflective Ability Scale (GRAS), the Social Skills Inventory and the Inventory of Interpersonal Problems. Supervisors completed an adapted version of the GRAS about their supervisees. Patient outcome data (depression; PHQ-9 and anxiety; GAD-7) was available for N=3112 patients. This provided outcome data for N=42 therapists, N=18 of which also had supervisor reported information. Multilevel modelling assessed the relationships between survey results and patient outcomes. **Results:** Therapists accounted for 2.69% (PHQ-9) and 2.78% (GAD-7) of variance in treatment outcomes. Social skills significantly predicted treatment outcomes for depression, but not anxiety. Self-reflection and interpersonal problems were not associated with treatment outcomes. Sensitivity analysis suggested that important outlying variables may undermine relationships between self-reported self-reflection and patient outcomes. There were too few corresponding supervisor-reported data to conduct meaningful analysis. **Conclusions:** Therapist effects may in part be explained by the therapists' reported social skills.

Keywords: “therapist effects” “interpersonal skills” “social skills” “interpersonal problems” “self-reflection” “depression” “anxiety”

Practitioner points

- Therapist social skills may partly explain variability in patient outcome.
- If findings are replicated, then therapist characteristics could be used as targets for training or recruitment

Introduction

Therapist effects, the influence that therapists have on variability in patient outcomes, have been documented and evidenced in a large number of psychotherapy studies. Crits-Christoph et al., (1991) conducted a meta-analysis and found that therapist effects had a pooled medium effect size accounting for 8.6% of variance in patient outcomes. In a recent UK-based study, Green et al., (2014) found that therapist effects accounted for 9% of variability in treatment outcomes. Patients treated by the most effective therapists attained reliable recovery rates of 46.4% for depression and 58.3% for anxiety symptoms, which were twice as large as the reliable recovery rates observed in patients treated by the least effective therapists (19% for depression and 25.7% for anxiety). Similar findings have been observed across different treatment contexts and treatment modalities; as asserted by Nissen-Lie, Monsen, and Rønnestad (2010) it seems that “some therapists win and some do not, independent of the therapeutic method they use”. More recent systematic reviews have supported this, suggesting that therapist effects account for approximately 5% of variability in treatment outcomes (Baldwin & Imel, 2013; Johns et al., 2019). That therapist effects exist can therefore be considered a reliable and robust phenomenon (Johns et al., 2019). Comparatively, this effect is smaller than other factors influencing outcome, for example some estimates suggest that extra-therapeutic factors (e.g. patient personality/ life circumstances) account for around 40% of variance in patient outcomes (Lambert & Barley, 2001; Lambert, 2013)

Whilst the therapist effect phenomenon is therefore evident, there remains limited consensus on understanding why these differences between therapists occur. The existent literature suggests that demographic factors, such as age, gender, professional identity and experience do not seem to predict patient

outcomes (Beutler et al., 2004; Goldberg et al., 2016). The emerging literature has therefore explored therapist characteristics as better potential explanatory variables. Factors such as resilience, mindfulness, organisational skills and reflective functioning have been evidenced (e.g. Cologon, Schweitzer, King, & Nolte, 2017; Green, et al., 2014; Pereira, Barkham, Kellett, & Saxon, 2017; Ryan, Safran, Doran, & Muran, 2012). This fits with a model of therapist skills in which there is differentiation between technical/ conceptual skills and procedural and reflective skills (Bennett-Levy, 2006). In particular, highly refined interpersonal and self-reflection skills have been suggested to be characteristic of highly effective therapists (Bennett-Levy, 2019; Wampold et al., 2017).

Interpersonal Skills

An interpersonally skilled individual is skilled in receiving and accurately integrating and interpreting emotional and social information from others and the self, and also is skilful in the communication and regulation of emotional and social information relevant to the social context (Riggio & Carney, 2003). This description could be usefully applied to the role of a therapist, whose job involves striving to understand, empathise, and communicate effectively with their patients, who also present with a range of emotional disorders and socio-economic contexts. Interpersonal skills likely present an interface through which therapist declarative knowledge can be integrated and communicated with the client (Bennet-Levy, 2006). It is perhaps then unsurprising that the literature is indicative of interpersonal functioning as a potential predictor of treatment effectiveness.

The emerging literature is indicative that interpersonal skills are likely to be important in enabling better patient outcomes. One way this has been investigated has been with the facilitative interpersonal skills (FIS) task in which

participants are provided with video-vignettes of difficult therapeutic scenarios and asked how they would respond. Responses are video-recorded and rated by researchers using a structured rating scheme (Anderson & Patterson, 2013). In a series of studies, Anderson and colleagues found that FIS ratings are predictive of patient outcomes (Anderson, Crowley, et al, 2016; Anderson, McClintock, et al., 2016; Anderson et al., 2009). Similarly, Schöttke, et al., (2017) found that observer-rated interpersonal skills were predictive of therapeutic outcomes when participants were observed during a group task. However, observer ratings were not predictive of patient outcomes following a one-one interview.

Heinonen and Nissen-Lie (2020) suggest that therapists' reports of their social skills are not relevant to the prediction of treatment outcomes. However, one study they cite (Bambling & King, 2013) was not primarily looking at social skills as a predictor of outcome and actually allocated one patient per therapist in an attempt to minimise therapist effects. Additionally, the three studies investigating the FIS (described above), provide inconsistent findings for self-reported social skills using the Social Skill Inventory (SSI; Riggio, 1986). Whilst two studies do suggest FIS rather than SSI is associated with treatment outcomes (Anderson et al., 2009; Anderson, McClintock, 2016), this was contradicted by Anderson, Crowley et al., (2016). They utilised the SSI alongside the FIS to screen whether participants were placed in a high or low interpersonal skill group. Whilst this found those in the high skilled group achieved better outcomes, subsequent analysis with observer and self-reported social skills as continuous variables found that self-reported social skills were significantly associated with treatment outcomes, whilst the observer report was not. In addition, other studies are indicative that self-reported skill can predict outcomes. Rieck and Callahan (2013) for example, found that emotional intelligence (a key component of social

skills; Riggio & Carney, 2003) was predictive of treatment outcome, particularly when considered with neuroticism as a personality trait. Heinonen et al., (2012) also found that self-rated relational skills were predictive of greater symptom decrease in short term but not long-term psychotherapies.

Therefore, whilst the literature for self-reported social skills does not conclusively assert this as a predictor of treatment outcome, it could be argued that this is due to poor study design and lack of application of appropriate modelling techniques. This is important to clarify, as increasingly therapist effect research utilises multilevel modelling analyses requiring large sample sizes at the patient and therapist level (Schiefele et al., 2017). Typically, this is only available in naturalistic datasets, where it may be more difficult to implement tasks such as the FIS with sufficient number of therapists.

Self-reflection

Self-reflection is the intentional examination and judgement of one's experiences, feelings, action and thoughts (Carver & Scheier, 1998). This process enables self-monitoring and can provide feedback which can be used to adapt, develop and respond in different ways in the future (Grant et al., 2002). It is similar to, but not the same as rumination, which is typically negative, self-perpetuating and based on failure experiences (Nolen-Hoeksema et al., 2008). Whilst they can co-occur and are both forms of self-focussed attention, self-reflection is an adaptive process, whereas rumination is not (Takano & Tanno, 2009). Bennett-Levy (2006) proposes a model of therapist development in which self-reflection plays a central role in the integration of a therapist's personal and therapeutic self. It has been suggested that self-reflection is particularly important when working with clients who present with psychological complexity (Bennett-Levy, 2019; Muran, Safran, Eubanks, & Gorman, 2018; Ridley, Kelly, & Mollen,

2011) and is also seen as central to the meta competencies of therapists (American Psychological Association, 2012; British Psychological Society, 2019).

A number of studies have investigated concepts similar to self-refection. Meichenbaum and Lilienfeld (2018) assert the role of “healthy self-doubt”, defined as involving self-reflection on one’s biases, limitations and techniques increasing awareness of the self. This is supported by the literature; for example, one study illustrated that self-doubt predicted better treatment outcomes, whilst negative personal reactions predicted worse outcomes (Nissen-Lie et al., 2013). A second study similarly illustrated self-doubt predicted better treatment outcomes, but only with high self-affiliation (Nissen-Lie et al., 2017). Both of these studies likely reflect a pronounced level of self-reflection in effective therapists, and it is suggested that this may be related to these therapists displaying an ability to critically evaluate their own practice facilitating adaptation in their approach.

Aims of the Current Study

Although previous studies have considered aspects of therapists’ reflective capacity, self-reflection has not been directly assessed. Additionally, whether therapist reports of their own interpersonal skills are associated with treatment outcomes is unclear from the literature. Therefore, the current study aimed to assess both social skills and reflective abilities and their relative contribution to patient outcome. Given the potential for bias in any self-assessment, this study also sought clinical supervisors’ assessment of therapist self-reflective abilities. Including the perception of clinical supervisors has been infrequent in the therapist effects literature. Furthermore, interpersonal skills will be assessed in more detail both in terms of both skills and problems. Therefore, the central aim of the current study is to establish if supervisor-reported and

therapist self-reported self-reflective skills and self-reported interpersonal functioning are associated with therapists' variability in treatment outcomes.

Hypotheses

1. Therapist effects will significantly account for a proportion of variability in treatment outcomes.
2. Therapists' self-reported self-reflective skills will be significantly associated with treatment outcomes.
3. Therapists' supervisor rated reflective skills will be significantly associated with treatment outcomes.
4. Therapists' self-reported interpersonal functioning will be significantly associated with treatment outcomes.
5. There will be a significant association between supervisor and therapist-rated abilities.

Methods

Design

The current study collected data from patients, from therapists and their supervisors. Therapist and supervisor data were collected cross-sectionally, utilising an electronic survey. Patient treatment outcomes data were collected retrospectively and sourced from electronic clinical record systems. This methodology is consistent with previous studies investigating therapist effects (e.g. Green et al., 2014; Pereira et al., 2017).

Sampling

A sample size estimate was made based on Maas and Hox (2005) simulation study which suggested that 50 participants were required at the

therapist level. Number of patients per therapist was calculated using criterion proposed by Schiefele et al., (2017) with an assumed effect size of 5% (based on Baldwin & Imel, 2013); this suggested 15 patients per therapist was required. Therefore, an estimated sample size for the current study was 50 therapists with 15 patients per therapist, yielding a total sample size of 750 patients.

Setting and Interventions

Lutz and Barkham (2015) recommend conducting therapist effects research in naturalistic settings in order to yield sufficiently large and adequately powered samples. This is consistent with Johns et al., (2019) who call for therapists' effects research to aim for greater sample sizes at each of the levels. Therefore, this study was based in Improving Access to Psychological Therapies (IAPT) services. IAPT services provide a suitable avenue for collecting data, particularly as client outcome data is collected in standard practice and has the potential to contribute a large number of therapists. In addition, within IAPT services it is expected that therapists receive a minimum of one-hour clinical and case-management supervision per week (Turpin & Wheeler, 2011). IAPT supervisors are therefore well placed to offer opinions about their supervisees.

IAPT is structured in a stepped-care model, in which patients can be 'stepped-up' from each level in order to access increasingly lengthier and costly treatments if necessary. Previous research has illustrated therapist effects at step 2 of IAPT services (Green et al., 2014) which involves brief (<8 sessions) guided self-help interventions based on CBT principles and at step 3 (Delgadillo et al., 2018; Saxon et al., 2017), which encompasses a wider variety of treatment options and interventions which might be typically imagined of psychotherapy, such as counselling for depression, interpersonal psychodynamic psychotherapy

and cognitive behavioural therapy. This study included therapists from both steps 2 and 3.

Recruitment

Therapists were recruited from four IAPT services in the North of England across two NHS trusts. All therapists and supervisors were eligible and asked to participate in the research. A total of 168 therapists were invited to the research and 61 completed at least one of the three measures (36.3% of those invited). A total of 51 supervisors were invited and 19 participated in the research (37.3% of those invited). Therapist and supervisor characteristics are summarised in Table 1.

Patient outcomes data were requested for all participating therapists for patients discharged within an 18-month period. Patient characteristics are summarised in Table 2. To be included in the study patients must have 1) had a recorded pre and post score on either Patient Health Questionnaire- 9 item scale (PHQ9) or the Generalised Anxiety Disorder- 7 item scale (GAD7; as these measures are completed retrospectively the first measure reflects the patient's pre-treatment symptoms and at least one additional measure is required to capture within treatment changes); 2) received face-to-face one-to-one therapy, and 3) attended at least two sessions with the participating therapist. In Figure 1 is a flow diagram of the process of patient selection which illustrates how the final sample for multilevel modelling was arrived at[3]. After applying the inclusion/ exclusion criteria, the final sample included 42 therapists, 13 supervisors and 3112 patients (ranging from 7 to 213 per therapist with a mean of 74 patients per therapist).

[3] Patient outcome data was only available for 3 out of 4 NHS services in time to include in this study. This was due to delays associated with COVID-19.

