Exploring Treatment Integrity and Drop-Out Rates in Low-Intensity CBT-based Interventions

Jonah Gosling

A thesis submitted in partial fulfilment of the requirements for the award of Doctor of Clinical Psychology at The University of Sheffield

The University of Sheffield

Faculty of Science

Clinical Psychology Unit, Department of Psychology

May 2022

Declaration

I declare that this thesis has been submitted for the award of Doctorate in Clinical Psychology at the University of Sheffield. It has not been submitted for any other qualification or to any other academic institution.

Word Count

Section One: Literature Review

Excluding tables, references and appendices	7,660
Including tables and references, excluding appendices	11,751

Section Two: Empirical Research Report

Excluding tables, references and appendices	8,000
Including tables and references, excluding appendices	11,694

Total

Excluding tables, references and appendices	15,660
Including tables and references, excluding appendices	23,445

Acknowledgements

Firstly, I would like to thank the participants. I would also like to thank the IAPT service for allowing us to conduct the research and for providing me with timely assistance. Thanks to my supervisors, Dr Stephen Kellett and Dr Melanie Simmonds-Buckley, who were available, supportive and encouraging throughout the lengthy research process. Also, many thanks to Dan and Kasia for going the extra mile and helping us with recruitment, logistics and general research planning.

Thank you to the psychYOLOgists group for your humour and grounding influence. Many thanks to Kane, who never stops calling me even when I go missing, and Nick, for always laughing with me like we're still 11. Also, a big thank you to my Mum and Dad, who both taught me the importance of hard work, resilience and persistence. All of which have been invaluable. To my son, Ezra, thanks for our countless trips to feed the horses and early 5am wake-ups! You are my inspiration and you have no idea how much you motivate me to make you proud. And finally, a huge thank you to my wife, Jenny, who feels as though she has completed the doctorate herself whilst supporting me and our young family. For that I am forever grateful and will bestow on her the title, Dr Jenny Gosling. Jenny, you are my rock and without you, this would not have been possible.

Lay Summary

Despite psychological interventions increasingly being used to provide accessible psychotherapy to the general population, the number of people dropping out of the support and the differences in their improvements requires research attention. Thus, this thesis aimed to investigate these phenomena by 1) combining the findings of multiple studies investigating the drop-out rate of patients accessing low-intensity interventions for depression in routine care, and 2) conducting a study developing and evaluating a facilitator group psychoeducational measure.

A key component of whether a psychological intervention is effective is related to how often patients attend and/or drop-out. The drop-out rate for depression interventions has been researched, but this had mainly been under controlled study conditions. Unfortunately, these findings are unlikely representative of care provided by services in the real-world. Therefore, section one of this thesis reports a review, which combines the findings of 11 studies investigating the drop-out rate of patients accessing low-intensity interventions for depression in routine care. The review calculated a combined drop-out rate and investigated factors that may impact variations in drop-out rates between studies. Finally, the combined drop-out rate in this study was compared to other studies combining research for different treatments. Differences between the drop-out rates of the studies in this review were heavily influenced by country and therapy format. The current reviews' drop-out rate was comparable to others within the field, but was considerably higher than one review that only included studies under controlled conditions. Findings propose that services should be wary of patients dropping out of low-intensity interventions for depression (perhaps more than existing research indicates), whilst looking to use strategies to reduce drop-out likelihood.

Differences in how much patients improve after accessing therapeutic groups is thought to be moderated by the facilitator's delivery. More specifically, a combination of how closely they follow the treatment manual, how skillfully they deliver the content and how much they differentiate the treatment from others (also known as 'treatment integrity'). However, a measure to mark the treatment integrity of facilitators delivering such groups does not exist. Consequently, section two of the thesis is a study that developed and evaluated a novel group psychoeducational treatment integrity measure (GPTIM). A measure and manual were developed for the measure after consulting key references. Experts were consulted on the relevance of the measure items, prior to another batch of experts trialing the measure on a batch of recorded sessions. Data collected from the trial were then used to test the reliability and validity of the GPTIM. Results indicated that the measure had an array of acceptable properties, but scores between raters for the same recorded session were poorly correlated. Findings indicate that the GPTIM demonstrated promising potential, but requires fine-tuning prior to its use within routine service, training and research contexts. Limited sample sizes and the brevity of the rater training may have impacted the conclusions and require further exploration.

Taken together, the two studies aim to inform and improve the psychological support that services and therapists deliver. One study provides clarity regarding 'real-world' drop-out rates, whilst the other presents a novel treatment integrity measure that (with further testing) could be utilised by services to aid psychoeducational group delivery. Research with an increased number of studies should further investigate the drop-out rates of low-intensity interventions for depression within routine care, which could provide practicable strategies to support its acceptability. Furthermore, additional research is also required to fine-tune the GPTIM, which could result in its application within routine care settings.

Contents

Section One: Literature Review

The acceptability of low intensity interventions for depression delivered in routine practice: Systematic review and proportional meta-analysis

Section Two: Empirical Research Report

Developing and evaluating a measure of

treatment integrity for the delivery of group psychoeducation for

common mental health problems

This page is intentionally blank

Section One: Literature Review

The acceptability of low intensity interventions for depression delivered in routine practice: Systematic review and proportional meta-analysis

Abstract

Objectives

Although low intensity interventions for depression are widely provided in routine services, their acceptability to patients has not been systematically reviewed. Therefore, this paper aimed to review the acceptability of low intensity interventions for depression in routine care settings and benchmark the meta-analysed dropout rate against other commonly delivered treatments for depression.

Design

A pre-registered (osf.io/2qhpf/) systematic review and meta-analysis.

Method

Studies gathering routine care outcome data for patients receiving low intensity interventions for depression and reporting drop-out rates were synthesised. Eleven studies were identified through comprehensive systematic searches and were quantitatively synthesised using a proportional random-effects meta-analysis. A moderator analysis explored variations in effect sizes for each study. The dropout rate for LI treatments was then benchmarked versus other meta-analysed dropout rates from other therapies in order to provide context.

Results

The pooled treatment drop-out rate was 28% (*k*=11; 95% CI's 23 to 33). Country and therapy format were significant drop-out rate moderators. Benchmarking demonstrated that

the LI dropout rate was generally comparable to the dropout rates generated by other therapies.

Conclusions

Dropout from LI treatment for depression continues to be a concern, despite the benchmarking evidence. The limitations of the review and the clinical and research implications are also explored.

Practitioner Points and Limitations

- Practitioners should be aware of the possibility of patients dropping out of LI interventions for depression.
- Services should identify patients most at risk of drop-out (i.e. those with multiple risk factors such as being male, younger, having co-morbid anxiety etc.) and attempt to employ strategies to reduce these figures (i.e. more therapist check-ins, strengthening patient hope, discussing treatment expectations, enhancing patient motivation etc.).
- Arbitrary cut-offs were created for the risk of bias tool, which may have implications regarding the resultant moderator analysis.
- Interventions to reduce drop-out need to be developed that are matched to the intensity of the intervention being delivered.

Keywords

Acceptability; Drop-out rates; Low-intensity interventions; Depression; Psychotherapy.

Introduction

The acceptability of psychotherapeutic interventions should be considered when attempting to effectively design and implement any psychological approaches and acceptability outcomes are a key component of the evidence base for any intervention. Sekhon, Cartwright & Francis (2017) defined acceptability as a necessary, but not sufficient, condition for the evidence base for any healthcare intervention. Being able to usefully deliver a psychological intervention in routine care services therefore depends heavily on how patients experience the intervention in terms of its content and the context of delivery. When interventions are acceptable to patients, then treatment adherence typically follows, resulting in the associated improved clinical outcomes (Homel et al., 2013). The main way in which the acceptability of psychological interventions has been indexed is to assess treatment refusal and treatment drop-out rates (Kriston et al., 2014). This assumes that patients that refuse treatment or drop-out of treatment are unsatisfied with the treatment offered and therefore the treatment has low acceptability to them. Unfortunately, there is a tendency for research to combine treatment refusal and drop-out rates, which results in the concealment of individual rates (Santana et al., 2013).

Treatment refusal and treatment drop-out is challenging for patients, healthcare providers, researchers and society (Hunsley, 2003). In comparison to those who have completed treatment, patients who drop-out are less likely to clinically improve (Klein et al., 2003; Lopes et al., 2018; Melartin et al., 2005) and report higher levels of dissatisfaction (Bjork et al., 2009; Knox et al., 2011). Drop-out is also associated with burnout and poor morale for service providers (Barrett et al., 2008) and creates service inefficiency through incomplete or ineffective ‘treatment episodes’ creating a ‘revolving door’ (Barrett et al., 2008; Swift & Greenberg, 2014). Resource allocation is heavily impacted by drop-out and it can

result in poorer long-term outcomes for patients resulting in a higher societal cost (Reis & Brown, 1999; Cooper et al., 2018). Treatment refusal and drop-out remain issues regardless of the type of psychological intervention being delivered (Nathan & Gorman, 2015). However, certain interventions delivered for particular mental health difficulties, for example cognitive behaviour therapy (CBT) for depression, have been found to have higher drop-out rates and therefore significant differences between approaches have been observed (Cuijpers et al., 2008)

Treatment refusal is widely defined as those instances in which patients are offered an intervention at a screening appointment, but then fail to begin treatment (Swift & Greenburg, 2015). There is less consensus regarding the definition of treatment drop-out, creating difficulties with cross-study comparisons (Zimmerman et al., 2017). The most common definition is when a patient unilaterally terminates an intervention prematurely, either before recovering from the problems that created the need for support or before completing the intervention's specified protocol (Swift & Greenberg, 2014). Different ways to measure and operationalise drop-out are found across the breadth of the literature (Swift & Greenberg, 2012). Although each definition has its advantages and disadvantages, drop-out rates therefore differ from study to study and therefore findings should be interpreted with caution in reference to the specific applied criteria (Hatchett & Park, 2003; Zimmerman et al., 2017). Swift and Greenburg (2014) noted that drop-out has been operationalised variously as a specified number of sessions (e.g. attended less than 3 sessions), failure to complete a treatment protocol (e.g. attending less than the prescribed sessions), failure to show for a scheduled appointment, and based on a therapist's judgement or the presence/absence of clinically significant change (e.g. not moved into the non-clinical range)

When studies investigating drop-out have been combined via proportional meta-analyses, contrasting results have also been found. In an early attempt, Wierzbicki & Pekarik (1993) analysed 125 studies and reported a mean pooled drop-out rate of 47%. However, it has been suggested that substantial methodological and statistical issues may have impacted the findings (Cooper & Conklin, 2015). When using more advanced meta-analytical techniques, study quality assessments and improved standards of reporting, recent major reviews have revised down the average drop-out rate to 20% of patients (Fernandez et al., 2015; Swift & Greenberg, 2012). Whilst some moderator analyses have found that drop-out rates did not differ according to diagnostic groups (Leichsenring et al., 2019) or psychotherapeutic approaches (Cooper & Conklin, 2015; Simmonds-Buckley et al., 2019; Swift & Greenberg, 2014), others have reported higher drop-out rates for the treatment of personality disorders, eating disorders and post-traumatic stress disorder (Imel et al., 2013; Swift & Greenberg, 2014).

Depression is a common mental health problem and its effective management is a central challenge for healthcare providers worldwide (Bower et al., 2013; Sinyor et al., 2016). The American Psychiatric Association (APA; 2013) characterises depression as a serious mood disorder that can result in one experiencing less enjoyment in activities, difficulties concentrating, persistent sadness, trouble sleeping and more. Treatments for depression continue to be a prominent topic within clinical research, both due to the condition's omnipresence within the general population and the fact that many patients experience a recurrence of symptoms or fail to respond to treatment (Cox et al., 2014). Meta-analyses have also been conducted investigating the acceptability, alongside the effectiveness, of interventions for depression (Cujipers et al., 2019; Cujipers et al., 2020). Such reviews typically only included randomised controlled trials (RCTs) within their analyses and often

discounted routine care data gathered outside of a controlled environment. The principal issue with this relates to the distinction between evidence-based practice (EBP; the act of basing professional practice on efficacy research, such as RCTs) and practice-based evidence (PBE; the act of deriving professional practice from real-world findings, such as routine care data). Since the 1980's, the EBP movement has heavily driven policy decisions in western cultures, despite condemnation from those maintaining the importance and crucial role of PBE (Barkham & Mellor-Clark, 2003). Many question the clinical applicability of EBP research, affirming that effective findings within clinical trials are not always translated to the real-world (Green, 2006). The reasons why an individual drops-out of a RCT delivered treatment, may be different to the reasons that contribute to drop-out in routine services. Crucially, although EBP is necessary, it is not an adequate sole condition for delivering an evidence base within practice settings (Bower, 2003). Expansion of PBE research can fill the gap between recommended care and improved care, whilst discovering the efficacy of interventions and approaches within routine care settings (Horn & Gassaway, 2007; Westfell, 2007). The treatment of mild to moderate depression has gained a lot of attention within comparative research (meta-analyses and systematic reviews; Cuijpers et al., 2008), but such findings have concentrated on EBP and there is a lack of clarity and exploration around drop-out rates (especially highlighting drop-out rate differences between control trials and routine care data).

In response to the incidence of depression within society, low-middle and high income countries have started to provide low-intensity (LI) interventions that are delivered by non-specialists/lay therapists (Bockting et al., 2016). With the increased emergence and uptake of these interventions, literature relating to LI-type CBT approaches started to materialise; which allowed for the identification of shared features (Bennett-Levy et al., 2010). Although not the

case for all LI interventions, common aspects of these approaches include brevity, minimal professional contact/lay therapist delivery, delivery via remote resources (i.e., book, leaflet, the internet) and a cognitive behavioural underpinning (Bockting et al., 2016; Bower et al., 2013; British Psychological Society, 2011). Research has found that such interventions are effective for treating depression (Andrews et al., 2018; Chowdhary et al., 2015; Cuijpers et al., 2010; Rahman et al., 2008). However, acceptability evidence has found that at least a quarter of patients accessing LI interventions drop-out of treatment (Richards & Borglin, 2011). With large numbers of patients currently accessing LI interventions for depression, an increased significance has now been placed upon understanding why drop-out rates for these specific interventions are so high. Improvements in our awareness of this phenomenon could allow us to stem drop-out rates during the earlier stages of treatment and create more efficient and effective services (Delgadillo et al., 2013).

A way of comparing acceptability evidence is to complete a meta-analysis, which synthesises all available evidence within a given subject area. Meta-analyses based on routine care data typically have to employ different data synthesis techniques due to the absence of control groups (e.g., present in RCTs). An approach which is regularly employed when researchers are combining data from single-group studies (such as the data drawn from routine care), is a proportional meta-analysis (El Dib et al., 2013). This method produces a pooled overall proportion that is calculated from individual studies reporting data dichotomously or in the form of percentages (Barker et al., 2021). Importantly, proportional meta-analyses are recommended for systematic reviews investigating prevalence and have been highlighted as useful in providing answers to under investigated findings (Barker et al., 2021). This review follows the Barker et al (2021) guidelines for a proportional meta-analysis and therefore limited the scope of the review to a single diagnosis in order to increase

specificity. This approach also countered the issue of much of the LI evidence base being based on mixed anxiety and depression samples.

So, the primary objective of this review was to complete a quantitative proportional meta-analysis of the treatment drop-out rates for LI interventions for depression within routine practice. This was the first attempt to estimate the treatment acceptability of depression specific to LI treatments delivered in routine care settings. As part of this process, the secondary objective was to benchmark the LI depression intervention drop-out rate against other depression interventions and psychotherapeutic approach rates. The specific aims of the meta-analysis were to: 1) estimate a pooled proportion of treatment drop-out rates for LI interventions for depression using a proportional meta-analysis approach, 2) to explore variables that might moderate drop-out rates for LI interventions for depression, 3) benchmark the pooled LI intervention for depression drop-out rate against other meta-analysed drop-out rates from other depression interventions and psychotherapeutic approaches.

Method

Study Identification and Eligibility

This meta-analytic review was pre-registered on Open Science Platform Home (OSFHOME) platform and is presented in accordance with PRISMA guidelines (Page et al., 2021). The PICO (population, intervention, comparison, outcome) tool was utilised to develop the review question and determine the search strategy (see Appendix A) and following this, a literature search was conducted to source eligible studies. The three electronic databases used were Scopus, Web of Science and Ovid by Medline. All searches were completed in line with pre-defined inclusion/exclusion criteria and were conducted between 1st and 8th

November 2021. No publication date limits were applied to the searches. Example search terms and Boolean logic used are displayed in Table 1. Within these search terms, truncation (*) was used to guarantee the inclusion of word variations. Alongside database searches, forward and reverse citation searches were conducted and hand searching of related literature was also performed. The terms “drop-out” and “attrition” were not included as search terms, as the majority of suitable studies did not include these terms within their title, abstracts or as key words. Thus, in an effort to widen the scope of the review, these terms were omitted.

Table 1

Search Terms Used in Database Searches

	Filter 1: CMHP	Filter 2: Psychotherapy	Filter 3: Data Source
Search Fields	**Title, abstract or key words**	**Title, abstract or key words**	**Title, abstract or key words**
	“depress**” “mood”	“psychoeducation” “therapy” “self care” “self help” “self management” “bibliotherapy”	“routine care/delivery/practice” “usual care/delivery/practice” “standard care/delivery/practice”

Selection of Studies

Study selection was independently conducted by one reviewer (JG), who screened the titles and abstracts of all potential studies. The reviewer selected the studies based upon the inclusion and exclusion criteria. Studies which appeared to meet the inclusion and exclusion criteria were further screened by reading the full text articles, resulting in a preliminary ‘included studies list’. Once this initial selection process was complete, an independent

reviewer (MSB) screened 30% of titles and abstracts from the included full-text studies. Agreement between the reviewer and the independent reviewer was found to be perfect (100%) for full text screening. All included studies 1) were conducted on a 16+ years old population, 2) were either individual or group psychological intervention, 3) were either face-to-face or virtual/video/online, 4) provided treatment for the primary presenting problem of depression (diagnosis or diagnosis not required), 5) evaluated a low-intensity intervention (e.g. pure self-help, guided self-help, computerised CBT etc.), 6) was delivered in routine care, 7) reported the drop-out rate and 8) was available in English. Exclusion criteria were 1) unpublished research, 2) clinical trials (including randomised controlled trials), 3) single case experimental designs or individual case studies, 4) high-intensity interventions, 5) drop-out rate not reported or unable to be calculated and 6) the article was unavailable in English.

Data Extraction

The primary author extracted data from the included studies using a bespoke data extraction instrument which was piloted and found to be reliable. Extracted data included 1) study details (primary author, year of publication), 2) country, 3) study design, 4) treatment setting, 5) mean age (standard deviation [range]), 6) percentage of males, 7) therapy duration (therapy format, [number of sessions attended]), 8) treatment sample number, 9) drop-out cases (proportion), 10) drop-out definition and 11) risk of bias. To assess the robustness of the extraction practices and the bespoke extraction tool, a proportion of the included studies (*k*= 4; 36%) were extracted by a second independent reviewer (MSB) who was unaware of the first reviewers extractions. The agreement rate for this process was 90%. Any extractions

that were dissimilar were jointly considered and the final extractions were based upon these discussions.

Outcomes

Due to the previously acknowledged variability regarding a fixed and universally agreed definition of ‘treatment drop-out’, the author operationalised this for the purposes of the review. The definition used was: the proportion of a studies sample that started treatment who subsequently dropped-out, in accordance with the studies classification of what drop-out is. According to this definition, for all studies, treatment drop-out rates were extracted. On occasion, studies may not provide a definition of treatment drop-out. In such instances, patients attending less than 50% of sessions will be classified as dropping out. As a result of the review definition, treatment drop-out incorporates patients discontinuing treatment and patients failing to complete pre-defined assessment measures. Previous studies have distinguished between these two definitions (i.e. Dixon & Linardon, 2020), but this distinction is less applicable for routine care data. However, due to the recognised differences between these, sensitivity analysis were performed for all of the studies to assess the effect of drop-out type on the pooled proportions. The two groups included within the drop-out rate definition sensitivity analysis were 1) did not complete treatment, and 2) service user who did not complete post-test.