Figure 1

*Adapted CONSORT Flow Diagram Illustrating Sample Selection Process (*Schulz, Altman & Moher*, 2010)*

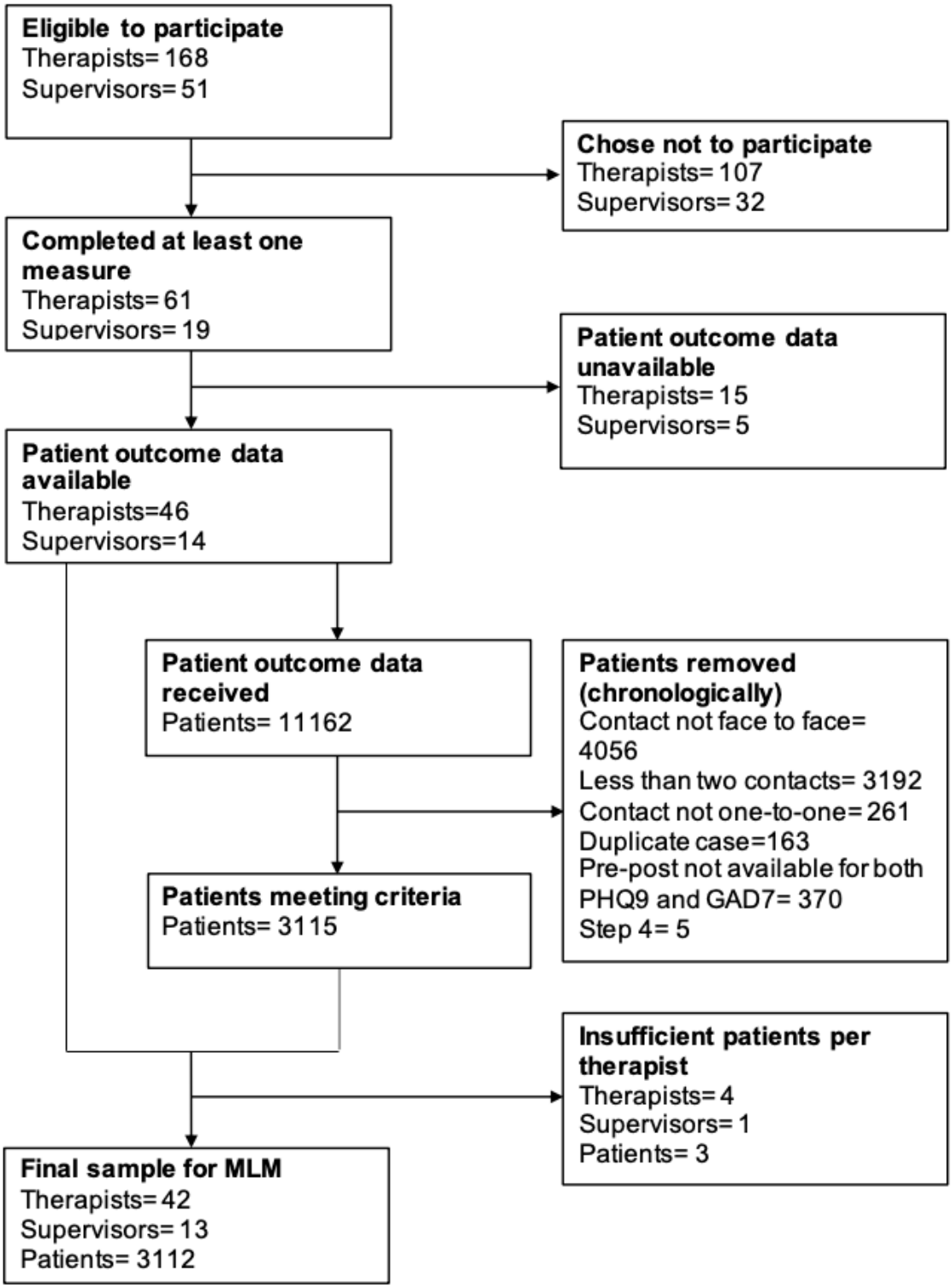


Table 1

Therapist and Supervisor Characteristics

	Full sample (SD;Range)[a]	Included in MLM (SD;Range)[a]
Therapist	N=61	N=42
Age M	43.05 (11.65;24-70)	42.33 (12.40; 24-70)
Females	N=54	N=40
Professional Role:		
CBT	N=30	N=17
PWP	N=18	N= 16
Counsellor	N=13	N= 9
Counselling Psychologist	N= 1	N=0
Clinical experience (years)	7.43 (6.42;0-24)	7.07 (5.97; 0-24)
Supervision received per week (hours)	0.83 (0.66;0-3.7)	0.86 (0.54;0-3)
Reflective Ability[b]	97.85 (6.09; 86-114)	98.33 (6.20; 86-114)
Self-reflection	42.98 (2.86; 36-49)	42.98 (3.02; 36-49)
Empathetic Reflection	26.72 (2.12; 21-30)	26.95 (2.08; 21-30)
Reflective Communication	28.15 (2.84; 22-35)	28.40 (2.49; 22-35)
Interpersonal Problems[c]	40.48 (16.72;6-76)	39.43 (17.33; 6-76)
Interpersonal Skills[d]	281.77 (27.22; 226-341)	285.82 (27.31; 227-341)
Supervisor	N= 19	N=13
Age M (SD;Range)	41.42 (8.04;25-53)	38.92(8.47;25-53
Females	N=17	N=12
Professional Role		
CBT	N=11	N=9
Counsellor	N=2	N=1
PWP	N=6	N=3
Clinical experience (years)	7.95 (3.48;1.2-13.7)	8.08 (3.28; 4-17)
Number of supervisees	3.79 (2.74; 1-11)	4.08 (2.93;1-11)
Supervision Given per week (hours)	3.15 (2.73;0.4-11.4)	2.99 (2.79; 0.4-11.4)
	45 therapists[e]	18 therapists
Therapist Reflective Ability[b]	94.29 (12.3;67-112)	91.72 (12.63; 67-112)

[a]Unless otherwise stated [b]GRAS [c] IIP-32 [d] SSI [e]Only 24 of these therapists participated in the study

Measures

Supervisor/ Therapist Data (see Appendix E)

Groningen Reflection Ability Scale (GRAS). The GRAS consists of 23 items scored on a scale between 0-5, designed to assess self-reported capacity to personally reflect. The measure has good construct validity and

Table 2

Patient Demographic and Clinical Information

	Mean (S.D; Range)[a]
Patients	N=3112
Age	40.25 (15.42;16-97)
Female	N= 2025 (66.0%)
Pre PHQ9	14.52 (5.92; 0-27)
Post PHQ9	9.32 (6.95; 0-27
Pre GAD7	13.48 (5.05; 0-21)
Post GAD7	8.63 (6.18; 0-21)
Number of therapy sessions:	
Total Sample	6.51 (4.81; 2-35)
Step 2	4.33 (2.46; 2-24)
Step 3	8.80 (5.53; 2-35)

[a] Unless otherwise stated

internal reliability (Cronbach's alpha=0.83 and 0.74). Factor analysis suggested the GRAS measures a single construct, but has three related subscales of self-reflection, empathetic reflection and reflective communication (Aukes et al., 2007). Authors indicate that use of subscales may only be used for research purposes (see in Appendix E). The wording of the GRAS was altered to reflect the supervisor completion context. The GRAS displayed acceptable internal reliability in the current sample (Therapist Reported Alpha=0.74 and Supervisor Reported Alpha= 0.94). Subscales were included in analysis, however, Alpha's for these scales suggest that they were not robust (Self-reflection Alpha= 0.56, Empathetic Reflection Alpha= 0.52 and Reflective Communication Alpha= 0.36)[4].

Social Skills Inventory (SSI). The SSI has 90 items scored on a 1-5 Likert scale and was designed to assess interpersonal skills. The SSI has two primary subscales, (social and emotional) which are further separated into six subscales (Social Control, Social Expressivity, Social Sensitivity, Emotional Control, Emotional Expressivity and Emotional Sensitivity). Higher scores indicate higher

[4] All Cronbach's Alpha outputs can be seen in Appendix F

interpersonal functioning. The SSI has good test re-test reliability (ranging from 0.81-0.96; Riggio, 1989), good convergent and discriminant validity (Riggio, 1986) and good internal consistency (Cronbach's Alpha=0.84; Anderson, McClintock, et al., 2016). The SSI displayed comparably good internal reliability in the current sample (Alpha= 0.86).

Inventory of Interpersonal Problems (IIP-32). The IIP-32 has 32 items scored on 0-4 Likert scale and factors into eight sub-scales. The IIP-32 assesses interpersonal problems, higher scores indicate greater interpersonal difficulties. The IIP-32 has good internal reliability (Alpha Coefficient= 0.87) and satisfactory test-retest reliability (r= 0.70; Barkham & Hardy, 1996). The IIP-32 displayed comparably good internal reliability in the current sample (Alpha= 0.87).

Patient Data

Patient Health Questionnaire (PHQ-9). The PHQ-9 consists of nine items scored on a scale between 0-3, designed to assess self-reported severity of symptoms associated with depression experienced over the previous two weeks. The PHQ-9 has been shown to have high internal validity (Cronbach's Alpha= 0.89), high test-retest reliability (r=0.84) and displayed satisfactory construct and convergent validity in a primary care population. Diagnostic criterion has been assessed with this measure which recommended that a score of ≥10 score is utilised as a cut off for clinically important symptoms of depression with adequate sensitivity and specificity (88% and 88% respectively; Kroenke et al., 2001). A reliable change index score for this measure is recommended as ≥6 (Richards & Borglin, 2011).

Generalised Anxiety Disorder scale (GAD-7). The GAD-7 consists of seven items scored on a scale between 0-3, designed to assess self-reported severity of symptoms associated with Generalised Anxiety Disorder experienced

over the previous two weeks. The GAD-7 has been shown to have high internal validity (Cronbach's Alpha=0.92), high test-retest reliability (r=0.82) and displayed strong construct and convergent validity in a primary care population (Spitzer et al., 2006). Diagnostic criterion has been assessed with this measure which recommended that a score of ≥8 score is utilised as a cut off for clinically important symptoms of anxiety with adequate sensitivity and specificity (77% and 82% respectively; Kroenke et al., 2007). A reliable change index score for this measure is recommended as ≥5 (Richards & Borglin, 2011).

Pre-processing and Data Preparation

Prior to formal hypothesis testing, the dataset was inspected to check statistical assumptions and to prepare and quality control the available data. There was missing item-level data in the therapist-reported measures (0.22% of items missing) and patient-level data with evident inputting errors (3.50% of cases overall). These were resolved by imputation using an expectation maximization approach (Schafer & Olsen, 1998). Some patients were treated by two participating therapists; and as this violates the independence of observations assumption, only the most recent contact with a therapist was included in the study. Preliminary model-fitting indicated that entering post-treatment symptom measures (PHQ-9, GAD-7) as dependent variables in regression analysis raised problems related to heteroscedasticity (see Appendix G). These model-fitting steps led to the decision to introduce pre-post treatment change scores (baseline minus final score) as the dependent variable in the primary analysis, which met the statistical assumptions of multilevel modelling.

Analytic Strategy

Pre-post treatment effect sizes (Cohen's d) were calculated using G*Power software (Faul et al., 2007). Reliable changes in depression and anxiety symptoms were explored using the clinical cut-off criterion and reliable change indices described above. All pre-processing steps were conducted in IBM SPSS, and the dataset was imported into MLwiN where all subsequent multilevel modelling (MLM) analyses were conducted. The multilevel models consisted of two levels, including patients at level 1 and therapists at level 2. The following process was conducted with the PHQ-9 change score and then repeated with the GAD-7 change score as the dependent variables. As is convention in MLM, the model was built upwards iteratively. Firstly, a null model with no predictors was modelled, building up to a case-mix adjusted model. The case-mix model was built by entering level 1 variables (continuous variables which were grand-mean centred), retaining only significant predictors using a backward elimination approach. Additionally, NHS trust and therapist professional role were entered to hold these constant in the model. Interactions between variables in the model were then tested iteratively and significant interactions retained. Development of this model was guided by inspection of beta coefficients and standard errors to assess significance of individual variables and interactions, and the -2-log likelihood ratio statistic to assess overall goodness-of-fit of the model. Random slopes were tested on the relevant baseline measure scores (PHQ-9, or GAD-7). Once these models were built, the intra-class correlation coefficient (ICC) was calculated as a measure of variance explained by the therapist-level (i.e., therapist effects). Residuals were calculated and these were plotted against the grand means of the predictor variables which were visually inspected; this

suggested there could be outlying therapists, but these were retained in the primary analysis to maintain the integrity of the data.

Next, therapist covariates were tested, but only retained if they contributed significantly to the model. Then level-2 predictors were fit to the model to asses for significance; this was firstly the full-scale scores of the primary measures of interest (i.e. GRAS, SSI and IIP32). Then, these were subsequently removed and exploratory analysis using subscale scores were entered into the model. Due to the small number of supervisor-reported data, this was tested in a separate model to therapist-reported outcomes. The ICC was recalculated at this stage to reassess the extent to which the added variables accounted for the therapist effect.

To assess the impact of the two potentially outlying therapists, a separate exploratory sensitivity analysis was conducted. This was done by holding the two therapists constant in the model, as suggested by Rasbash et al., (2020). Exploratory analyses of the relationships between independent variables were conducted by running correlations between the variables. This utilised all available therapist data (N=61) and all available data from supervisors which corresponded to a therapist (N=24).

Results

Large pre-post treatment effect sizes were observed for both the PHQ-9 (d= 0.85) and the GAD-7 (d= 0.87) in the full sample. On the PHQ-9 1326 out of 2914 (45.5%) patients displayed a reliable improvement (≥6) in depression symptoms, whilst 73 (2.5%) patients displayed a reliable deterioration in depression symptoms. On the PHQ-9, 2312 out of 2914 (79.3%) patients were at caseness on the measure at the start of treatment. Of these patients, 1008 (43.6%) had a clinically significant reduction on the PHQ-9 and can therefore be

considered to have achieved reliable and clinically significant improvement (Jacobson & Truax, 1991). Similarly, on the GAD-7 measure, 1511 out of 3109 (48.6%) patients displayed a reliable improvement (≥5) in anxiety symptoms, whilst 98 (3.2%) patients displayed reliable deterioration in anxiety symptoms. On the GAD-7 2459 out of 3109 (79.1%) patients were at caseness at the start of treatment. Of these patients, 1114 (45.3%) attained reliable and clinically significant decreases in anxiety.

Multilevel Modelling (MLM) of Therapist Effects.[5]

MLM results show a significant therapist effect on both depression and anxiety outcomes. The size of therapist effect was small on both outcomes, with an effect size of 2.69% (ICC= 0.0269; p= 0.005) and 2.78% (ICC= 0.0278; p= 0.005) for depression (PHQ-9) and anxiety (GAD-7) respectively. Pre-treatment depression scores significantly predicted outcomes on both depression and anxiety measures (p< 0.001 and p< 0.001), whilst the pre-treatment anxiety measure only predicted outcomes on the anxiety measure (p< 0.001 and p= 0.082). Additionally, being a female patient predicted significantly poorer outcomes on the anxiety outcome (p= 0.001) but not depression outcome (p= 0.214). Across both measures, CBT therapists achieved significantly better outcomes than psychological wellbeing practitioners (Depression: p= 0.008; Anxiety: p= 0.009), but significantly worse outcomes than counsellors (Depression: p= 0.024; Anxiety: p= 0.011). Additionally, on both measures being older (Depression: p= 0.022; Anxiety: p= 0.003) and from a white ethnic background (Depression: p= 0.013; Anxiety: p= 0.035) was associated with better treatment outcomes. The two participating NHS trusts also demonstrated

[5] All models reported either presented in Tables 3 and 4 in text or can be found in Appendix H

significantly different treatment outcomes (Depression: p< 0.001; Anxiety: p< 0.001).

Multilevel Model: Depression

Fully adjusted MLM models are presented in Table 3 and Figure 2 contains a caterpillar plot displaying the variability of therapist effectiveness. The first depression model reflects the assessment of the primary hypothesis. This indicated that therapists' social skills were significantly associated with treatment outcomes (B= 0.016; S.E= 0.007; p= 0.022). A positive regression coefficient reflects that higher scores on the social skills measure were associated with greater changes in depression at the end of therapy. There was no significant effect of therapist-reported interpersonal problems (B= -0.001; S.E= 0.011; p= 0.928) or self-reflection skills on treatment outcomes (B= 0.021; S.E= 0.033; p= 0.525). Visual analysis of GRAS scores plotted against predicted residuals (calculated from case-mix model) suggested a potential positive relationship apart from two therapists' data points contrary to the trend at each extreme (see

Figure 2

A Plot of Therapist Rank Versus Therapists' PHQ-9 Outcome Residuals

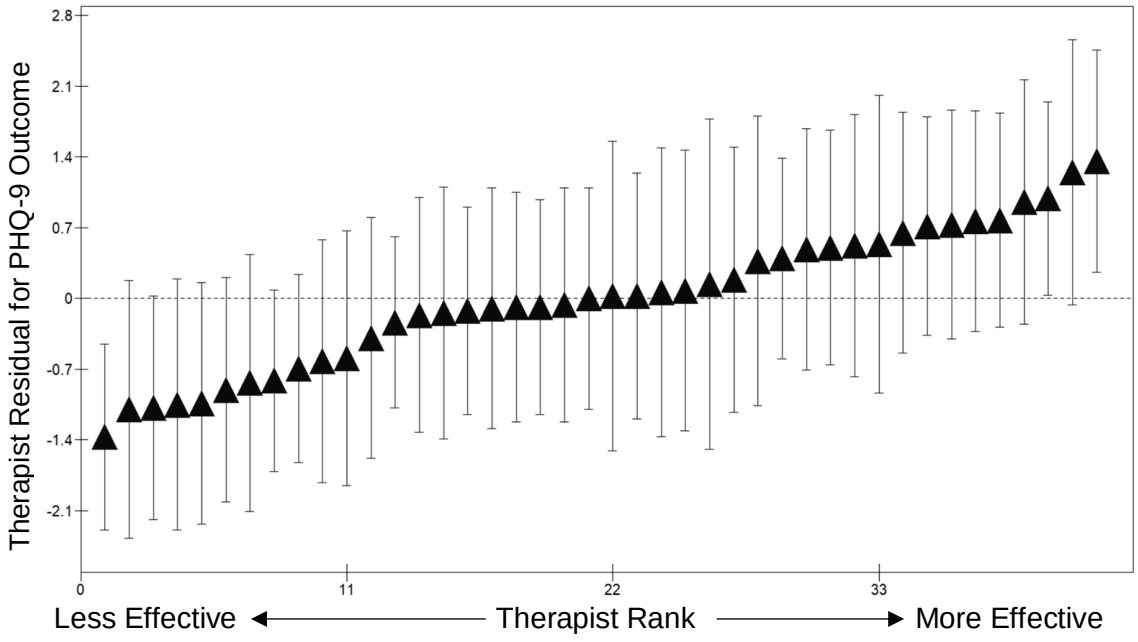


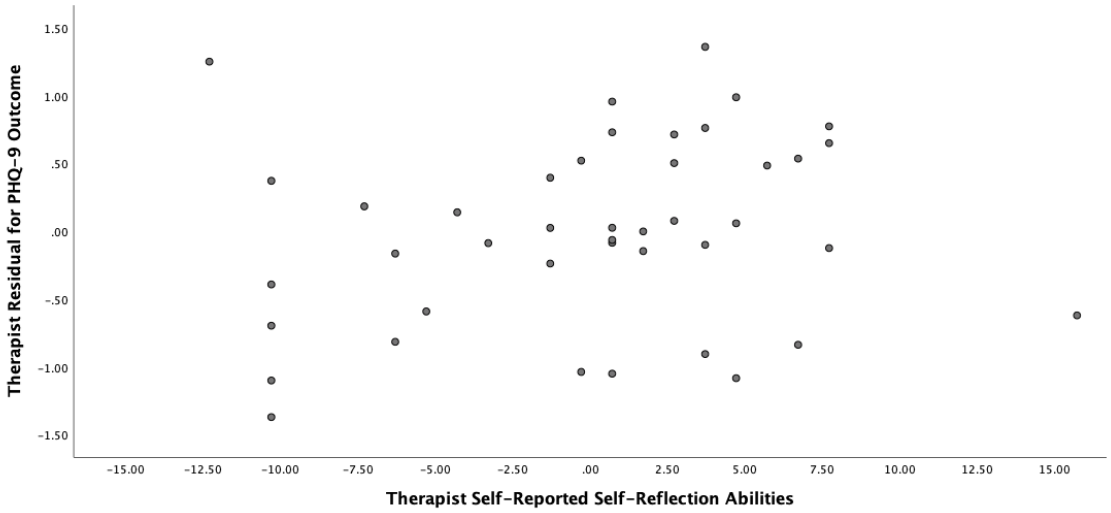
Figure 3). The separate sensitivity analyses of the primary predictors with these therapists absorbed into the model as dummy variables (held constant) found that both higher therapist-reported self-reflection skills (B= 0.070; S.E= 0.031; p= 0.024) and social skills (B= 0.015; S.E= 0.006; p= 0.012) were associated with significantly greater change on the PHQ-9. There was no significant effect of interpersonal problems (B= -0.005; S.E= 0.009; p= 0.579) in this model.

Exploratory analysis of subscale scores indicated non-significant effects of all subscales of the SSI, and all but the empathetic reflection scale of the GRAS on treatment outcomes. The empathetic reflection scale was significantly associated with treatment outcomes even adjusting p for multiple comparisons ($p< 0.001$) and remained so when non-significant subscales were removed, and social skills was re-entered into the model (B= 0.206; S.E= 0.0076; p= 0.007); this final model is presented in Table 3. However, due to low Cronbach's Alpha of this scale, this should be interpreted tentatively. The recalculated therapist effect using this model was 1.37% (ICC= 0.0137) in comparison to the original therapist effect without predictors. This indicates that these variables accounted for 49.10% of the variance in treatment outcome explained by the therapist.

Supervisor-rated self-reflective abilities were not significantly predictive of patient outcomes (B= 0.016; S.E= 0.026; p= 0.538). Due to small number of therapists in this analysis (N=18) and some issues with normality of residuals, this result should be interpreted very tentatively.

Figure 3

Therapist Self-reflection Mean-Centred (GRAS) Plotted Against Therapists' Residual for PHQ-9 Treatment Outcome



Multilevel Model: Anxiety

Fully adjusted MLM models are presented in Table 4 and Figure 4 contains a caterpillar plot displaying the variability of therapist effectiveness. The first model reflects assessment of the primary hypothesis. In this model none of the primary variables were significantly associated with treatment outcome (social skills: B= 0.008, S.E= 0.006; p= 0.182; self-reflection: B= 0.037, S.E= 0.031; p= 0.233; and interpersonal problems: B= 0.000, S.E= 0.011; p= 1.000). Again, visual analysis of the GRAS scores plotted against anxiety treatment outcome appeared to display a trend, with two therapists' data points contrary to this trend (see Figure 5). Sensitivity analysis with these therapists held constant, found that self-reflection was significantly associated with treatment outcomes (B= 0.082, S.E= 0.028; p= 0.003) but there was not a significant effect of social skills (B= 0.008, S.E= 0.005; p= 0.110) or interpersonal problems (B= -0.003, S.E= 0.008; p= 0.708).

Table 3

Summary of Multilevel Models for Depression

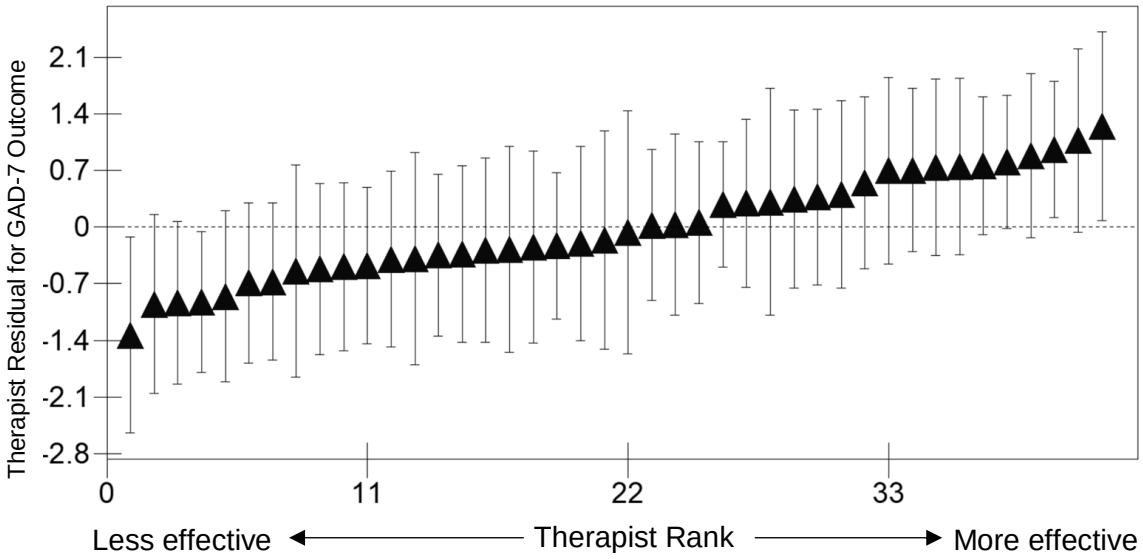
		Primary Predictors		Final Model	
		B (S.E)	*P*	B (S.E)	*p*
Intercept		7.357(0.373)	<0.001	7.307(0.341)	<0.001
Pre-treatment PHQ-9		0.616(0.033)	<0.001	0.614(0.033)	<0.001
Unemployed at referral		-2.121(0.312)	<0.001	-2.110(0.312)	<0.001
Professional Role [a]:	Counsellor	1.482(0.490)	0.002	1.306(0.441)	0.003
	PWP	-1.469(0.405)	<0.001	-1.346(0.371)	<0.001
Patient age		0.016(0.007)	0.022	0.016(0.007)	0.022
NHS Trust		-3.000(0.405)	<0.001	-2.917(0.370)	<0.001
BAME [b]		-1.040(0.419)	0.013	-1.040(0.419)	0.013
Pre-treatment PHQ-9) * PWP [c]		-0.147(0.037)	<0.001	-0.145(0.037)	<0.001
NHS Trust* Unemployed at referral		1.392(0.605)	0.021	1.390(0.605)	0.022
Pre-treatment PHQ-9 * BAME		-0.196(0.069)	0.005	-0.198(0.069)	0.004
Pre-treatment PHQ-9 * NHS Trust		-0.190(0.041)	<0.001	-0.192(0.041)	<0.001
Pre-treatment PHQ-9 * Unemployed at referral		-0.163(0.047)	0.001	-0.162(0.047)	0.001
Self-Reflection Ability (GRAS)		0.021(0.033)	0.525	-	-
Interpersonal Problems (IIP32)		-0.001(0.011)	0.928	-	-
Interpersonal Skills (SSI)		0.016(0.007)	0.022	0.016(0.006)	0.008
Empathetic Reflection (GRAS subscale)		-	-	0.206(0.076)	0.007

Note: * indicates an interaction term. B= regression coefficient S.E= Standard Error

[a] Compared to CBT therapist. [b] Black, Asian and Minority Ethnicity compared to White ethnicity. [c] Compared to CBT and Counsellor grouped.