Risk of Bias

To assess for risk of bias (RoB), the Psychotherapy Outcome Study Methodology Rating Form (POSMRF; Öst, 2008; see Appendix B) evaluated the methodological quality of

included studies. The 22-item measure is based on the work of Tolin (1999) and was chosen as it is appropriate for rating study quality across a wide variety of research designs. The scale has been shown to have good internal consistency and good inter-rater reliability (Öst, 2008). The measure assesses: (1) clarity of sample description, (2) severity/ chronicity of the disorder, (3) representativeness of the sample, (4) reliability of the diagnosis, (5) specificity of outcome measures, (6) reliability and validity of outcome measures, (7) use of blind evaluators, (8) assessor training, (9) assignment to treatment, (10) design, (11) power analysis, (12) assessment points, (13) manualized, replicable, specific treatment programs, (14) number of therapists, (15) therapist training/ experience, (16) checks for treatment adherence, (17) checks for therapist competence, (18) control of concomitant treatments, (19) handling of attrition, (20) statistical analyses and presentation of results, (21) clinical significance, (22) equality of therapy hours between groups. Each item is rated as either poor (0), fair (1) or good (2) and so overall scores range between 0-44. As comparison groups were not available due to the design of the included studies, item 22 was omitted for this meta-analysis. Overall, this resulted in each study being rated between 0-42. The POSMRF does not have classification cut-offs. To ensure quality comparisons could be made between included studies, cut-offs were calculated in a comparable manner to other systematic reviews that have used the POSMRF (Sloan et al., 2017; Swain et al., 2013). This process involved the calculation of standard deviations (SD) for ‘well below average’ (<-1 SD), ‘below average’ (-1-0 SD), ‘above average’ (0-1 SD) and ‘well above average’ (>1). All POSMRF scores were rated by second raters (MSB, SK and MK). The level of agreement was ‘’substantial” (Kappa = 0.78) in line with the criteria set out by Landis and Koch (1977). For the ratings that were not fully agreed, a discussion occurred and a joint agreement was achieved before ratings were finalised.

Data Synthesis and Analyses

The analysis process was conducted using the MetaXL 2.0 add-on tool within the Microsoft Excel computer programme (available at www.epigear.com). From the extracted data, patients who dropped out of their intervention were calculated by dividing the number of drop-out cases by the total sample. Drop-out rates were synthesised using a random-effects single proportion meta-analysis using the inverse of the variance to weight individual studies. As a result of restrictions on variability estimates for proportional data that is situated near the extremes, the Freeman-Tukey (double arcsine) transformation was utilised before synthesis of the data (Barendregt et al., 2013). Proportional data situated near the extremes (which includes close proximity to 0.0 or 1.0) must be transformed, as it can bias the inverse variance weighting of studies. Back-transformation and percentage conversion were then completed for the proportions to provide a clearer interpretation of the weighted pooled drop-out rate.

Study heterogeneity can greatly affect meta-analytical conclusions and therefore homogeneity should not be assumed (Kontopantelis et al., 2013). Thus, all conducted analysis employed the Q and I^2 statistics to assess for study heterogeneity. The Q statistic was calculated as the significance level of Q indicates whether inconsistency between studies exceeds what would be expected from sampling error alone (Higgins et al., 2003). Despite this, the Q statistic is recognised as having low statistical power and is perceived as vulnerable to the number of studies included within the analysis (Jackson, 2006). To safeguard against this, the I^2 statistic was also used which calculates the extent of variability across the included studies that is not due to sampling errors (Harrer et al., 2021). The I^2 statistic was selected due to the methodological issues with purely calculating the Q statistic and for the fact that it is not sensitive to the number of studies included within the analysis

(Harrer et al., 2021). The I^2 classification values are >25% (low heterogeneity), >50% (medium heterogeneity) and >75% (high heterogeneity) (Higgins et al., 2003).

Publication bias was assessed by visually inspecting funnel plot asymmetry (Light & Pillemer, 1984). Publication bias relates to the preferential publication of studies with captivating, favourable and/or clear-cut results (Murtaugh, 2002). Although there are debates about the impact of publication bias affecting the publication of single proportional data (i.e., data is not dependent on significance levels), the current study conducted visual inspection of publication bias as it may be possible that studies reporting higher drop-outs may not have been reported and/or published (Wang, 2018). The presence of publication bias can result in the findings of a meta-analysis being misleading and therefore an indication of its possible existence is imperative (Peters et al., 2006).

A moderator analysis was conducted to explore if particular variables explain variations in effect sizes for each individual study. For this study, the moderator analysis investigated the sources of heterogeneity between the included studies and was completed for six categorical study characteristics. Analysis of categorical variables was performed using a subgroup analysis, which is only likely to be a useful measure of heterogeneity when done on at least 10 studies (Deeks et al., 2019). The categorical variables included within the subgroup analysis investigated were 1) drop-out definition (did not complete treatment/did not complete post-test), 2) country (UK/Scandanavian/other), 3) format (therapist-assisted CCBT or iCBT/non therapist-assisted CCBT or iCBT/group CBT/other), 4) delivery type (internet/face-to-face), 5) patient setup (individual/group) and, 6) RoB (“Well below average” or “below average”/”well above average” or “above average”).

Benchmarking

The pooled LI depression treatment drop-out rate found within this meta-analysis was also benchmarked against other approaches to treating depression and other psychotherapy treatments in general. Benchmarking is the process of statistically comparing clinical outcomes against other psychotherapeutic outcomes (Lueger & Barkham, 2010). The same databases used for the main literature search (Scopus, Web of Science and Ovid by Medline) were also used to complete the benchmarking search. As part of the process, meta-analyses assessing a range of approaches to treat depression/subthreshold depression, along with those using specific therapeutic approaches (i.e. DBT, ACT etc.) for a range of presentations, were included. To benchmark the current drop-out rate against meta-analyses combining mixed approaches to treat depression, we included reviews investigating CBT for unipolar depression (Hans & Hiller, 2013), computerised CBT for depression (Kaltenthaler et al., 2008), general subthreshold treatment for depression (Cuijpers et al., 2007) and general depression treatment (Swift et al., 2014). Following this, four meta-analyses that calculated drop-out rates for different psychotherapies were included for benchmarking: CBT (Fernandez et al., 2015), dialectical behaviour therapy (DBT) (Dixon & Linardon, 2020), acceptance and commitment therapy (ACT) (Ong et al., 2018) and general psychotherapy (Swift et al., 2017). Once included meta-analyses were decided upon, data were extracted regarding the pooled drop-out rate and the confidence intervals (CIs) of the studies treatment group and were benchmarked against the LI depression treatment drop-out rate.

Results

Study Selection

A summary of the search strategy and findings are detailed in Figure 1 (PRISMA flowchart; Page et al., 2021). Electronic database searches identified 1730 studies after duplicates were removed. No additional studies were found through hand searching or from forward and reverse citation searchers. Following this, titles and abstracts were screened. Out of 1730 studies, 1670 were excluded (97%). The remaining 60 studies were then examined for inclusion via a full-text screening process and 49 studies were removed (see Appendix C for titles and assigned reasons for exclusions of all full-text articles). This process resulted in 11 studies deemed eligible for inclusion within the meta-analysis.

Summary Study Characteristics

The characteristics of the included routine care studies are reported in Table 2. The studies were mostly conducted in Scandanavia (*k*=5, 45%) and the United Kingdom (*k* = 2; 18%). The remaining single articles were conducted in Netherlands (9%), Israel (9%), Australia (9%) and Germany (9%). In terms of design, two utilised cohort designs (*k*=2; 18%), one was an effectiveness design (*k*=1; 9%) and the rest were purely pre-post designs (*k*=8; 73%). None of the studies reported a treatment refusal rate and the criteria for defining treatment drop-out varied (e.g. completing the first session but discontinuing the treatment at a later time point (*k*=8; 73%) and failure to complete the post-intervention/follow-up measures (*k*=3; 27%).

Treatment Characteristics

Settings included outpatient hospitals, outpatient psychiatric/psychological clinics, a secondary care service, an online mental health clinic, a stroke service, a specialist CBT service and a self-help service. Treatment was largely delivered based on the principles of CBT (*k*=10; 91%) and this was offered either on a one-to-one basis (*k*=9; 82%) or in a group (*k*=2; 18%). The two studies that provided group treatment also delivered the intervention face-to-face (18%), whilst the remaining nine studies were all delivered online (82%). Five of the included studies delivered therapist-assisted computerised CBT (CCBT) or internet-based CBT (iCBT) (45%), with two delivering CCBT or iCBT without therapist assistance (18%). Two studies delivered group CBT (18%), one study provided videoconference CBT (CBT; 9%) and another offered a self-help forum (9%). Treatment durations varied, with some providing brief (≤8 sessions; *k*=3; 27%), medium (9-12 sessions; *k*=1; 9%) and longer (13-16 sessions; *k*=3; 27%) contracts; although this was not always clear from the studies information. Four studies outlined that mixed intervention durations (36%) were used. Furthermore, the time duration of sessions varied in length for each study, ranging from 50-180 minutes.

Patient Characteristics

One study was conducted on a purely female sample, whilst all other studies had both male and female participants. Males made up 31% of the included studies sample (SD = 15.26). All but one study reported the average age of participants. Mean age was 41.64 years (SD = 8.91), with a range of 16-70 years. Seven studies (64%) reported that participants also presented with a comorbid anxiety disorder, with a range between 24.7% and 74.2%. One

study (9%) reported that 31.8% of their participants had ≥2 diagnoses, but did not define the diagnoses.

Risk of Bias Assessment

A mean quality rating score of 17.91 (SD = 3.09; range = 13-24) was found. In relation to the aforementioned POSMRF classification calculation, this created the following cut-offs: 'well below average' (<15; *k*=1), 'below average' (15-18; *k*=6), 'above average' (18-21; *k*=1) and 'well above average' (>21, *k*=3). Studies scored highest (mostly rating 'good') for the specificity, validity and reliability of outcome measures used. General high scores were also noted for the studies statistical analysis and presentation of their results and the use of manualised, replicable and specific treatment programs. RoB was observed most (mostly rated 'poor') for the absence of blind evaluators, lack of any power analysis and poor control of contaminant treatment (e.g. medications). In addition, items which also appeared to increase the RoB for most studies included treatment adherence and competency and assessor training.