Figure 4

A Plot of Therapist Rank Versus Therapists' GAD-7 Outcome Residuals

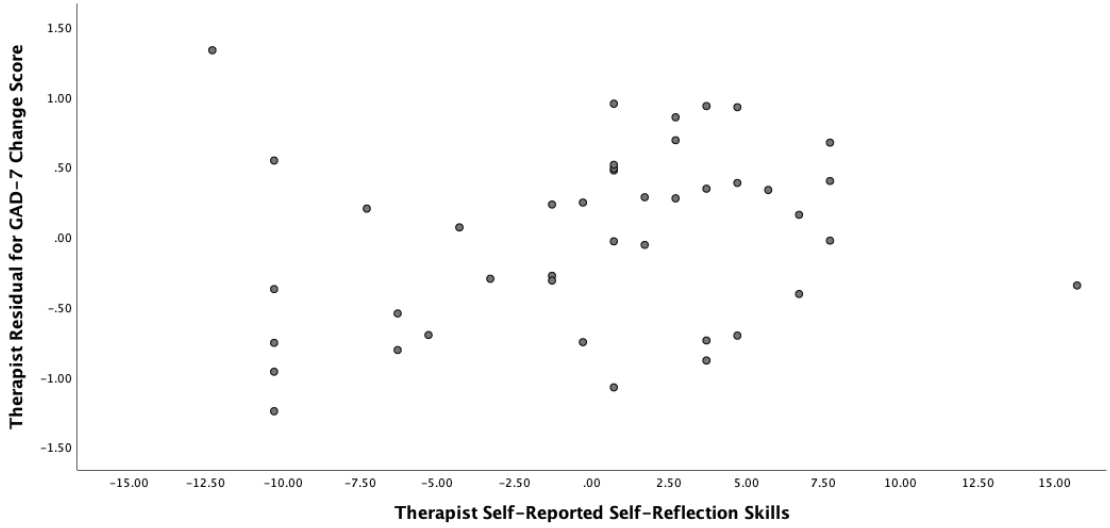


When subscales were entered into the model only empathetic reflection was significantly associated with treatment outcome. This was again highly significantly associated with anxiety outcomes (B= 0.252; S.E= 0.070; p< 0.001). Again, given the low Cronbach's Alpha of this scale this should be interpreted cautiously. This was the only predictor retained in the final anxiety model (see Table 4). The recalculated therapist effect using this model was 1.56% (ICC= 0.0156), indicating that this variable accounted for 43.88% of the variance explained by the therapist.

Supervisor variables presented a similar pattern to the depression model. The full scale was not predictive of treatment outcomes (B= 0.020; S.E= 0.025; p= 0.424). Again, due to small number of therapists (N=18) and issues with normality of residuals, this result should be interpreted very tentatively.

Figure 5

Therapist Self-reflection Mean-Centred (GRAS) Plotted Against Therapists' Residual for GAD-7 Treatment Outcome



Secondary Correlational Analysis

Secondary correlational analysis displayed moderate inverse relationships between supervisor-reported self-reflection and therapist-reported interpersonal problems (r= -0.638, p= 0.001). However, supervisor-reported self-reflection was not related to therapist-reported self-reflection (r= 0.363; p= 0.139) or therapist-reported social skills (r= 0.082; p= 0.71). Therapist-reported self-reflection correlated with both interpersonal problems (r= -0.442, p< 0.001) and social skills (r= 0.343, p= 0.010). Social skills and interpersonal problems were not significantly correlated when rated by therapists (r= -0.225, p= 0.095).

Table 4

Summary of Multilevel Models for Anxiety

	Primary Predictors		Final Model	
	B (S.E)	p	B (S.E)	p
Intercept	7.488(0.379)	<0.001	7.123(0.335)	<0.001
Pre-treatment PHQ9	-0.202(0.024)	<0.001	-0.200(0.024)	<0.001
Pre-treatment GAD7	0.678(0.027)	<0.001	0.670(0.027)	<0.001
Unemployed at referral	-1.794(0.279)	<0.001	-1.767(0.279)	<0.001
Professional Role[a]: Counsellor	0.866(0.461)	0.060	0.815(0.395)	0.039
Professional Role[a]: PWP	-1.264(0.379)	0.001	-0.917(0.328)	0.005
Patient Age	0.020(0.006)	0.001	0.018(0.006)	0.003
NHS Trust	-2.061(0.379)	<0.001	-1.688(0.307)	<0.001
BAME[b]	-0.764(0.363)	0.035	-0.730(0.349)	0.036
Female[c]	-0.657(0.203)	0.001	-0.509(0.198)	0.010
Pre-treatment PHQ9* Pre-treatment GAD7	-0.016(0.003)	<0.001	-0.015(0.003)	<0.001
Pre-treatment GAD7* Unemployed at referral	-0.160(0.051)	0.002	-0.152(0.049)	0.002
Pre-treatment PHQ9* NHS Trust	-0.109(0.036)	0.002	-0.099(0.034)	0.004
NHS Trust* Unemployed at referral	1.620(0.550)	0.003	1.499(0.512)	0.003
Self-Reflection (GRAS)	0.037(0.031)	0.233	-	-
Interpersonal Problems (IIP32)	0.000(0.011)	1.000	-	-
Interpersonal Skills (SSI)	0.008(0.006)	0.182	-	-
Empathetic Reflection (GRAS Subscale)	-	-	0.252(0.07)	<0.001

Note: * indicates an interaction term. B= regression coefficient. S.E= Standard Error

[a] Compared to CBT therapists. [b] Black, Asian and Minority Ethnicity compared to White Ethnicity. [c] Compared to males.

Discussion

This study aimed to explore the role of therapists' interpersonal, social and reflective skills on the outcomes achieved by their patients in a routine practice setting. The therapist effect of 2.69% for anxiety outcomes and 2.78% for depression outcomes was smaller than the average therapist effect reported of between 5-10% (Baldwin & Imel, 2013; Johns et al., 2019). However, this finding was comparable to a recent therapist effects study, which found therapist effects in the ranges of 1-3% (Delgadillo et al., 2020). The findings indicated that therapists' perception of their social skills was predictive of treatment outcomes for depression, but not anxiety. Therapists' self-reflection and interpersonal problems did not predict treatment outcomes for either depression or anxiety. Exploratory analyses suggested that therapists that rated themselves as higher in empathetic reflection generated consistently better treatment outcomes for depression and anxiety.

Patients attained better depression outcomes if they were treated by therapists with higher self-reported social skills. This finding is at odds with Heinonen and Nissen-Lie (2020), who suggested that self-reported social skills had little value in predicting treatment outcomes. This diversity could reflect the very different methods used in the studies. The current study also had a much larger sample at both the patient and therapist level than studies by Anderson and colleagues (Anderson, Crowley, et al, 2016; Anderson, McClintock, et al., 2016; Anderson, et al., 2009). This finding only applied to outcomes on the PHQ-9, therefore possibly reflecting an effect which is more readily detectable when considered related to patient's depression symptoms. This contrasts to the studies cited by Heinonen and Nissen-Lie (2020) in which outcomes were

measured on more broad indices of psychological distress (e.g. Outcome Questionnaire-45; Lambert et al., 1994).

Depression outcomes are similar irrespective of treatment modality used (Barth et al., 2013; Cuijpers et al., 2008), which suggests that shared 'common factors' across therapies creating treatment outcome. It could be argued that therapist social skills represent a common factor in treatment of depression, which facilitates communication of empathy and nurturance of the therapeutic relationship (Wampold, 2015). It is possible that this is particularly important given that symptoms associated with depression (e.g. loss of interest or loss of energy; National Institute for Clinical Excellence, 2009) may mean that additional skill is required to build engagement and motivation with a patient.

Additionally, of interest, none of the social skills subscale scores were independently predictive of treatment outcomes, suggesting the importance of global social skills as opposed to any specific strength in the components which contribute to the global score. It was anticipated that interpersonal problems would inversely correlate with social skills however this was not observed. This suggests that these are separate constructs as opposed to being opposite ends of the same construct. This was reflected in MLM as interpersonal problems demonstrated no relationship to patient outcomes.

Therapist-reported self-reflection was not associated with treatment outcomes when all therapists were included in the analysis. However, sensitivity analysis holding constant two therapists who appeared to be outlying, did find a significant positive effect of therapist-reported self-reflection on treatment outcomes for both depression and anxiety. The sensitivity analysis was not intended to overwrite the primary finding that there was no effect. However, this

finding is of interest as it indicates two patterns of responding which may undermine attempts to draw conclusions about the relationships between therapist-reported self-reflection and patient outcomes. For high therapist-reported reflective ability but poor treatment outcomes, it could be argued that this reflects a deficit in metacognitive awareness (Kruger & Dunning, 1999). This may mean that a very high score (i.e. as was the case for one therapist) is evidence of a lack of, rather than presence of, reflective ability. On the contrary, where therapists report low reflective ability, but achieve good outcomes, it is possible this reflects a greater awareness of personal and therapeutic limitations. This could therefore indicate useful self-doubt, which has previously been associated with positive outcomes (Meichenbaum & Lilienfeld, 2018; Nissen-Lie et al., 2017).

Although self-reflection as a full scale was not predictive of outcomes, the empathetic reflection subscale did predict treatment outcomes for both depression and anxiety. However, this finding should be read with caution, given the low internal reliability of this scale in the study and the scale primarily intending to assess a single construct. This scale is described as being the intersubjective and social extension of internal self-reflection and consists of the ability to consider the positions of others (Aukes et al., 2007). This scale may tap into the associated abilities of mentalisation and reflective functioning, which has been previously identified as being important in relation to treatment outcomes (Cologon et al., 2017). It is worth noting that the two outlying scores accounted for in the sensitivity analysis of the full scale were not extreme on this subscale, which could offer a statistical rather than theoretical explanation as to why this subscale was predictive of outcomes whilst the full scale was not.

The current study attempted to utilise the perspective of participating therapist's supervisors. Unfortunately, there was only corresponding data for 18 therapists included and 24 in total. Supervisor ratings of therapist's reflective abilities was not associated with patient outcomes. However, little weight should be placed on these findings, as only N=18 participants at level 2 is too few to draw reliable conclusions from (Maas & Hox, 2005). It is interesting that the supervisor and therapist reflection ratings did not correlate, suggesting that each party was rating differently. This was also reflected in supervisors, on average, reporting lower reflection scores than therapists. Supervisor reported therapist-reflection did correlate moderately with the therapist-reported measure of interpersonal problems. Whilst it is beyond the scope of this paper to explore this fully, it is indicative supervisor report can detect important characteristics about therapists.

Limitations

There are limitations in this study which should be considered. Due to a worldwide pandemic recruitment in this study was curtailed. At the point recruitment was ended, there were sufficient therapists. However, data collection from supervisors was terminated early. Unfortunately, the small number of associated supervisor data means it is difficult to make conclusions about 'observer reports' in this study. Additionally, challenges related to COVID-19 led to delays in receiving data, and therefore only data from three out of four services recruited from was received in time for analysis. This meant that 15 therapists from one service could not be used in the multilevel model. Despite sufficient patient data at level 1, this means that the study was potentially underpowered at level 2.

Whilst naturalistic data collection has strengths related to representing real world clinical practice there are some drawbacks of utilising this data. Firstly, there is no control over quality of data recording. This was reflected in the current sample where data from one Trust site appeared to have erroneous data recorded which required imputation. Additionally, collection of outcome measures cannot be standardised and could influence scores reported (e.g. if measures are completed in the presence of the therapist this could lead to socially desirable responses inflating the effect of the therapy). Furthermore, as there was no way to consider allocation of patient to therapist, it was possible that certain patients were systematically allocated to certain therapists (e.g. the more experienced therapists seeing the more complex clients).

Diagnosis was recorded by services, but these were not formal diagnosis and therefore were not considered in the analysis. Whilst all available data was included, this may not have been the primary focus of the intervention, for example, a patient with an anxiety disorder whose primary treatment was related to anxiety was also included in the depression analysis if data was available. This could have led to underestimates of treatment effectiveness. Additionally, whilst there was a significant difference between NHS trusts there was insufficient data to explore this further. For example, there was no information in the dataset related to deprivation of patients, nor was there any details about factors such as therapist's caseload and service waiting times which may have added context to this discrepancy. Therapists were invited to participate in this study voluntarily, this may have led to a self-selecting bias (Sharma, 2017), and responses being biased by participants responding in a socially desirable way (Grimm, 2010). Finally, this research was conducted within an IAPT setting, and because there is evidence that therapist effects are different in different treatment contexts (Firth

et al., 2020), caution should be applied in generalising these findings to other contexts.