Figure 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Diagram Summarising Screening (Page et al., 2021)

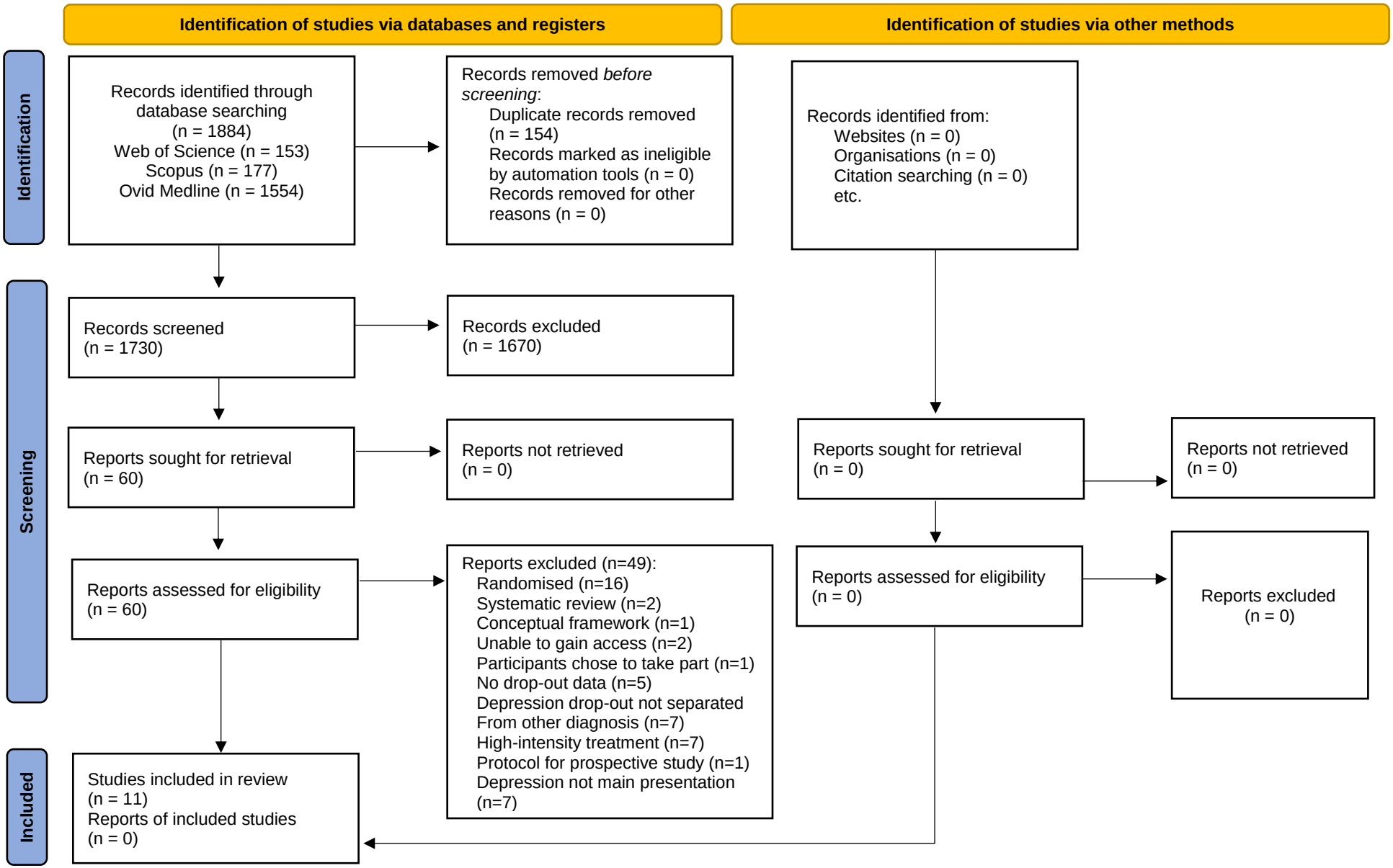


Table 2

Characteristics of Included Studies

Study, Author, Year	Country	Study design	Treatment setting	Mean age (SD) [range]	Sex (% male)	Treatment duration (format) [sessions]	Treatment Sample (n)	Drop-out cases (proportion)	Drop-out definition	Risk of bias (RoB)
Ruwaard et al. 2012	Netherlands	PBE (Uncontrolled pre-post study)	Online mental health clinic	40 (11)	32%	16-week iCBT treatment (therapist-assisted)	413	112 (0.27)	SUs who did not complete post-test	24/42 'Well above average'
Hedman et al. 2014	Sweden	PBE (Cohort study)	Outpatient psychiatric clinic	37.9 (11.8)	32.8%	iCBT (therapist-assisted)	1203	298 (0.25)	Did not complete treatment	21/42 'Well above average'
Nordgreen et al. 2019	Denmark	PBE (Open effectiveness study)	University Hospital clinic	35 (12)	42%	14-week iCBT (therapist-assisted)	105	31 (0.30)	SUs who did not complete post-test	20/42 'Above average'
Thimm & Antonsen 2014	Norway	PBE (Pre-Post)	University Hospital Psychiatric Centre	41.6 [20-69]	29%	15 weekly group CBT sessions	143	25 (0.17)	Did not complete treatment	21/42 'Well above average'
Jakobsen et al. 2017	Norway	PBE (Pre-post)	University outpatient clinic	43.5 (11.6)	27.3%	iCBT (therapist-assisted)	22	5 (0.22)	Did not complete treatment	17/42 'Below average'
Baumel et al. 2018	Israel	Non-randomised PBE (Pre-Post)	Adult Outpatient Hospital Clinic	31.95 (5.57)	0%	30-day digital self-help platform & forum	20	3 (0.15)	Did not complete treatment	16/42 'Below average'

Study, Author, Year	Country	Study Design	Treatment setting	Mean age (SD) [range]	Sex (% male)	Treatment duration (format) [mean sessions]	Treatment Sample (n)	Drop-out cases (proportion)	Drop-out definition	Risk of bias (RoB)
Mathieson et al. 2018	Denmark	PBE (Cohort study)	Secondary care mental health clinic	36.03 (10.97) [19-67]	21.7%	iCBT (therapist-assisted)	60	24 (0.4)	Did not complete treatment	16/42 'Below average'
Ward et al. 2016	Australia	PBE (Pre-Post)	Stroke service and community team	66 (11.6)	64.6%	Brainstorm group CBT program	48	11 (0.23)	SUs who did not complete post-test	16/42 'Below average'
Voderholzer et al. 2021	Germany	PBE (Pre-Post)	Primary German healthcare data	44.46 (12.86) [20-64]	28.8%	MindDoc V-CBT 8 sessions	91	15 (0.16)	Did not complete treatment	15/42 'Below average'
Cavanagh et al. 2011	United Kingdom	PBE (pre-post), open trial	Self-help service (tier 2)	NR	33%	8 weekly BtB CCBT sessions	295	115 (0.39)	Did not complete treatment	13/42 'Well below average'
Learmonth et al. 2008	United Kingdom	PBE (Pre-Post)	Specialist CBT Unit	40 (12) [18-70]	38.7%	8 weekly BtB CCBT sessions	394	161 (0.41)	Did not complete treatment	18/42 'Below average'

Note. Abbreviations: NR: not reported; N: number of patients; SU: service users; PBE: practice-based evidence; CBT: cognitive behavioural therapy; CCBT: computerised cognitive behavioural therapy; iCBT: internet-based cognitive behavioural therapy; V-CBT: videoconference cognitive behavioural therapy; BtB: Beating the Blues.

Meta-analysis of LI Treatment for Depression Drop-out Rate

The analysis incorporated data from n=2794 treatment seeking participants with depression. The weighted mean drop-out rate was 28% (95% CI's 23 to 33) and this is displayed in the forest plot in Figure 2. There was significant high study heterogeneity (I^2 = 86%; Q = 73.51, p < .001). To further examine drop-out estimates, one study was removed at a time to assess the stability of the estimate. Completing this technique outlined a pooled drop-out rate ranging between 26-29%. Due to the lack of range, this is indicative of stable drop-out rate estimates. Inspection of the funnel plot (shown in Figure 3) displayed some asymmetry, suggestive of publication bias.

Moderator Analysis

To explore heterogeneity between studies, a moderator analysis was conducted to investigate how certain variables influenced the relationship between LI interventions for depression and their drop-out rates (Helm & Mark, 2012). The moderator analysis is reported in Table 3. No effect was found for RoB, delivery method, drop-out definition or delivery type on the proportion of people who started treatment for depression and resultantly dropped-out. Marginally higher drop-out rates were observed for studies with "well-below average" and "below average" RoB ratings and for studies that defined drop-out as when a patient did not complete a specified number of sessions. Studies utilising individual interventions reported higher (but non-significant) drop-out rates in comparison to group interventions. Internet interventions were also found to have higher (but non-significant) drop-out rates compared to face-to-face interventions. Moderator significance was difficult to ascertain regarding the delivery and patient setup variables, as these separated the studies into the same subgroups.

There were significantly higher drop-out rates for UK studies (40%) compared to Scandanavian (26%) and other countries (22%). Therapy format was found to be a significant moderator; non-therapist assisted CCBT/iCBT had a significantly higher drop-out rates (40%) compared with all other approaches. The drop-out rate for therapist-assisted CCBT/iCBT (28%) was significantly higher than for group CBT (19%) and other therapy (17%) formats.

Drop-out Rate Benchmarking

Drop-out benchmarks are illustrated in Figure 4. For interventions treating depression, the pooled treatment drop-out estimates ranged from 19.2% (general depression treatment) to 32% (computerised treatments for depression). The LI interventions for depression pooled drop-out rate was higher than all the other treatments for depression, apart from the computer-based treatments. It was also observed that the 95% confidence intervals did not overlap between the current study drop-out rate and the general depression treatment rate; meaning that their drop-out rate was significantly different. For general psychotherapies, pooled treatment drop-out rate estimates ranged from 15.8% (ACT) to 28% (DBT), with the pooled drop-out rate for general treatment estimated at 21.9%. For this subsection of the benchmarking process, the LI interventions for depression pooled drop-out rate was equal to the DBT rate. The 95% confidence interval for the current studies pooled drop-out rate did not overlap with the ACT or general psychotherapy rates meaning that they were significantly different.

Figure 2

Forest Plot of Drop-out Rates for LI Depression Treatment

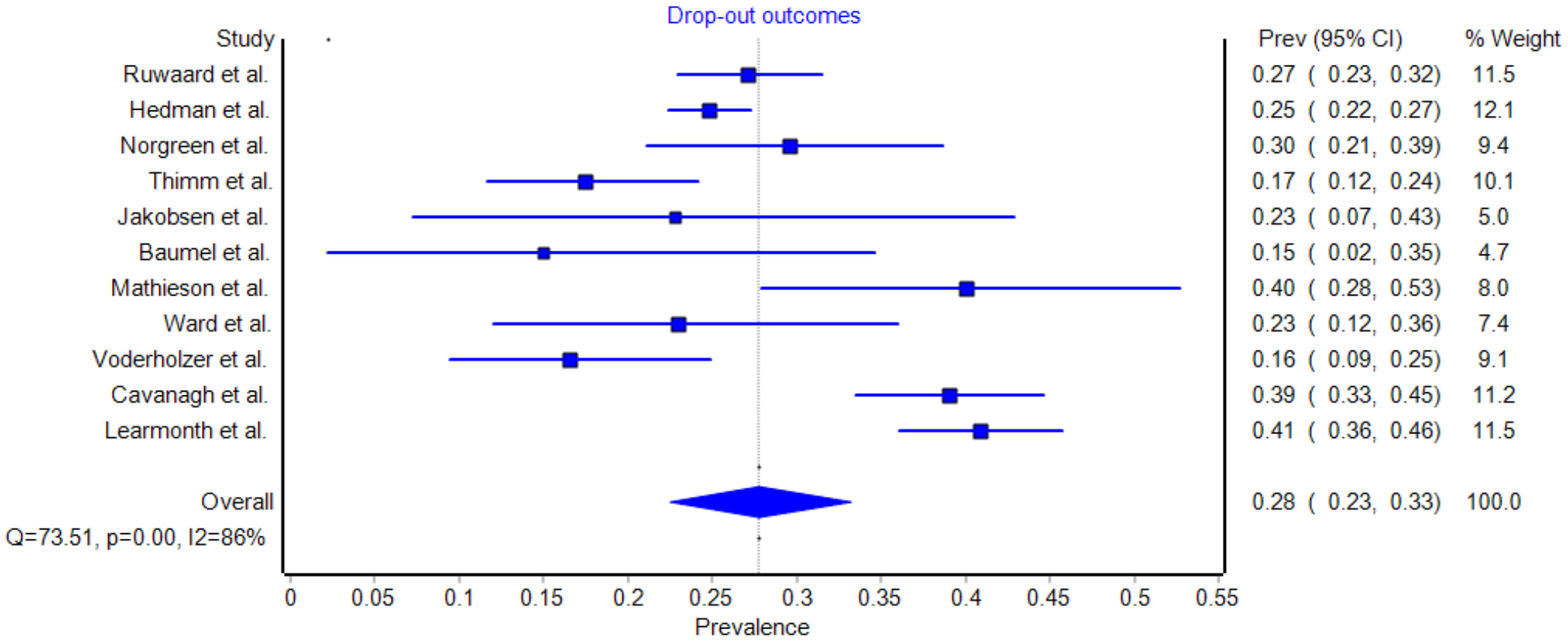


Figure 3

Funnel Plot for Drop-out Rates for LI Depression Treatment

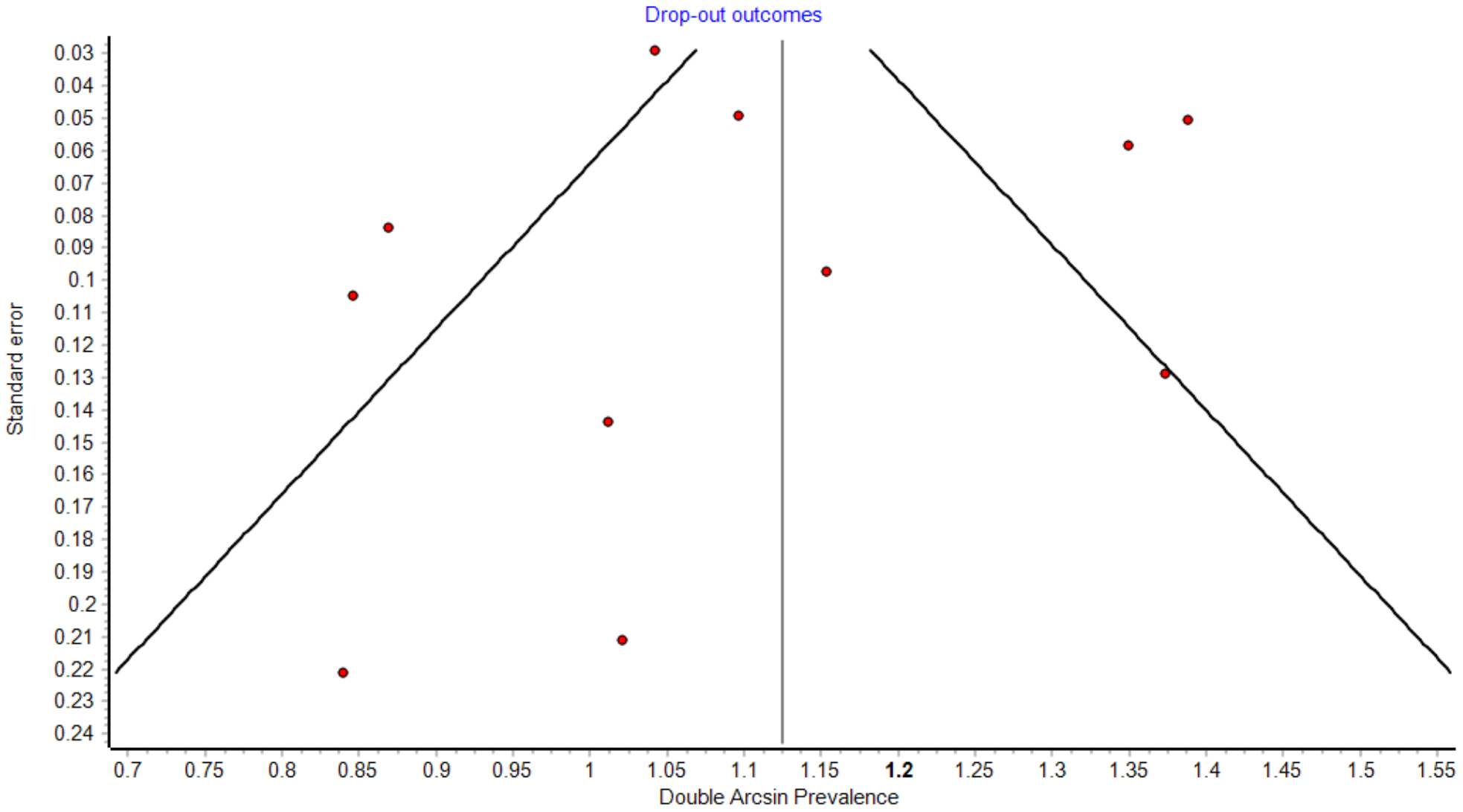


Table 3

Moderator Analyses for Categorical (Subgroups) Moderators of LI Intervention for Depression Drop-out Rates

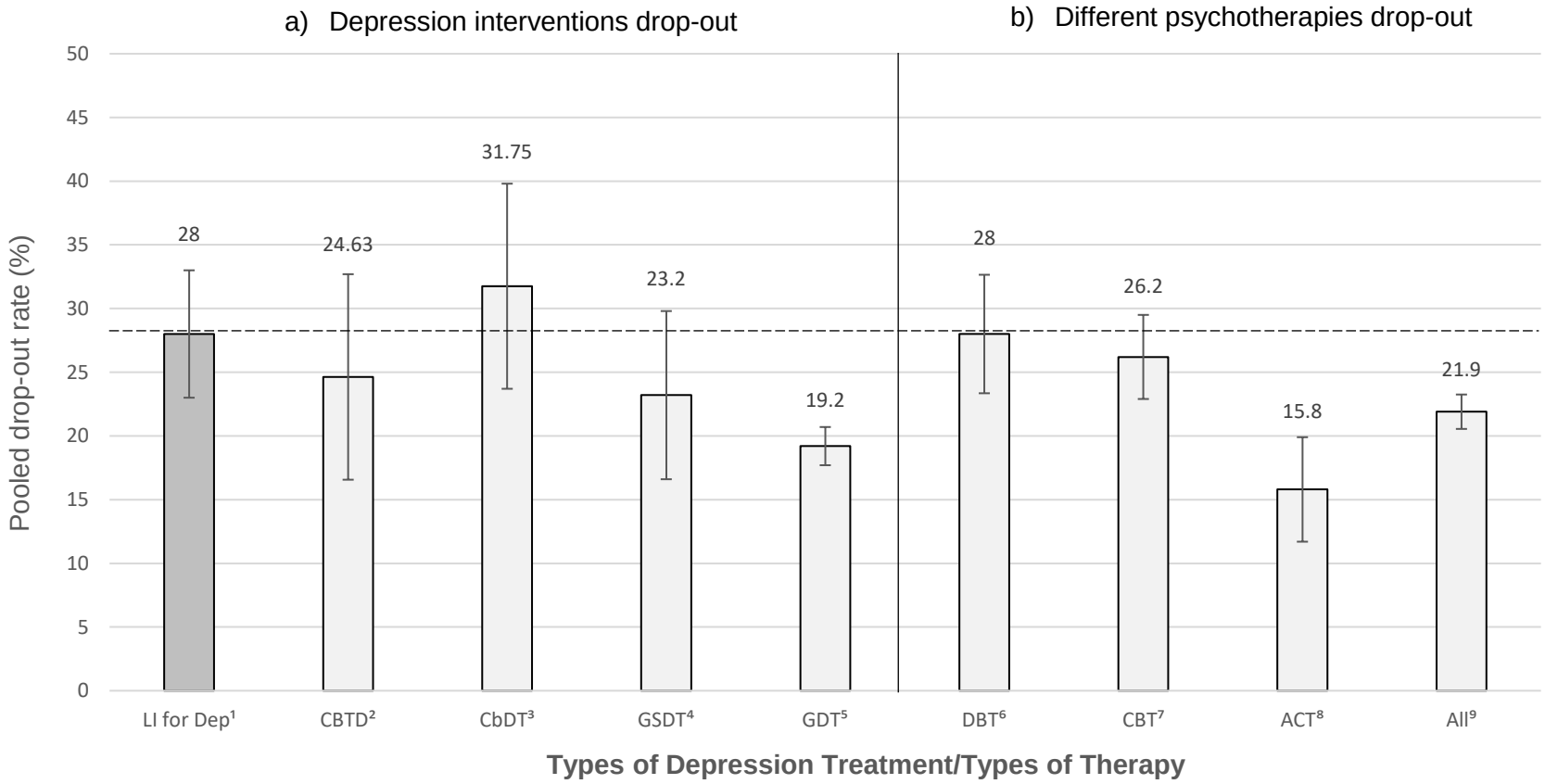
Categorical variable	Subgroup	*n*	Proportion	95% CI	I^2(%)[a]	Subgroup heterogeneity (*p*)
Country	UK	2	0.40	0.36 to 0.44	0	.062
	Scandanavian	5	0.26	0.20 to 0.32	67	.02*
	Other	4	0.22	0.16 to 0.29	47	.13
Drop-out definition	Dropped out/did not complete	8	0.28	0.21 to 0.35	90	.000***
	SU did not complete post-test	3	0.27	0.24 to 0.31	0	.21
Format	TA CCBT/iCBT	5	0.28	0.24 to 0.31	47	.11
	Non-TA CCBT/iCBT	2	0.40	0.36 to 0.44	0	.62
	Group CBT	2	0.19	0.14 to 0.25	0	.39
	Other	2	0.17	0.10 to 0.24	0	.98
Delivery	Internet	9	0.29	0.24 to 0.36	87	.000***
	Face-to-face	2	0.19	0.14 to 0.25	0	.39
Patient setup	Individual	9	0.29	0.24 to 0.36	87	.000***
	Group	2	0.19	0.14 to 0.25	0	.39
RoB	'Well below' & 'below average'	7	0.30	0.22 to 0.38	82	.000***
	'Above average' and 'well above average'	4	0.25	0.21 to 0.29	56	.08

Note. Abbreviations: *n:* number of comparisons, SMD: standardised mean difference; CI: confidence interval; R^2: percentage of variation explained; SE: standard error; RoB: risk of bias; UK: United Kingdom; TA: therapist-assisted; CBT: cognitive behavioural therapy; CCBT: computerised cognitive behavioural therapy; iCBT: internet cognitive behavioural therapy; SU: service user.