Clinical Implications

These findings have important clinical implications. This study contributes to the emerging literature identifying those therapists' characteristics which are associated with patient outcomes. This indicates that focus on the training and supervision of therapists could be better based on the evidence from the therapist effects literature. However, before substantial changes are made to clinical practice the findings of this, and similar studies, should be replicated to ensure that observed findings are robust and replicable.

Future Research and Training Directions

1. There were too little available data from supervisors in this study. This remains a useful resource. It would be beneficial for further research to employ a similar design but expand the number of supervisors. Unlike the current study it may be advisable to collect data in two sequential steps, first therapist-reported data and second supervisor-rated for participating therapists.
2. There may be specific patterns of therapists responding which undermine associations between therapist-reported self-reflection and treatment outcomes. It would be helpful to establish if these are consistent and replicable patterns of responding. This would be assisted by the collection of observers reported measures of self-reflection, which could be compared with therapist-reported data.
3. This study fits alongside an emerging literature assessing therapist level variables as predictors of treatment outcomes. Further research on

therapist effects should endeavour to collect data directly from therapists to test associations with the outcomes the therapist achieves.

4. If therapist skills/ abilities are consistently associated with treatment outcomes, then an important question is whether these are skills/ abilities that can trained. The answer to this question should determine whether training programmes focus on developing these abilities or whether focus ought to be on recruiting people into psychotherapeutic roles on the basis of these skills/ abilities.
5. Change in depression/ anxiety scores is one way of assessing the effectiveness of therapists. Other treatment outcomes, such as dropout may also be of interest. It would be interesting to understand whether greater social skills could also predict therapists' patient dropout rate.
6. Collecting data about supervisor's views of therapists interpersonal functioning was considered in this study, however feedback from a sample of supervisors during the design of the study suggested that supervisors would feel unable to respond about more general components of interpersonal functioning considered in the measures used in the study. Future research may consider development of an observer-reported measure more specifically designed for interpersonal functioning in the therapeutic context.

Conclusion

This study investigated whether therapists self-perceived social skills, interpersonal problems and self-reflection, and supervisor reported self-reflection, could predict patient outcomes. The findings indicated that therapist-reported social skills were associated with patient treatment outcomes for depression but not anxiety. Interpersonal problems were not associated with

treatment outcomes. Therapist-reported self-reflection was not associated with treatment outcomes, but a sensitivity analysis controlling for outlying variables did find an association between self-reported self-reflection and patient outcome. This raises the question of whether clinically important outlying variables may undermine therapist-reported self-reflection as a variable of interest in understanding therapist effects. Empathetic self-reflection was associated with treatment outcomes for outcomes on both depression and anxiety but the subscale displayed low internal reliability and therefore strong conclusions cannot be made from this finding. There was insufficient data to make meaningful conclusions about whether supervisor reported therapist-reflection was associated with treatment outcomes. These findings may have important clinical implications in terms of the type of therapists' abilities that are assessed and trained. However, further well-powered research is required to replicate and build on the findings in this study.

References

American Psychological Association. (2012). *Revised competency benchmarks in professional psychology*. http://www.apa.org/ed/graduate/benchmarks-evaluation-system.aspx

Anderson, T, Crowley, M. E. J., Himawan, L., Holmberg, J. K., & Uhlin, B. D. (2016). Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome. *Psychotherapy Research*, *26*(5), 511–529. https://doi.org/10.1080/10503307.2015.1049671

Anderson, T, McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016). A Prospective Study of Therapist Facilitative Interpersonal Skills as a Predictor of Treatment Outcome. *Journal of Consulting and Clinical Psychology*, *84*(1), 57–66. https://doi.org/10.1037/ccp0000060

Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. Journal of Clinical Psychology, 65(7), 755–768. https://doi.org/10.1002/jclp.20583

Anderson, Timothy, & Patterson, C. (2013). *Facilitative Interpersonal Skill Task and Rating Method*. http://www.safranlab.net/uploads/7/6/4/6/7646935/fis_task_and_coding_manual_2013_10__1_.pdf

Aukes, L. C., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R. P., & Slaets, J. P. J. (2007). The development of a scale to measure personal reflection in medical practice and education. *Medical Teacher*, *29*(2–3), 177–182.

https://doi.org/10.1080/01421590701299272

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In Michael J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). John Wiley & Sons, Inc.

Bambling, M., & King, R. (2013). Extended problem solving treatment for depression. *Counselling and Psychotherapy Research*, *13*(4), 317–323. https://doi.org/10.1080/14733145.2012.754918

Barkham, M., & Hardy, G. E. (1996). The IIP-32: A short version of the Inventory of Interpersonal. *British Jowrnal of Clinical PIyrhology*, *35*(1), 21–35. https://doi.org/10.1111/j.2044-8260.1996.tb01159.x

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., Jüni, P., & Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine*, *10*(5). https://doi.org/10.1371/journal.pmed.1001454

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

Bennett-Levy, J. (2019). Why therapists should walk the talk: The theoretical and empirical case for personal practice in therapist training and professional development. *Journal of Behavior Therapy and Experimental Psychiatry*, *62*, 133–145. https://doi.org/10.1016/j.jbtep.2018.08.004

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M J Lambert (Ed.), B*ergin and Garfield's handbook of psychotherapy and behavioral change* (5th ed., pp.

227–306). Wiley.

British Psychological Society. (2019). *Standards for the accreditation of doctoral programmes in clinical psychology.* British Psychological Society. https://www.bps.org.uk/sites/bps.org.uk/files/Accreditation/Clinical%20Accreditation%20Handbook%202019.pdf

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge University Press.

Cologon, J., Schweitzer, R. D., King, R., & Nolte, T. (2017). Therapist reflective functioning, therapist attachment style and therapist effectiveness. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(5), 614–625. https://doi.org/10.1007/s10488-017-0790-5

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials*. BMJ, 340,* c332. https://doi.org/10.1136/bmj.c332

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., Luborsky, L., McLellan, T., Woody, G. E., Thompson, L., Gallagher, D., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, *1*(2), 81–91. https://doi.org/10.1080/10503309112331335511

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*, *76*(6), 909–922. https://doi.org/10.1037/a0013075

Delgadillo, J., Branson, A., Kellett, S., Myles-Hooton, P., Hardy, G. E., &

Shafran, R. (2020). Therapist personality traits as predictors of psychological treatment outcomes. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2020.1731927

Delgadillo, J., Saxon, D., & Barkham, M. (2018). Associations between therapists' occupational burnout and their patients' depression and anxiety treatment outcomes. *Depression and Anxiety*, *35(9),* 1–7. https://doi.org/10.1002/da.22766

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behaviour Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Firth, N., Saxon, D., Stiles, W. B., & Barkham, M. (2020). Therapist effects vary significantly across psychological treatment care sectors. *Clinical Psychology and Psychotherapy*. https://doi.org/10.1002/cpp.2461

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*. https://doi.org/10.1037/cou0000131

Grant, A. M., Franklin, J., & Langford, P. (2002). The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality*, *30*(8), 821–836. https://doi.org/10.2224/sbp.2002.30.8.821

Green, H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT Psychological Wellbeing Practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy*, *63*, 43–54.

https://doi.org/10.1016/j.brat.2014.08.009

Grimm, P. (2010). Social Desirability Bias. In *Wiley International Encyclopedia of Marketing*. Wiley. https://doi.org/10.1002/9781444316568.wiem02057

Heinonen, E., Lindfors, O., Laaksonen, M. A., & Knekt, P. (2012). Therapists' professional and personal characteristics as predictors of outcome in short- and long-term psychotherapy. *Journal of Affective Disorders*, *138*(3), 301–312. https://doi.org/10.1016/j.jad.2012.01.023

Heinonen, E., & Nissen-Lie, H. A. (2020). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy Research*. *30(4),* 417-432. https://doi.org/10.1080/10503307.2019.1620366

Jacobson, N. S., & Truax, P. (1991). Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology*. *59(1),* 12-19. https://doi.org/10.1037/0022-006X.59.1.12

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to Baldwin and Imel's (2013) review. *Clinical Psychology Review*, *67*, 78–93. https://doi.org/10.1016/j.cpr.2018.08.004

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9 validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment,

comorbidity, and detection. *Annals of Internal Medicine, 146(5),* 317–325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), *1121*

Lambert, M. J, Lunnen, K., Unphress, V., Hansen, N., & Burlingame, G. M. (1994). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.1)*. IHC Center for Behavioral Healthcare Efficacy.

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M J Lambert (Ed.), *Handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). John Wiley & Sons Inc.

Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy: Theory, Research and Practice, 38(4),* 357-361. 10.1037/0033-3204.38.4.357

Lutz, W., & Barkham, M. (2015). Therapist Effects. In *The encyclopedia of clinical psychology*. https://doi.org/doi:10.1002/9781118625392.wbecp109

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Meichenbaum, D., & Lilienfeld, S. O. (2018). How to spot hype in the field of psychotherapy: A 19-item checklist. *Professional Psychology: Research and Practice*, *49*(1), 22–30. https://doi.org/10.1037/pro0000172

Muran, C. J., Safran, J. D., Eubanks, C. F., & Gorman, B. S. (2018). The effect

of alliance-focused training on a cognitive-behavioral therapy for personality disorders. *Journal of Consulting and Clinical Psychology*, *86*(4), 384–397. https://doi.org/10.1037/ccp0000284

National Institute for Clinical Excellence. (2009). *Depression in adults: recognition and management*. Retrieved June 9, 2020, from https://www.nice.org.uk/guidance/CG90

Nissen-Lie, H. A., Monsen, J. T., & Rønnestad, M. H. (2010). Therapist predictors of early patient-rated working alliance: A multilevel approach. *Psychotherapy Research*, *20*(6), 627–646. https://doi.org/10.1080/10503307.2010.497633

Nissen-Lie, H. A., Monsen, J. T., Ulleberg, P., & Rønnestad, M. H. (2013). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research*, *23*(1), 86–104. https://doi.org/10.1080/10503307.2012.735775

Nissen-Lie, H. A., Rønnestad, M. H., Høglend, P. A., Havik, O. E., Solbakken, O. A., Stiles, T. C., & Monsen, J. T. (2017). Love yourself as a person, doubt yourself as a therapist? *Clinical Psychology and Psychotherapy*, *24*(1), 48–60. https://doi.org/10.1002/cpp.1977

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*. *3,* 400-424. https://doi.org/10.1111/j.1745-6924.2008.00088.x

Pereira, J. A., Barkham, M., Kellett, S., & Saxon, D. (2017). The role of practitioner resilience and mindfulness in effective practice: A practice-based feasibility study. *Administration and Policy in Mental Health and*

Mental Health Services Research, *44*(5), 691–704. https://doi.org/10.1007/s10488-016-0747-0

Rasbash, J., Steele, F., Browne, W. J., & Goldsten, H. (2020). *A user's guide to MLwiN, v3.05*. Centre for Multilevel Modelling, University of Bristol. http://www.bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/3-05/manual-web.pdf

Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders*, *133 (1-2),* 51-60. https://doi.org/10.1016/j.jad.2011.03.024

Ridley, C. R., Kelly, S. M., & Mollen, D. (2011). Microskills training: evolution, reexamination, and call for reform. *The Counseling Psychologist*, *39*(6), 800–824. https://doi.org/10.1177/0011000010378438

Rieck, T., & Callahan, J. L. (2013). Emotional intelligence and psychotherapy outcomes in the training clinic. *Training and Education in Professional Psychology*, *7*(1), 42–52. https://doi.org/10.1037/a0031659

Riggio, R E. (1989). *Manual for the Social Skills Inventory*. Consulting Psychologists Press.

Riggio, R. E., & E., R. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology*, *51*(3), 649–660. https://doi.org/10.1037/0022-3514.51.3.649

Riggio, R. E, & Carney, D. R. (2003). *Social Skills Inventory Manual*. Mind Garden Inc.

Ryan, A., Safran, J. D., Doran, J. M., & Muran, J. C. (2012). Therapist mindfulness, alliance and treatment outcome. *Psychotherapy Research*, *22*(3), 289–297. https://doi.org/10.1080/10503307.2011.650653

Saxon, D., Firth, N., & Barkham, M. (2017). The relationship between therapist effects and therapy delivery factors: Therapy modality, dosage, and non-completion. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(5), 705–715. https://doi.org/10.1007/s10488-016-0750-5

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. In *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr3304_5

Schiefele, A. K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., Kopta, M., Schulte, D., Saxon, D., Nielsen, S. L., & Lambert, M. J. (2017). Reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(5), 598–613. https://doi.org/10.1007/s10488-016-0736-3

Schöttke, H., Flückiger, C., Goldberg, S. B., Eversmann, J., & Lange, J. (2017). Predicting psychotherapy outcome based on therapist interpersonal skills: A five-year longitudinal study of a therapist assessment protocol. *Psychotherapy Research*, *27*(6), 642–652. https://doi.org/10.1080/10503307.2015.1125546

Sharma, G. (2017). Pros and cons of different sampling techniques. *International Journal of Applied Research*, *3*(7), 749–752.

http://www.allresearchjournal.com/archives/2017/vol3issue7/PartK/3-7-69-542.pdf

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder. *Archives of Internal Medicine*, *166*(10), 1092. https://doi.org/10.1001/archinte.166.10.1092

Takano, K., & Tanno, Y. (2009). Self-rumination, self-reflection, and depression: Self-rumination counteracts the adaptive effect of self-reflection. *Behaviour Research and Therapy*, *47*(3), 260–264. https://doi.org/10.1016/j.brat.2008.12.008

Turpin, G., & Wheeler, S. (2011). *IAPT Supervision Guidance*. http://webarchive.nationalarchives.gov.uk/20160302160224/http://www.iapt.nhs.uk/silo/files/iapt-supervision-guidance-revised-march-2011.pdf

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238

Wampold, B. E., Baldwin, S. A., Holtforth, M. grosse, & Imel, Z. E. (2017). What characterizes effective therapists? In L. G. Castonguay & C. E. Hill (Eds.) *How and why are some therapists better than others? Understanding therapist effects.* (pp. 37–53). American Psychological Association.