[a]Pooled within-group estimates of between-study variance/heterogeneity, significance based on *p* value of associated Q statistic. *significant at $p < .05$ threshold. ** significant at $p < .0.01$ threshold, ***significant at $p < 0.001$ threshold.

Figure 4

Interventions for Depression In-Treatment Drop-Out rates (indicated by diagonal line patterned bars) Benchmarked Against Meta-Analysed Pooled Drop-out Rates for Depression Treatments (2-5) and Different Psychotherapies (6-9) (dashed line represents the pooled rate from the current review; error bars indicate 95% confidence intervals for pooled estimates)



Note. [1]Current study; [2] Cognitive Behavioural Therapy for Unipolar Depression (Hans & Hiller, 2013); [3]Computer-based Depression Treatments (Kaltenthaler et al., 2008); [4]General Subthreshold Depression Treatment (Cuijpers et al., 2007); [5]General Depression Treatment (Swift & Greenberg, 2014); [6]Dialectical Behaviour Therapy (DBT; Dixon & Linardon, 2020); [7]Cognitive Behavioural Therapy (CBT; Fernandez et al., 2015); [8]Acceptance and Commitment Therapy (ACT; Ong et al; 2018); [9]Psychotherapy and/or pharmacotherapy (Swift et al., 2017).

Discussion

This meta-analysis synthesised published studies that included drop-out rates for LI treatments for depression in routine care. Overall, the aims of the study were achieved and the results indicated that the pooled drop-out rate for LI interventions for depression was comparable to the dropout rates generated by other CBT approaches and other therapies. However, although not always significant, drop-out rates for LI treatments for depression were higher than most of the benchmarked depression interventions and other psychotherapeutic interventions. Importantly, differences between drop-out rates can be observed between meta-analyses only including routine care data and those only including data from control trials. Significant moderators within the meta-analysis included country and therapy format, which appeared in line with previous research.

For those patients that started LI interventions for depression, a pooled average of 28% subsequently dropped out of treatment. This drop-out rate was not significantly different from other rates reported for unipolar depression (Hans & Hiller, 2013), computer-based depression treatment (Kaltenthaler et al., 2008) or general depression (Swift & Greenberg, 2014) treatments. Interestingly, it was significantly higher than the drop-out rates reported for the subthreshold depression treatment review (Cuijpers et al., 2017). Although this may mean that services can deliver LI interventions to a large number of patients in an accessible and inexpensive format, this rate appears elevated and so places the acceptability of such approaches into sharper focus and question. Nine of the included eleven studies implemented online interventions (CCBT, iCBT or forum-based support), with two of those not containing any therapeutic contact. Such

delivery formats have previously been associated with higher drop-out rates (Kaltenhaler et al., 2008; Richards & Richardson, 2012). Benchmarking also outlined that the LI interventions for depression drop-out rate was equal to DBT and slightly higher than the CBT rate, whilst being significantly higher than ACT and general studies. Country was a significant moderator, suggesting lower drop-out rates for LI interventions for depression in countries outside of the UK. Furthermore, the second significant moderator was delivery format, which suggested higher drop-out rates when CCBT/iCBT without therapist assistance was provided, in comparison to therapy-assisted CCBT/iCBT, group CBT or other approaches.

Through the process of benchmarking, it was demonstrated that the drop-out rate of LI interventions for depression were slightly higher than most types of depression intervention (including CBT for unipolar depression, treatment of subthreshold depression and general depression treatment) and higher compared to the majority of therapy types (CBT, ACT and all). While the current pooled drop-out rate was slightly higher than most of the depression interventions, a large amount of overlap can be observed between confidence intervals and therefore significant differences in drop-out between these treatments cannot be declared. Despite this, the finding that the upper confidence interval limit of the general depression treatment drop-out rate did not overlap with the lower limit of the current studies' pooled estimate, indicates that the general depression treatments' drop-out rate was significantly lower. Differences between these two rates may be due to the inclusion of high-intensity approaches (i.e. integrative, psychodynamic, DBT and more) within the general depression treatments review. Namely, the inclusion of integrative approaches resulted in the lowest drop-out

rate of 10.9%, suggesting such methods may be more suited to depressed patients (Swift et al., 2014). The manualised and adherent nature of LI interventions may account for the disparity between drop-out rates, as these approaches can result in patients receiving limited/no therapist contact and restricted treatment flexibility. However, these conclusions should be interpreted with caution, as the review did not make direct comparisons and subsequently drop-out rate variations could be ascribed to a combination of methodological (review methods and drop-out definitions) and/or therapeutic reasons (components contributing to higher or lower drop-out).

High study heterogeneity necessitated additional exploration with the completion of a moderator analysis. Country was found to be a significant moderator, with studies from the UK having higher drop-out rates in comparison to studies conducted in Scandinavia and other countries. Another significant moderator was therapy format, with studies using non-therapist-assisted CCBT/iCBT, attaining a significantly higher drop-out rate than all other formats. Inspection of the studies categorised into the country and therapy format subgroups indicates confounding results between the two moderators making it difficult to separate their effects. For both UK studies, only data from the Beating the Blues (BtB) CCBT program was included (a non-therapist assisted 8-session therapy recommended by NICE). Whereas studies from Scandinavia and other countries utilised either therapist-assisted CCBT/iCBT, group CBT, videoconference CBT or online forums (all of which incorporated elements of human contact or support). In relation to this, research has established that the presence of human support (either in an administrative or therapeutic capacity) can reduce drop-out rates by up to 40% (Richards & Richardson, 2012).

The moderator findings of this review verify the importance of human contact during therapy in relation to drop-out rates, whilst also explaining why these rates are significantly higher for UK studies compared to other regions/countries. The literature also suggests reasons why drop-out rates may be higher for these types of therapy. One study investigating patients' experiences of non therapist-assisted support in the UK, found that patients had little/no knowledge about CCBT/BtB before they were referred, were sometimes incorrectly matched to the programmes and often desired unavailable built-in support features (Folker et al., 2018; Du et al., 2021). It should be said that moderator analyses are restrained by the number of studies included in the review, and therefore interpretations can be vulnerable to low power (which could be the case in this instance; Deeks et al., 2019).

The presence of some asymmetry in the funnel plot suggests that publication bias may have been present between the included studies. In this instance, although the funnel plot appears suggestive of publication bias, the accuracy of this interpretation is debatable. As the included studies are largely concentrating on the efficacy of LI treatments for depression, they are less likely to be impacted by publication bias based upon their drop-out rates (Liu, 2010). It is important to note, nevertheless, that the funnel plot appears to lack representation from smaller studies showing larger drop-out rates. This may indicate that researchers conducting such studies, may be choosing to avoid publication or might not be reporting drop-out rates. To assess for this, future research could look to include unpublished literature and service evaluations, to better reduce the risk of under-estimating drop-out rates. In addition, further analyses could look to increase the number of studies by broadening the inclusion criteria of the review,

as larger study numbers often make it easier to identify gaps that can more reliably evidence the presence of publication bias (Cooper et al., 2009). As an aside, it has been suggested that there is no evidence that publication bias tests, such as funnel plots, adjust appropriately for proportional data (due to their assumption that 'positive results' are more likely to be published) and therefore some propose it would be wise to assess this qualitatively instead (Barker et al., 2021).

Interestingly, the pooled drop-out rate for this review (28%; 95% CI's 23 to 33) was higher than the rate of an equivalent systematic review that included randomised studies (19.2%; 95% CI's 17.8 to 20.8; Cuijpers et al., 2007). As this review only included research from studies gathering data within routine care conditions, it was expected that the drop-out rate would be significantly higher due to a number of factors (i.e., less screening and use of inclusion/exclusion criteria, sample heterogeneity, less rigour about following up and keeping patients in a study, demand characteristics etc.). In comparison, the drop-out rate was 8.8% higher for this review and a lack of overlap between the confidence intervals deemed the difference significant. This provides evidence that the drop-out rates found for reviews including RCTs (such as Cuijpers et al., 2007), may not accurately reflect what happens in routine care. Importantly, this highlights three salient points. Firstly, this review appears to support concerns regarding the applicability of EBP findings to real-world settings (Barkham & Mellor-Clark, 2003; Bower, 2003; Green, 2006; Horn & Gassaway, 2007; Westfell, 2007). Such findings uphold the proposal that therapists and researchers should be cautious when applying EBP research to routine care settings without comparable PBE. Secondly, it appears as though drop-out rates are higher in routine care settings, therefore LI interventions for

depression are less acceptable than the literature had previously presented. Finally, previous recommendations from EBP reviews aiming to increase the acceptability of LI interventions for depression may need to be re-thought and fine-tuned, whilst further PBE research should be expedited to consider service delivery issues in the immediate future.

Limitations

Importantly, despite finding multiple relevant studies to include, the review excluded many studies as the drop-out data were either not sufficiently reported or was not reported at all. Although this meant that the inclusion/exclusion criteria were strictly followed, it limited the number of studies that could be compared and may have limited and/or skewed the moderator analysis. As outlined, the POSMRF was selected as the RoB tool due to its appropriateness for multiple research designs and its respected psychometric properties. On the other hand, due to some of its included items, the POSMRF can be overly castigatory towards research utilising self-help interventions (Spauls et al., 2021). This may have resulted in the included studies receiving lower quality scores than they would if a standard RoB tool was used. Furthermore, the calculation of cut-off boundaries may have allowed for further comparisons between studies to be made, but this could be interpreted as fairly subjective and is purely based upon the studies included (for example, the inclusion of studies with higher scores might have deem those rated ‘well above average’ of lower quality). Thus, future research using the POSMRF should ensure that standardised cut-offs have been developed.

Otherwise, researchers could appropriately omit items so the measure is suitably tailored for certain reviews.

Despite its recommended use within the literature, another limitation relates to the use of the I^2 statistic. It has been stated that as the statistic was developed to be used for comparative data, when it is alternatively utilised for proportional data, the I^2 score can be resultantly high (Barker et al., 2021). Therefore, higher scores may not be truly representative of inconsistent data if this statstic is applied to an inappropriate data set, which means the test results should be understood with caution (Barker et al., 2021). During the study selection phase, studies were chosen based upon the main authors' notion of what a LI intervention was, which was based on historic searches of previous LI reviews and articles attempting to conceptualise its meaning (see search strategy; Cremers et al., 2019; Bower et al., 2013). Despite this, it is recognised that there can be debate regarding what is classed as a LI intervention (Shafran et al., 2021). Although the author applied a consistent framework and a set agreed definition to decide upon the eligibility of studies, this may have resulted in studies that some would deem a LI intervention being missed or studies which some would not have deemed LI being included.

Clinical & Research Implications

This review outlined various implications. Importantly, it can be said that there is a need for services that deliver non therapist-assisted LI treatments for depression to consider utilising strategies to minimise the likelihood of patient drop-out. Secondly,

there is a requirement for more PBE studies investigating patients with depression accessing interventions. This is necessary so that future analyses can more comprehensively and accurately compare real-world drop-out rates to clinical trials. Additionally, due to the significantly higher drop-out rates found in this review compared to Scandanavia and other countries, services in the UK should to be mindful of the formats they deliver LI interventions for depression (i.e. BtB). This could involve opting for other formats which are still LI in nature, but with elements that may reduce drop-out rates (i.e. more therapist contact/check-ins). In relation to this and in the interest of reducing drop-out rates, services could also consider providing different interventions for the treatment of depression, such as ACT, due to its significantly lower drop-out rates.

In relation to other research, it is known that drop-out rates for LI interventions are highest during the earliest stages, which infers that the crucial time frame opportunity is sessions 1-4 (Delgadillo et al., 2013). Hence, services should be aware of specific factors that increase the likelihood of patients dropping out (i.e. being male, younger, accessing a time-limited intervention, in a University-based settings, having co-morbid anxiety and a lower educational level (Swift & Greenberg, 2012; Zimmerman et al., 2017). In circumstances where these elements are present, early strategies should be employed to actively try and reduce anticipated drop-outs. For example, services could more openly discuss both the possibility of dropping out with their patients and treatment expectations/reservations, whilst increasing therapist check-ins (Cavanagh, 2012; Swift & Greenberg, 2012). It has also been suggested that strengthening patient hope, enhancing patient motivation and discussing behaviour

change principles may have a positive impact on drop-out (Swift & Greenberg, 2015). More recent research has also advocated for the implementation of systemic change, which would involve ensuring that professionals referring people for LI interventions for depression (i.e. GPs, healthcare professionals etc.) are knowledgeable about the format and content of LI interventions (Du et al., 2021).

Future research could aim to investigate the impact of specific strategies on reductions in particular patient drop-out rates. Namely, this could involve pinpointing people at specific risk of prematurely discontinuing therapy and trialing certain strategies. For example, more visual and audio-based interventions for those with lower educational levels, or more accessible formats for younger people (i.e. videos or gaming) and more (Karyotaki et al., 2015). Research could also look to investigate the incidence of drop-out rates at particular time points in association with the implementation of certain strategies. Additionally, the value of adding therapist or coach support for self-help LI interventions needs to be further researched, particularly for online/remote interventions (i.e. iCBT and CCBT; Karyotaki et al, 2015). Further research in this area will help to reduce and/or prevent treatment discontinuation, which appears to be becoming more poignant in an age where computerised and remote treatments (which are recognised for early treatment termination) are becoming commonplace.

Conclusions

LI interventions for depression are a widely used and accepted form of therapy adopted by psychological services in primary care. The current review provides some evidence that the routine care drop-out rates from these interventions are higher than those reported from randomised studies. On the other hand, the drop-out rate for LI interventions for depression were comparable with some forms of psychotherapy (DBT and CBT), but had higher rates than others (ACT and general). It is apparent, however, that more routine care studies are required so more rigorous and comprehensive meta-analyses can be conducted. In relation to the review aims, the current study has been successful in calculating a pooled drop-out rate and completing a moderator analysis. The moderator analysis highlighted the significant influence of both country and therapy format. Future research should aim to explore achieving more effective referral pathways and treatment matching. Research should evaluate strategies prior to (i.e. pre-therapy discussions regarding treatment expectations/reservations) and during (i.e. more accessible materials, patient reminders, therapist check-ins, more integrative/open approaches etc.) LI interventions for depression, in an effort to generate improved attendance and completion rates. Services should also be mindful of the unprecedented uptake in the use of CCBT/iCBT as a result of the COVID-19 pandemic and the associated increased risks of drop-out for such delivery formats.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

American Psychological Association. (n.d.). Depression. In *APA Dictionary of Psychology*. Retrieved April 24, 2022, from https://dictionary.apa.org/depression

Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., & Newby, J. M. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. *Journal of Anxiety Disorders*, *55*, 70-78. https://doi.org/10.1016/j.janxdis.2018.01.001

Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *10*(6), 319-327. https://doi.org/10.1002/cpp.379

Barendregt, J. J., Doi, S. A., Lee, Y. Y., Norman, R. E., & Vos, T. (2013). Meta-analysis of prevalence. *Journal of Epidemiology & Community Health*, *67*(11), 974-978. http://dx.doi.org/10.1136/jech-2013-203104

Barker, T. H., Migliavaca, C. B., Stein, C., Colpani, V., Falavigna, M., Aromataris, E., & Munn, Z. (2021). Conducting proportional meta-analysis in different types of systematic reviews: a guide for synthesisers of evidence. *BMC Medical Research Methodology*, *21*(1), 1-9. https://doi.org/10.1186/s12874-021-01381-z

Barrett, M. S., Chua, W. J., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy: Theory, Research, Practice, Training*, *45*(2), 247. https://doi.org/10.1037/0033-3204.45.2.247

*Baumel, A., Tinkelman, A., Mathur, N., & Kane, J. M. (2018). Digital peer-support platform (7Cups) as an adjunct treatment for women with postpartum depression: feasibility, acceptability, and preliminary efficacy study. *JMIR mHealth and uHealth*, *6*(2), e9482. https://doi.org/10.2196/mhealth.9482

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088-1101. https://doi.org/10.2307/2533446

Bennett-Levy, J., Richards, D. A., Farrand, P., Christensen, H., Griffiths, K. M., Kavanagh, D. J., & Proudfoot, J. (2010). Low intensity CBT interventions: a revolution in mental health care. *Oxford Guide to Low Intensity CBT Interventions*, *3*, 3-18.

Björk, T., Bjorck, C., Clinton, D., Sohlberg, S., & Norring, C. (2009). What happened to the ones who dropped out? Outcome in eating disorder patients who complete or prematurely terminate treatment. *European Eating Disorders Review*, *17*, 109 – 119. https://doi.org/10.1002/erv.911

Bockting, C. L. H., Williams, A. D., Carswell, K., & Grech, A. E. (2016). The potential of low-intensity and online interventions for depression in low-and middle-income countries. *Global Mental Health*, *3*. https://doi.org/10.1017/gmh.2016.21

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., ... & Liu, E. T. H. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *British Medical Journal*, *346*. https://doi.org/10.1136/bmj.f540

Campbell, L. F., Norcross, J. C., Vasquez, M. J., & Kaslow, N. J. (2013). Recognition of psychotherapy effectiveness: The APA resolution. *Psychotherapy*, *50*(1), 98. https://doi.org/10.1037/a0031817

Castonguay. L., Barkham., M., Lutz., W., & McAleavey, A. (2013). Practice-Based Evidence. In M. J., Lambert (Eds.). *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 191-222). John Wiley & Sons.

Cavanagh, K. (2010). Tune on, tune in and (don't) drop out: engagement, adherence, attrition, and alliance with internet-based interventions. In: J. Bennett-Levy., D. Richards., P. Farrand., Christensen., K. M. Griffiths., D.J. Kavanagh., B. Klein., M.A. Lau., J. Proudfoot., L. Ritterband., J. White., & C. Williams (Eds). *Oxford guide to low intensity CBT interventions* (pp. 227-234). Oxford University Press.

*Cavanagh, K., Seccombe, N., & Lidbetter, N. (2011). The implementation of computerized cognitive behavioural therapies in a service user-led, third sector self help clinic. *Behavioural and Cognitive Psychotherapy*, *39*(4), 427-442. https://doi.org/10.1017/S1352465810000858

Chowdhary, N., Anand, A., Dimidjian, S., Shinde, S., Weobong, B., Balaji, M., & Patel, V. (2016). The Healthy Activity Program lay counsellor delivered treatment for severe depression in India: systematic development and randomised evaluation. *The British Journal of Psychiatry*, *208*(4), 381-388. https://doi.org/10.1192/bjp.bp.114.161075

Cooper, A. A., Kline, A. C., Baier, A. L., & Feeny, N. C. (2018). Rethinking research on prediction and prevention of psychotherapy drop-out: A mechanism-oriented approach. *Behavior Modification*, 1-24. https://doi.org/10.1177/0145445518792251

Cooper, A. A., & Conklin, L. R. (2015). Drop-out from individual psychotherapy for major depression: A meta-analysis of randomized clinical trials. *Clinical Psychology Review*, *40*, 57-65. https://doi.org/10.1016/j.cpr.2015.05.001

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The Handbook of Research Synthesis and Meta-Analysis, 2nd Ed.* Russell Sage Foundation.