Appendices

Appendix A: Letters/Emails Providing Evidence of Ethical Approval

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Mr Ryan Lawty
Trainee Clinical Psychologist
Sheffield Health and Social Care NHS Foundation Trust
Cathedral Court, Clinical Psychology Unit
Sheffield
S1 2LT

Email: hra.approval@nhs.net
HCRW.approvals@wales.nhs.uk

23 September 2019

Dear Mr Lawty

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Therapist interpersonal and reflective functioning and their relationship to psychological treatment outcomes**
IRAS project ID:	**261991**
Protocol number:	**N/A**
REC reference:	**19/HRA/2818**
Sponsor	**University of Sheffield**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The *"After HRA Approval – guidance for sponsors and investigators"* document on the HRA website gives detailed guidance on reporting expectations for studies with HRA and HCRW Approval, including:

- Registration of Research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **261991**. Please quote this on all correspondence.

Yours sincerely

Joanna Ho
Approvals Specialist

Email: hra.approval@nhs.net

Copy to: *Mr Amrit Sinha, Sponsor Representative, University of Sheffield*

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
IRAS Application Form [IRAS_Form_25042019]		25 April 2019
IRAS Application Form XML file [IRAS_Form_25042019]		25 April 2019
IRAS Checklist XML [Checklist_07052019]		07 May 2019
Letter from sponsor		
Organisation Information Document [Organisation Information Document]	1	23 July 2019
Other [Ethics amendments response document]	1.1	01 July 2019
Participant consent form [Consent form]	1	01 March 2019
Participant information sheet (PIS) [Participant information sheet]	1.1	17 June 2019
Research protocol or project proposal [Protocol]	1.4	17 June 2019
Schedule of Events or SoECAT [Schedule of Events]	1	23 September 2019
Summary CV for Chief Investigator (CI) [CI CV]	1	12 April 2019
Validated questionnaire [Questionnaires]	1	01 March 2019

IRAS project ID	261991

Information to support study set up

The below provides all parties with information to support the arranging and confirming of capacity and capability with participating NHS organisations in England and Wales. This is intended to be an accurate reflection of the study at the time of issue of this letter.

Types of participating NHS organisation	Expectations related to confirmation of capacity and capability	Agreement to be used	Funding arrangements	Oversight expectations	HR Good Practice Resource Pack expectations
This is non-commercial multicentre study where all participating NHS organisations will be undertaking all research activities as described in the IRAS application. There is only one site-type in this study.	Research activities should not commence at participating NHS organisations in England or Wales prior to their formal confirmation of capacity and capability to deliver the study.	An Organisation Information Document has been submitted as an agreement between sponsor and the participating NHS organisation(s); sponsor is not requesting and does not expect any other site agreement to be used.	No funding will be provided to participating NHS organisations as indicated in the Organisation Information Document.	The Chief Investigator will be responsible for research activities at participating NHS organisations.	Local staff substantively employed by the participating NHS organisation will be undertaking research activities as described in the IRAS application. No HR access arrangements are therefore expected for this study. Where arrangements are not already in place, network staff employed by another Trust or University (or similar) undertaking any of the research activities listed in A18 or A19 of the IRAS form (except for administration of questionnaires or surveys), would be expected to obtain an honorary research contract from one NHS organisation (if university employed), followed by Letters of Access for subsequent organisations. This would be on the basis of a Research Passport (if university employed) or an NHS to NHS confirmation of pre-engagement checks letter (if NHS employed). These should confirm enhanced DBS checks, including
					appropriate barred list checks, and occupational health clearance. For research team members only administering questionnaires or surveys, a Letter of Access based on standard DBS checks and occupational health clearance would be appropriate.

Other information to aid study set-up and delivery

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales in study set-up.
The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

The University Of Sheffield.

Ryan Lawty <rlawty1@sheffield.ac.uk>

RDaSH Confirmation of Capability & Capacity - Therapist interpersonal and reflective functioning

MCKIE, Jeannie (ROTHERHAM DONCASTER AND SOUTH HUMBER NHS FOUNDATION TRUST) <j.mckie@nhs.net> 18 November 2019 at 15:49
To: Ryan Lawty <rlawty1@sheffield.ac.uk>
Cc: Jaime Delgadillo <j.delgadillo@sheffield.ac.uk>

Dear Ryan,

RDaSH Study ID: 0203

IRAS ID: 261991

Short study title: Therapist interpersonal and reflective functioning and their relationship to psychological treatment outcomes

Principle Investigator: Ryan Lawty

Grounded Research **Rotherham Doncaster and South Humber NHS Foundation Trust** (RDaSH) has completed a review of the Local Information Pack for the above study and can confirm capacity and capability for the study to be undertaken within the Trust.

Please find attached our agreed Organisation Information Document as confirmation. We agree to start this study from **18th November 2019** with a recruitment target of 12 therapists and 3 supervisors from RDaSH. Over recruitment at RDaSH is allowed within the overall study target.

That the research should:

- Be conducted in accordance with ICH GCP guidelines and that you and your team are familiar with issues of informed consent within research having completed the 'Good Clinical Practice' training at an appropriate level.
- Comply with the requirements of UK Policy Framework for Health and Social Care Research.
- Comply with regulatory requirements and legislation as applicable to your research, including: Clinical Trials, GDPR Data Protection, Health & Safety, Trust Caldicott Guidelines and the use of Human Tissue for research purposes

You must also:

- Ensure that all study personnel, not employed by RDaSH hold either a Honorary Research Contract with the Trust or a Letter of Access issued by the Trust, before they have access to any facilities, patients, staff, data, or tissue.

- Notify this office of all amendments (excluding category B amendments not relevant to this site). In accordance with National guidelines amendments may be implemented 35 days after notification to Grounded Research as long as regulatory approval has been received *and* additional review time has not been requested by the Trust. We will endeavour to provide confirmation of continued capability and capacity at the earliest opportunity.

- Notify Grounded Research when your research is completed. At the point of completion, please submit your findings, any publication or presentations of your findings.

- For monitoring purposes, you should maintain an up to date site file with all relevant information. This may be used for audit purposes in the future. Research documentation should be archived as stated in the IRAS form after the study has been completed.

- If you decide to terminate this research prematurely, you send a report to this office within 28 days, indicating the reason for the early termination.

- Advise this office of any unusual or unexpected results that raise questions about the safety of the research. Any adverse events experienced during the course of the research must be reported in accordance with the trial protocol, and notified to Grounded Research.

Please ensure a copy of this letter is filed in the Investigator Site File.

I would like to take this opportunity to wish you well with your project. If you have any questions or we can be of any further assistance to you, do not hesitate to contact Grounded Research.

Kind Regards

Jeannie

Jeannie McKie

Research Governance Manger

Grounded Research Team, Community Research Hub, RDaSH NHS Foundation Trust

e: j.mckie@nhs.net

t: 01302 798456

m: 07818560176

Please note my usual working days are Monday, Tuesday, Thursday, Friday

Note to NHS Staff Members: If you have ever been diagnosed with anxiety or depression please see the below link to a research study examining the genetic, social and environmental risk factors. (Please tick RDaSH as the NHS Trust which told you about the study) https://gladstudy.org.uk/

Organisation Information Document_RDaSH Assessed.docx
177K

Research and Innovation Department

NHS
Pennine Care
NHS Foundation Trust

Research Management Approval 6th November 2019

Dear Amrit Sinha (Sponsor Representative), Ryan Lawty (Chief Investigator)

Steve Kellett & Jaime Delgadillo, (Academic Supervisors)

Confirmation of Capacity and Capability at Pennine Care NHS Foundation Trust

Re: PCFT ref: 100449 IRAS ID: 261991

Full title of research: Therapist interpersonal and reflective functioning and their relationship to psychological treatment outcomes

Short title or acronym (if applicable): Therapist effects on psychotherapy outcomes

This email confirms that Pennine Care NHS Foundation Trust has the capacity and capability to deliver the above referenced study. Please find attached our agreed Organisation Information Document as confirmation.

We agree to start this study on **6th November 2019**, as previously discussed.

If you wish to discuss further, please do not hesitate to contact me or your assigned Resesearch Clinician, Simon Kaye on Tel: 0161 716 3993

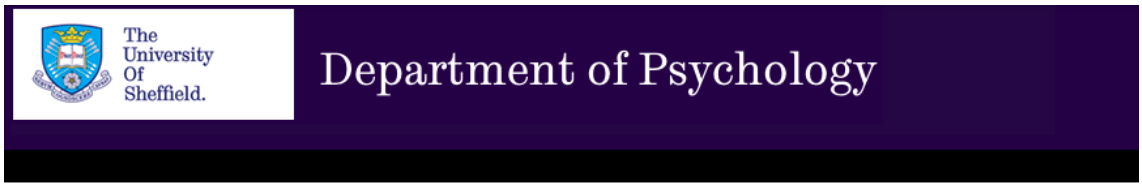
Kind regards

Susan Waine

Susan Waine
Research Governance Officer
Study Approvals

Pennine Care NHS Foundation Trust
Research and Innovation Department
225 Old Street
Ashton-under-Lyne
Lancashire
OL6 7SR
Email: researchdevelopment.penninecare@nhs.net
Website: www.penninecare.nhs.uk

Appendix B: Consent Form



Consent Page

I confirm that I have read and understood the information sheet.

○ Yes

I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline. Please contact Ryan Lawty by email (rlawty1@sheffield.ac.uk) if you wish to withdraw from the study.

○ Yes

I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access to my confidential responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.

○ Yes

I agree for the data collected from me to be used in future research.

○ Yes

I agree to take part in the research project.

○ Yes

Appendix C: Participant Information Page

Participant Information Page

1. Research Project Title
Therapist interpersonal functioning and self-reflection and their relationship to psychological treatment outcomes

2. Invitation paragraph
You are being invited to take part in a research project. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

3. What is the project's purpose?
The literature suggests that there are differences in outcomes between therapists. This project aims to understand in more detail why this might be the case by thinking about both the therapists and their supervisor's views about therapists' interpersonal skills and reflective functioning and how these relate to patient outcomes. It is hoped that this will help us to understand how therapists might be best trained and supported to improve mental health services in the future. This research is being completed as part of Doctor of Clinical Psychology Project at the University of Sheffield.

4. Why have I been chosen?
We have approached several Improving Access to Psychological Therapies (IAPT) services, all therapists and their supervisors within these services will be invited to take part in this research.

5. Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to indicate you consent to participation in the study. However, you can still withdraw at any time without any negative consequences, you do not have to give a reason. Please contact Ryan Lawty with the contact details provided below.

6. What will happen to me if I take part? What do I have to do?
You will be asked to complete a questionnaire; whether you are a therapist or supervisor will determine what you are asked about. Therapists will be asked to provide demographic information and then be asked to complete questions about their own interpersonal functioning and self-reflection. Supervisors will be asked to provide demographic information and then asked to complete questions about their supervisee's reflective functioning. Supervisors who supervise several therapists will be asked to complete the questionnaire for each of their supervisees. Therapists will be allocated a number only known by the supervisors within the therapist's service and the research team which supervisors will use to identify whom their responses relate to. Only the supervisor and the research team will be able to link the supervisor's response to the therapist. It is anticipated that it will take between 10-15 minutes to complete.

7. What are the possible disadvantages and risks of taking part?
We do not anticipate there being any significant disadvantages or risks to taking part in this research. However, to participate in a study which is considering therapist effectiveness may feel threatening, especially for therapists. We acknowledge that it may feel difficult to participate in a study of this nature, however, please be assured that your data will be kept confidential within the research team and will be completely anonymised once the study is written up. **Supervisee's responses will not be shared with their supervisor and visa-versa.**

8. What are the possible benefits of taking part?
Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will help to contribute to the understanding of how therapist characteristics might influence patient outcomes. It is hoped that this will help inform how therapists and supervisors are supported and trained in the future.

9. Will my taking part in this project be kept confidential?
All the information that we collect about you during the course of the research will be kept strictly confidential and will only be accessible to members of the research team.You will not be able to be identified in any reports or publication.

10. What is the legal basis for processing my personal data?
The University of Sheffield is the sponsor for this study based in the United Kingdom. We will be using information from you in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. The University of Sheffield will keep identifiable information about you until data is analysed and written up at which point it will become anonymous, identifiable information will be deleted at this point.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information at https://www.hra.nhs.uk/planning-and-improving-research/policies-standardslegislation/data-protection-and-information-governance/gdpr-guidance.

11. What will happen to the data collected, and the results of the research project?
Your data will be stored using a unique identity number, this will allow us to connect relevant supervisor, supervisee and patient data. Only members of the research team have access to individual responses and will be able to identify participants at this point, this database will be password protected. Once data is analysed and written up your responses will become anonymous.

The results will be written up as part of a Doctor of Clinical Psychology thesis. In addition, it is intended that the results will be disseminated by publication in a scientific journal.

Due to the nature of this research, it is likely that other researchers may find the data collected to be useful in answering future research questions. These organisations may be universities, NHS organisations or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research**.**

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research and cannot be used to contact you. You will be asked to provide consent for your data to be shared in this way.

12. Who is the Data Controller?
The University of Sheffield will act as the Data Controller for this study. This means that the University of Sheffield is responsible for looking after your information and using it properly.

13. Who has ethically reviewed the project?
This project has been ethically approved by the National Health Service Health Research Authority

14. What if something goes wrong and I wish to complain about the research?
In the first instance please contact us and we will try to resolve your concern. However, if you do not feel your complaint has been handled satisfactorily please contact Professor Gillian Hardy (Doctor of Clinical Psychology Programme Director at the University of Sheffield), who will escalate your concerns through the appropriate channels. Please do this in writing to the address below.

Professor Gillian Hardy
Clinical Psychology Unit
Cathedral Court
Vicar Lane
Sheffield
S1 2LT

If your complaint relates to how personal data has been handled, information about how to raise a complaint can be found in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

15. Who should I contact for further information?
Ryan Lawty
Clinical Psychology Unit
The University of Sheffield
1 Vicar Lane
Sheffield
S1 2LT
UK

Email: rlawty1@sheffield.ac.uk

Alternatively, you can contact:
Dr Jaime Delgadillo: j.delgadillo@sheffield.ac.uk or Dr Stephen Kellett: s.kellett@sheffield.ac.uk

Thank you for considering participating in the project.

Appendix D: Example Recruitment Emails

Therapists

Dear ${m://FirstName}

I would like to invite you to participate in the research into therapist effects on patient outcomes that your service has agreed to participate in. I am hopeful that you are already aware of this research and this email does not come as a surprise.

Below is the link for you to participate in the research, this will link you to an information page where you can find more details about the study. If you have any concerns or questions, please do not hesitate to contact me (rlawty1@sheffield.ac.uk). Please be reassured that nobody outside of the research team will be able to see your responses, this includes your supervisor.

I would like to thank you in advance for your participation, I understand you may be busy, but your time and effort is much appreciated.

Follow this link to the Survey:
${l://SurveyLink?d=Take the Survey}

Or copy and paste the URL below into your internet browser:
${l://SurveyURL}

Best wishes

Ryan Lawty
Trainee Clinical Psychologist

University of Sheffield

Follow the link to opt out of future emails:
${l://OptOutLink?d=Click here to unsubscribe}

Supervisors

Dear ${m://FirstName}

I would like to invite you to participate in the research into therapist effects on patient outcomes that your service has agreed to participate in. I am hopeful that you are already aware of this research and this email does not come as a surprise.

Below is the link for you to participate in the research, this will link you to an information page where you can find more details about the study. This link is for you to participate in the research as your role as supervisor. This is a reusable link and can therefore be used for you to respond relating to each of the therapists you supervise; you can simply close the page and reopen the link when you are ready. If you participate as a therapist first, please note that the supervisor version of the survey is much shorter and will therefore be much quicker to complete. Please be reassured that nobody outside of the research team will be able to see your responses, this includes your supervisees.

I sent an email to you today [DATE] with therapist codes which you will need to reference to identify which therapists your responses are about. If you are struggling to find this email, please contact me directly and I will share this again.

If you have any concerns or questions, please do not hesitate to contact me (rlawty1@sheffield.ac.uk).

Follow this link to the Survey:
${l://SurveyLink?d=Take the Survey}

Or copy and paste the URL below into your internet browser:
${l://SurveyURL}

Best wishes

Ryan Lawty
Trainee Clinical Psychologist

University of Sheffield

Follow the link to opt out of future emails:
${l://OptOutLink?d=Click here to unsubscribe}

Appendix E: Therapist/Supervisor Measures Used

Groningen Reflective Ability Scale: Self-reported Version

Here are some statements about your learning and functioning in practice. Using the 1-5 scale below, please indicate what *really reflects* your approach rather than what you think your experience should be. Please treat each item separately from every other item.

	1. Strongly disagree	2. Somewhat disagree	3. Neither agree nor disagree	4. Somewhat agree	5. Strongly agree
1. I want to know why I do what I do	○	○	○	○	○
2. I am aware of the emotions that influence my behaviour	○	○	○	○	○
3. I do not like to have my standpoints discussed	○	○	○	○	○
4. I do not welcome remarks about my personal functioning	○	○	○	○	○
5. I take a closer look at my own habits of thinking	○	○	○	○	○
6. I am able to view my own behaviour from a distance	○	○	○	○	○
7. I test my own judgement against those of others	○	○	○	○	○
8. Sometimes others say that I do overestimate myself	○	○	○	○	○
9. I find it important to know what certain rules and guidelines are based on	○	○	○	○	○
10. I am able to understand people with a different cultural/religious background	○	○	○	○	○
11. I am accountable for what I say	○	○	○	○	○
12. I reject different ways of thinking	○	○	○	○	○
13. I can see an experience from different standpoints	○	○	○	○	○
14. I take responsibility for what I say	○	○	○	○	○
15. I am open to discussion about my opinions	○	○	○	○	○
16. I am aware of my own limitations	○	○	○	○	○
17. I sometimes find myself having difficulty in illustrating an ethical standpoint	○	○	○	○	○
18. I am aware of the cultural influences on my opinions	○	○	○	○	○
19. I want to understand myself	○	○	○	○	○
20. I am aware of the possible emotional impact of information on others	○	○	○	○	○
21. I sometimes find myself having difficult thinking of alternative solutions	○	○	○	○	○
22. I can empathise with someone else's situation	○	○	○	○	○
23. I am aware of the emotions that influence my thinking	○	○	○	○	○

Groningen Reflective Ability Scale: Supervisor-reported Version

This page is about the approach of the therapist you are supervising. Please rate these questions to the best of your knowledge about this supervisee.

Here are some statements about their learning and functioning in practice. Using the 1-5 scale below, please indicate what *really reflects their approach*. Please treat each item separately from every other item.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
They want to know why they do what they do	○	○	○	○	○
They are aware of the emotions that influence their behaviour	○	○	○	○	○
They do not like to have their standpoints discussed	○	○	○	○	○
They do not welcome remarks about their personal functioning	○	○	○	○	○
They take a closer look at their own habits of thinking	○	○	○	○	○
They are able to view their own behaviour from a distance	○	○	○	○	○
They test their own judgements against those of others	○	○	○	○	○
Sometimes others say that they do overestimate themselves	○	○	○	○	○
They find it important to know what certain rules and guidelines are based on	○	○	○	○	○
They are able to understand people with a different cultural/religious background	○	○	○	○	○
They believe they are accountable for what they say	○	○	○	○	○
They reject different ways of thinking	○	○	○	○	○
They can see an experience from different standpoints	○	○	○	○	○
They take responsibility for what they say	○	○	○	○	○
They are open to discussion about their opinions	○	○	○	○	○
They are aware of their own limitations	○	○	○	○	○
They sometimes find themselves having difficulty in illustrating an ethical standpoint	○	○	○	○	○
They are aware of the cultural influences on their opinions	○	○	○	○	○
They want to understand themselves	○	○	○	○	○

Scoring: Items 3,4,8,12 17 and 21 are reverse scored

Self-reflection scale: 1,2,5,6,7,9,13,18,19,23

Empathetic reflection scale: 8,10,12,16,20,22

Reflective communication scale: 3,4,11,14,15,17,21

Correspondence from Author

GRAS Instructions (The Groningen Reflection Ability Scale)

Pleas note these instructions carefully before using the GRAS.

The GRAS score must be treated a single total score. The GRAS is a one-dimensional scale, although we found three subscales. These subscales can be used for research purposes only.

The GRAS can only be used for evaluations / decisions at group level (group mean score) and is not fit for *important* decisions at an individual level.

The items are scored on a 5 point Likert scale [1. totally disagree ~ 5. totally agree]

The scores of some negative formulated items must be reversed. These are the nrs. 3, 4, 8, 12, 17 and 21.

In order to avoid to the word 'reflection' in the instructions (avoidance of bias), we used the following instructions: "Instruction: Here are some statements about your learning and functioning in practice. Using the 1-5 scale below, please indicate what *really reflects* your approach rather than what you think your experience should be. Please treat each item separately from every other item."

We did not yet examine the relationship between the GRAS score and functioning in clinical practice (extern validity). We only compared the GRAS scores between experimental and control group (concurrent validity) and found a difference indeed. You can argue that, given this result, and given the assumption that reflection is important for learning in (clinical) practice, you may expect that a high GRAS score is a positive indication, and a low GRAS score is a negative indication for reflective learning and functioning in practice. We do not (yet) have a standard 'good' score.

Our data are based on Dutch items. Colleagues from the United States who used the GRAS, suggested some English textual fine-tunings which are accepted. German and Portuguese versions of the GRAS are available too.

We wish you success with using the GRAS and like to hear your experiences.

Leo Aukes

September 2010

Dr. L.C. Aukes
University Medical Center Groningen,
Center for Research and Innovation of Medical Education
PO Box 196 (Home postcode FC 40)
9700 AD Groningen, The Netherlands
Private: l.c.aukes@gmail.com Mobile +31 6 543 070 58

Social Skills Inventory (SSI)

Due to copyright full measure not displayed

mind garden

www.mindgarden.com

To Whom It May Concern,

The above-named person has made a license purchase from Mind Garden, Inc. and has permission to administer the following copyrighted instrument up to that quantity purchased:

Social Skills Inventory

The three sample items only from this instrument as specified below may be included in your thesis or dissertation. Any other use must receive prior written permission from Mind Garden. The entire instrument may not be included or reproduced at any time in any other published material. Please understand that disclosing more than we have authorized will compromise the integrity and value of the test.

Citation of the instrument must include the applicable copyright statement listed below.
Sample Items:

I usually feel uncomfortable touching other people.

I am interested in knowing what makes people tick.

I love to socialize.

Sincerely,

Robert Most
Mind Garden, Inc.
www.mindgarden.com

Interpersonal Problems Inventory (IIP-32)

Removed to ensure copyright compliance.

Appendix F: Cronbach's Alpha Outputs

Social Skills Inventory (SSI)

Reliability Statistics

Cronbach's Alpha	N of Items
.858	90

Interpersonal Problems Inventory (IIP-32)

Reliability Statistics

Cronbach's Alpha	N of Items
.869	32

Groningen Reflective Ability Scale (GRAS): Self-reported Version

Reliability Statistics

Cronbach's Alpha	N of Items
.737	23

Self-reflection

Reliability Statistics

Cronbach's Alpha	N of Items
.557	23

Empathetic Reflection

Reliability Statistics

Cronbach's Alpha	N of Items
.516	23

Reflective Communication

Reliability Statistics

Cronbach's Alpha	N of Items
.358	23

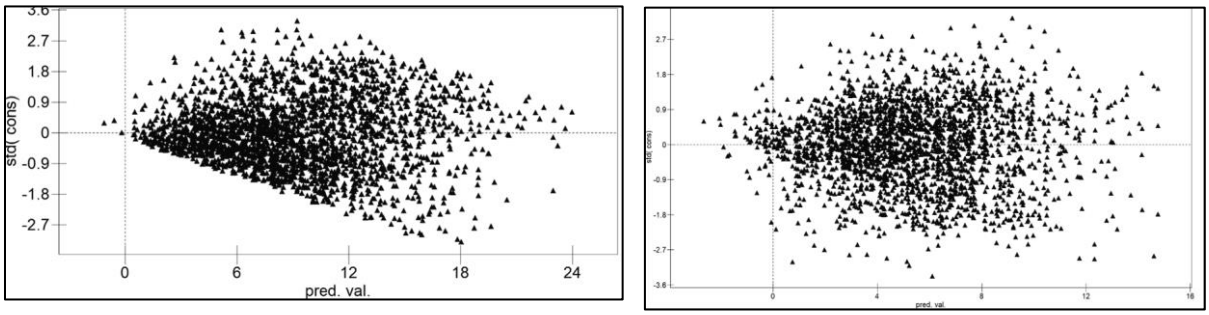
Groningen Reflective Ability Scale (GRAS): Supervisor-reported Version

Reliability Statistics

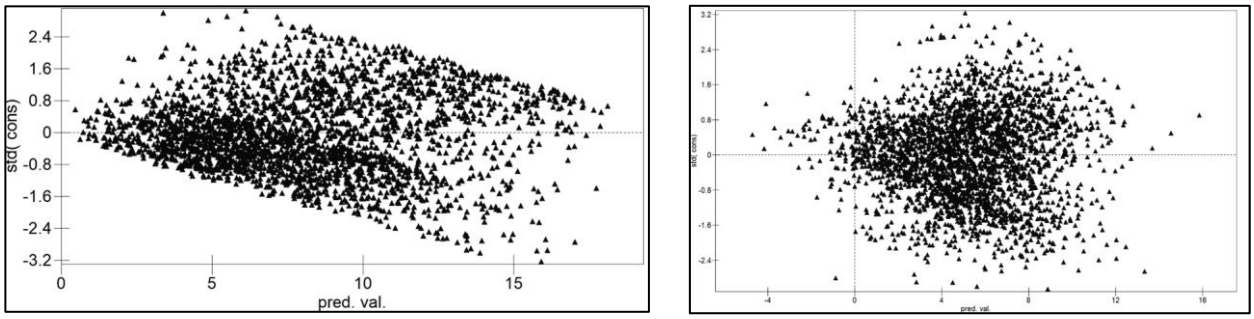
Cronbach's Alpha	N of Items
.935	23

Appendix G: Heteroskedasticity Assumption

Left image illustrates problem with heteroskedasticity when using PHQ-9 last score as the dependent variable. Right image illustrates this was corrected when PHQ-9 change score was used as the dependent variable.



Left image illustrates problem with heteroskedasticity when using GAD-7 last score as the dependent variable. Right image illustrates this was corrected when GAD-7 change score was used as the dependent variable.



Appendix H: MLwiN Model Outputs

Output dictionary	
cons	Intercept
(PHQ9_first_@_PT-gm)	Baseline PHQ-9
ProfessionalRole3Counsellor and Counsellor	Counsellor compared to CBT
ProfessionaRole3PWP and PWP	PWP compared to CBT
(PHQ9_first_@_PT-gm).PWP	Baseline PHQ-9*PWP compared to CBT and counsellor grouped
TrustID_1	NHS Trust
BAME	Black, Asian and Minority ethnicity compared to white ethnicity
Unemployed	Unemployed at referral to service
(Patient_Age-gm)	Patient age
(TotalIIP32_Self-gm)	IIP32- full scale
(SSI_Self_Total-gm)	SSI- full scale
(GRAS_Self_Total-gm)	GRAS- Self full scale
(SSI_Self_EE-gm)	SSI- Emotional Expressivity subscale
(SSI_Self_EC-gm)	SSI- Emotional Control subscale
(SSI_Self_ES-gm)	SSI- Emotional Sensitivity subscale
(SSI_Self_SC-gm)	SSI- Social Control subscale
(SSI_Self_SE-gm)	SSI- Social Expressivity subscale
(SSI_Self_SS-gm)	SSI- Social Sensitivity subscale
(GRAS_Self_SelfReflection-gm)	GRAS Self- Self Reflection subscale
(GRAS_Self_EmpatheticReflection-gm)	GRAS Self- Empathetic Reflection subscale
(GRAS_Self_ReflectiveCommunication-gm)	GRAS Self- Reflective Communication subscale
Patient_Gender:Female	Female compared to male
GRAS_Sup_Total-gm)	GRAS Supervisor- full scale
Gm	Grand mean centred
[Variable].[Variable]	Interaction term
B	Beta Coefficient
S.E	Standard Error
P	Probability Value

Depression: Case-mix Model

	B	S.E	p
cons	7.112	0.381	<0.001
(PHQ9_first_@_PT-gm)	0.607	0.033	<0.001
Counsellor	1.613	0.49	0.001
PWP	-1.232	0.418	0.003
(PHQ9_first_@_PT-gm).PWP	-0.138	0.036	<0.001
TrustID_1	-2.594	0.388	<0.001
BAME	-1.053	0.4	0.008
TrustID_1.Unemployed	1.485	0.565	0.009
(PHQ9_first_@_PT-gm).BAME	-0.191	0.065	0.003
(PHQ9_first_@_PT-gm).TrustID_1	-0.174	0.039	<0.001
(PHQ9_first_@_PT-gm).Unemployed	-0.152	0.045	0.001
Unemployed	-2.124	0.313	<0.001
(Patient_Age-gm)	0.015	0.007	0.032
ICC	0.0269328		
Therapist Effect	2.69%		

Depression: Sensitivity Analysis

	B	S.E	p
cons	7.288	0.32	<0.001
(PHQ9_first_@_PT-gm)	0.611	0.033	<0.001
Counsellor	1.42	0.406	<0.001
PWP	-1.233	0.345	<0.001
(PHQ9_first_@_PT-gm).PWP	-0.141	0.037	<0.001
TrustID_1	-3.146	0.355	<0.001
BAME	-1.058	0.419	0.012
TrustID_1.Unemployed	1.436	0.604	0.017
(PHQ9_first_@_PT-gm).BAME	-0.196	0.069	0.005
(PHQ9_first_@_PT-gm).TrustID_1	-0.192	0.041	<0.001
(PHQ9_first_@_PT-gm).Unemployed	-0.16	0.047	0.001
Unemployed	-2.11	0.312	<0.001
(Patient_Age-gm)	0.016	0.007	0.022
(TotalIIP32_Self-gm)	-0.005	0.009	0.579
(SSI_Self_Total-gm)	0.015	0.006	0.012
(GRAS_Self_Total-gm)	0.07	0.031	0.024

Depression: Subscale Exploratory Models

	B	S.E	p	B	S.E	p
cons	7.085	0.353	<0.001	7.002	0.364	<0.001
(PHQ9_first_@_PT-gm)	0.613	0.033	<0.001	0.604	0.033	<0.001
Counsellor	1.161	0.514	0.024	1.416	0.464	0.002
PWP	-1.035	0.393	0.008	-1.067	0.388	0.006
(PHQ9_first_@_PT-gm).PWP	-0.146	0.037	<0.001	-0.135	0.036	<0.001
TrustID_1	-2.652	0.383	<0.001	-2.481	0.372	<0.001
BAME	-1.046	0.419	0.013	-1.053	0.399	0.008
TrustID_1.Unemployed	1.34	0.605	0.027	1.453	0.565	0.010
(PHQ9_first_@_PT-gm).BAME	-0.198	0.069	0.004	-0.192	0.065	0.003
(PHQ9_first_@_PT-gm).TrustID_1	-0.191	0.041	<0.001	-0.176	0.039	<0.001
(PHQ9_first_@_PT-gm).Unemployed	-0.158	0.047	0.001	-0.151	0.045	0.001
Unemployed	-2.113	0.312	<0.001	-2.104	0.313	<0.001
(Patient_Age-gm)	0.016	0.007	0.022	0.014	0.007	0.046
(SSI_Self_EE-gm)	-0.017	0.026	0.513			
(SSI_Self_EC-gm)	0.016	0.024	0.505			
(SSI_Self_ES-gm)	0.035	0.031	0.259			
(SSI_Self_SC-gm)	0.043	0.027	0.111			
(SSI_Self_SE-gm)	0.024	0.024	0.317			
(SSI_Self_SS-gm)	-0.009	0.021	0.668			
(GRAS_Self_SelfReflection-gm)	-0.144	0.066	0.029	-0.084	0.069	0.223
(GRAS_Self_EmpatheticReflection-gm)	0.316	0.088	0.000	0.285	0.094	0.002
(GRAS_Self_ReflectiveCommunication-gm)	-0.029	0.077	0.706			

Depression: Supervisor Model

	B	S.E	p
cons	6.948	0.566	<0.001
(PHQ9_first_@_PT-gm)	0.582	0.049	<0.001
Counsellor	1.459	0.955	0.127
PWP	-1.215	0.685	0.076
(PHQ9_first_@_PT-gm).PWP	-0.12	0.058	0.039
TrustID_1	-2.803	1.006	0.005
BAME	-1.796	0.748	0.016
TrustID_1.Unemployed	-0.697	1.716	0.685
(PHQ9_first_@_PT-gm).BAME	-0.135	0.122	0.268
(PHQ9_first_@_PT-gm).TrustID_1	-0.172	0.105	0.101
(PHQ9_first_@_PT-gm).Unemployed	-0.18	0.065	0.006
Unemployed	-2.129	0.38	<0.001
(Patient_Age-gm)	0.009	0.009	0.317
(GRAS_Sup_Total-gm)	0.016	0.026	0.538

Anxiety: Case-mix Model

	B	S.E	p
cons	7.102	0.377	<0.001
(PHQ9_first_@_PT-gm)	-0.201	0.024	<0.001
TrustID_1	-1.642	0.355	<0.001
Unemployed	-1.778	0.28	<0.001
ProfessionalRole3:Counsellor	1.14	0.45	0.011
ProfessionalRole3:PWP	-1.001	0.383	0.009
(GAD7_first_@_PT-gm)	0.672	0.027	<0.001
Patient_Gender:Female	-0.501	0.198	0.011
(Patient_Age-gm)	0.018	0.006	0.003
(GAD7_first_@_PT-gm).(PHQ9_first_@_PT-gm)	-0.015	0.003	<0.001
(GAD7_first_@_PT-gm).Unemployed	-0.152	0.049	0.002
(PHQ9_first_@_PT-gm).TrustID_1	-0.096	0.034	0.005
TrustID:TrustID_1.Unemployed	1.492	0.512	0.004
BAME	-0.736	0.35	0.035
ICC	0.027771		
Therapist Effect	2.78%		

Anxiety: Sensitivity Analysis Model

	B	S.E	p
cons	7.423	0.324	<0.001
(PHQ9_first_@_PT-gm)	-0.202	0.024	<0.001
TrustID_1	-2.286	0.317	<0.001
Unemployed	-1.791	0.279	<0.001
ProfessionalRole3:Counsellor	0.843	0.364	0.021
ProfessionalRole3:PWP	-1.055	0.307	0.001
(GAD7_first_@_PT-gm)	0.676	0.027	<0.001
Patient_Gender:Female	-0.65	0.203	0.001
(Patient_Age-gm)	0.02	0.006	0.001
(GAD7_first_@_PT-gm).(PHQ9_first_@_PT-gm)	-0.016	0.003	<0.001
(GAD7_first_@_PT-gm).Unemployed	-0.155	0.051	0.002
(PHQ9_first_@_PT-gm).TrustID_1	-0.112	0.036	0.002
TrustID:TrustID_1.Unemployed	1.67	0.549	0.002
BAME	-0.775	0.363	0.033
(GRAS_Self_Total-gm)	0.082	0.028	0.003
(SSI_Self_Total-gm)	0.008	0.005	0.110
(TotalIIP32_Self-gm)	-0.003	0.008	0.708

Anxiety: Subscale Exploratory Models

	B	S.E	p	B	S.E	p
cons	7.306	0.35	<0.001	6.975	0.347	<0.001
(PHQ9_first_@_PT-gm)	-0.203	0.024	<0.001	-0.2	0.024	<0.001
TrustID_1	-1.737	0.344	<0.001	-1.52	0.324	<0.001
Unemployed	-1.778	0.279	<0.001	-1.754	0.279	<0.001
ProfessionalRole3:Counsellor	0.501	0.462	0.278	0.936	0.402	0.020
ProfessionalRole3:PWP	-0.977	0.352	0.006	-0.807	0.334	0.016
(GAD7_first_@_PT-gm)	0.677	0.027	<0.001	0.67	0.027	<0.001
Patient_Gender:Female	-0.671	0.203	0.001	-0.512	0.198	0.010
(Patient_Age-gm)	0.02	0.006	0.001	0.018	0.006	0.003
(GAD7_first_@_PT-gm).(PHQ9_first_@_PT-gm)	-0.016	0.003	<0.001	-0.015	0.003	<0.001
(GAD7_first_@_PT-gm).Unemployed	-0.159	0.051	0.002	-0.153	0.049	0.002
(PHQ9_first_@_PT-gm).TrustID_1	-0.109	0.036	0.002	-0.098	0.034	0.004
TrustID:TrustID_1.Unemployed	1.565	0.55	0.004	1.457	0.512	0.004
BAME	-0.751	0.363	0.039	-0.731	0.349	0.036
(SSI_Self_EC-gm)	-0.001	0.022	0.964			
(SSI_Self_EE-gm)	-0.034	0.023	0.139			
(SSI_Self_ES-gm)	0.036	0.028	0.199			
(SSI_Self_SC-gm)	0.031	0.024	0.196			
(SSI_Self_SE-gm)	0.017	0.022	0.440			
(SSI_Self_SS-gm)	-0.007	0.019	0.713			
(GRAS_Self_SelfReflection-gm)	-0.123	0.059	0.037	-0.096	0.059	0.104
(GRAS_Self_EmpatheticReflection-gm)	0.311	0.078	<0.001	0.318	0.081	<0.001
(GRAS_Self_ReflectiveCommunication-gm)	0	0.069	1.000			

Anxiety: Supervisor Model

	B	S.E	p
cons	7.228	0.562	<0.001
(PHQ9_first_@_PT-gm)	-0.204	0.03	<0.001
TrustID_1	-1.952	0.925	0.035
Unemployed	-1.705	0.329	<0.001
ProfessionalRole3:Counsellor	0.937	0.892	0.294
ProfessionalRole3:PWP	-0.995	0.638	0.119
(GAD7_first_@_PT-gm)	0.638	0.038	<0.001
Patient_Gender:Female	-0.835	0.278	0.003
(Patient_Age-gm)	0.016	0.008	0.046
(GAD7_first_@_PT-gm).(PHQ9_first_@_PT-gm)	-0.019	0.005	<0.001
(GAD7_first_@_PT-gm).Unemployed	-0.204	0.073	0.005
(PHQ9_first_@_PT-gm).TrustID_1	-0.093	0.085	0.274
TrustID:TrustID_1.Unemployed	0.38	1.509	0.801
BAME	-1.066	0.616	0.084
(GRAS_Sup_Total-gm)	0.02	0.025	0.424