Cox, G. R., Callahan, P., Churchill, R., Hunot, V., Merry, S. N., Parker, A. G., & Hetrick, S. E. (2014). Psychological therapies versus antidepressant medication, alone and in combination for depression in children and adolescents. *Cochrane Database of Systematic Reviews*, (11). https://doi.org/10.1002/14651858.CD008324.pub3

Cremers, G., Taylor, E., Hodge, L., & Quigley, A. (2022). Effectiveness and acceptability of low-intensity psychological interventions on the well-being of older adults: a systematic review. *Clinical Gerontologist*, *45*(2), 214-234. https://doi.org/10.1080/07317115.2019.1662867

Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological Medicine*, *40*(12), 1943-1957. https://doi.org/10.1017/S0033291710000772

Cuijpers, P., Smit, F., & Van Straten, A. (2007). Psychological treatments of subthreshold depression: a meta-analytic review. *Acta Psychiatrica Scandinavica*, *115*(6), 434-441. https://doi.org/10.1111/j.1600-0447.2007.00998.x

Deeks, J. J., Higgins, J. P., Altman, D. G., & Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions*, 241-284. https://doi.org/10.1002/9781119536604.ch10

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose–response in low intensity psychological interventions. *British Journal of Clinical Psychology*, *53*(1), 114-130. https://doi.org/10.1111/bjc.12031

Dixon, L. J., & Linardon, J. (2020). A systematic review and meta-analysis of drop-out rates from dialectical behaviour therapy in randomized controlled trials. *Cognitive Behaviour Therapy*, *49*(3), 181-196. https://doi.org/10.1080/16506073.2019.1620324

Du, E., Quayle, E., & Macleod, H. (2021). A qualitative study of patients' experiences and acceptance of computerised cognitive behavioural therapy in primary care, Scotland. *The Cognitive Behaviour Therapist*, *14*, 24. https://doi.org/10.1017/S1754470X21000210

El Dib, R., Nascimento Junior, P., & Kapoor, A. (2013). An alternative approach to deal with the absence of clinical trials: a proportional meta-analysis of case series studies. *Acta Cirurgica Brasileira*, *28*(12), 870-876. https://doi.org/10.1590/s0102-86502013001200010

Fernandez, E., Salem, D., Swift, J. K., & Ramtahal, N. (2015). Meta-analysis of drop-out from cognitive behavioral therapy: Magnitude, timing, and moderators. *Journal of Consulting and Clinical Psychology*, *83*(6), 1108b. https://doi.org/10.1037/ccp0000044

Folker, A. P., Mathiasen, K., Lauridsen, S. M., Stenderup, E., Dozeman, E., & Folker, M. P. (2018). Implementing internet-delivered cognitive behavior therapy for common mental health disorders: A comparative case study of implementation challenges perceived by therapists and managers in five European internet services. *Internet Interventions*, *11*, 60-70. https://doi.org/10.1016/j.invent.2018.02.001

Glover, G., Webb, M., & Evison, F. (2010). Improving access to psychological therapies: A review of the progress made by sites in the first rollout year. Retrieved April 2022, from http://www.iapt.nhs.uk/silo/files/iapt-a-review-of-the-progress-made-by-sites-in-the-first-roll8208-out-year.pdf.

Hallam, C., Simmonds-Buckley, M., Kellett, S., Greenhill, B., & Jones, A. (2021). The acceptability, effectiveness, and durability of cognitive analytic therapy: Systematic review and meta-analysis. *Psychology and Psychotherapy: Theory, Research and Practice*, *94*, 8-35. https://doi.org/10.1111/papt.12286

Hans, E., & Hiller, W. (2013). Effectiveness of and drop-out from outpatient cognitive behavioral therapy for adult unipolar depression: a meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology*, *81*(1), 75-88. https://doi.org/10.1037/a0031080

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, *9*(3), 329. https://doi.org/10.1093/clipsy.9.3.329

Harrer, M., Ebert, D. D., Furukawa, T. A., Cuijpers, P. (2021). *Doing Meta-Analysis with R: A Hands-On Guide.* CRC Press.

Hatchett, G. T., & Park, H. L. (2003). Comparison of four operational definitions of premature termination. *Psychotherapy: Theory, Research, Practice, Training*, *40*(3), 226. https://doi.org/10.1037/0033-3204.40.3.226

*Hedman, E., Ljótsson, B., Kaldo, V., Hesser, H., El Alaoui, S., Kraepelien, M., ... & Lindefors, N. (2014). Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. *Journal of Affective Disorders*, *155*, 49-58. https://doi.org/10.1016/j.jad.2013.10.023

Helm, R., & Mark, A. (2012). Analysis and evaluation of moderator effects in regression models: state of art, alternatives and empirical example. *Review of Managerial Science*, *6*(4), 307-332. https://doi.org/10.1007/s11846-010-0057-y

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hommel, K. A., Hente, E., Herzer, M., Ingerski, L. M., & Denson, L. A. (2013). Telehealth behavioral treatment for medication nonadherence: a pilot and feasibility study. *European Journal of Gastroenterology & Hepatology*, *25*(4), 469. https://doi.org/10.1097/MEG.0b013e32835c2a1b

Horn, S. D., & Gassaway, J. (2007). Practice-based evidence study design for comparative effectiveness research. *Medical Care*, *45*(10), S50-S57. https://doi.org/10.1097/MLR.0b013e318070c07b

Hunsley, J. (2003). Cost effectiveness and medical cost-offset considerations in psychological service provision. *Canadian Psychology*, *44*, 61-73. https://doi.org/10.1037/h0085818

Imel, Z. E., Laska, K., Jakupcak, M., & Simpson, T. L. (2013). Meta-analysis of drop-out in treatments for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, *81*(3), 394. https://doi.org/10.1037/a0031474

Jackson, D. (2006). The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in Medicine*, *25*(15), 2688-2699. https://doi.org/10.1002/sim.2481

*Jakobsen, H., Andersson, G., Havik, O. E., & Nordgreen, T. (2017). Guided Internet-based cognitive behavioral therapy for mild and moderate depression: a benchmarking study. *Internet Interventions*, *7*, 1-8. https://doi.org/10.1016/j.invent.2016.11.002

Kaltenthaler, E., Sutcliffe, P., Parry, G., Beverley, C., Rees, A., & Ferriter, M. (2008). The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. *Psychological Medicine*, *38*(11), 1521-1530. https://doi.org/10.1017/S0033291707002607

Karyotaki, E., Kleiboer, A., Smit, F., Turner, D. T., Pastor, A. M., Andersson, G., ... & Cuijpers, P. (2015). Predictors of treatment drop-out in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. *Psychological Medicine*, *45*(13), 2717-2726. https://doi.org/10.1002/sim.2481

Klein, E. B., Stone, W. N., Hicks, M. W., & Pritchard, I. L. (2003). Understanding drop-outs. *Journal of Mental Health Counseling*, *25*(2), 89-100. https://doi.org/10.17744/mehc.25.2.xhyreggxdcd0q4ny

Kontopantelis, E., Springate, D. A., & Reeves, D. (2013). A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. *PLOS One*, *8*(7), e69930. https://doi.org/10.1371/journal.pone.0069930

Kriston, L., Von Wolff, A., Westphal, A., Hölzel, L. P., & Härter, M. (2014). Efficacy and acceptability of acute treatments for persistent depressive disorder: a network meta-analysis. *Depression and Anxiety*, *31*(8), 621-630. https://doi.org/10.1002/da.22236

Layard, R., & Clark, D. M. (2014). *Thrive: the power of evidence-based psychological therapies*. Penguin UK.

Leichsenring, F., Sarrar, L., & Steinert, C. (2019). Drop-outs in psychotherapy: a change of perspective. *World Psychiatry*, *18*(1), 32. https://doi.org/10.1002/wps.20588

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

*Learmonth, D., Trosh, J., Rai, S., Sewell, J., & Cavanagh, K. (2008). The role of computer-aided psychotherapy within an NHS CBT specialist service. *Counselling and Psychotherapy Research*, *8*(2), 117-123. https://doi.org/10.1080/14733140801976290

Light, R. J., Richard, J., Light, R., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Liu, S. T. (2010). Comparative Study of Four Techniques for the Detection of Bluetongue Virus. *Doctoral Dissertation*. https://doi.org/10.13140/RG.2.2.27199.00161

Lopes, R. T., Gonçalves, M. M., Sinai, D., & Machado, P. P. (2018). Clinical outcomes of psychotherapy drop-outs: does dropping out of psychotherapy necessarily mean failure?. *Brazilian Journal of Psychiatry*, *40*(2), 123-127. https://doi.org/10.1590/1516-4446-2017-2267

Lueger, R. J., & Barkham, M. (2010). Using benchmarks and benchmarking to improve quality of practice and services. *Developing and delivering practice-based evidence: A Guide for the Psychological Therapies*, 223-256. https://doi.org/10.1002/9780470687994.ch9

Marks, I. M., Cavanagh, K., & Gega, L. (2007). Computer-aided psychotherapy: revolution or bubble?. *The British Journal of Psychiatry*, *191*(6), 471-473. https://doi.org/10.1192/bjp.bp.107.041152

*Mathiasen, K., Riper, H., Andersen, T. E., & Roessler, K. K. (2018). Guided internet-based cognitive behavioral therapy for adult depression and anxiety in routine secondary care: observational study. *Journal of Medical Internet Research*, *20*(11), e10927. https://doi.org/10.2196/10927

Melartin, T. K., Rytsälä, H. J., Leskelä, U. S., Lestelä-Mielonen, P. S., Sokero, T. P., & Isometsä, E. T. (2005). Continuity is the main challenge in treating major depressive disorder in psychiatric care. *The Journal of Clinical Psychiatry*, 66, 220-227. https://doi.org/10.4088/JCP.v66n0210

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, *14*(1), 1-10. https://doi.org/10.1186/s12913-014-0579-0

Murtaugh, P. A. (2002). Journal quality, effect size, and publication bias in meta-analysis. *Ecology*, *83*(4), 1162-1166.

National Institute for Health and Clinical Excellence (2009). *Depression: the treatment and management of depression in adults* (NICE Guideline No. 90). https://www.nice.org.uk/guidance/cg90

*Nordgreen, T., Blom, K., Andersson, G., Carlbring, P., & Havik, O. E. (2019). Effectiveness of guided Internet-delivered treatment for major depression in routine mental healthcare-An open study. *Internet interventions*, *18*, 100274. https://doi.org/10.1016/j.invent.2019.100274

O'Connor, S. R., Tully, M. A., Ryan, B., Bradley, J. M., Baxter, G. D., & McDonough, S. M. (2015). Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Research Notes*, *8*(1), 1-7.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Jama*, *295*(6), 676-680.

Rahman, A., Malik, A., Sikander, S., Roberts, C., & Creed, F. (2008). Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: a cluster-randomised controlled trial. *The Lancet*, *372*(9642), 902-909. https://doi.org/10.1016/S0140-6736(08)61400-2

Reis, B. F., & Brown, L. G. (1999). Reducing psychotherapy drop-outs: Maximizing perspective convergence in the psychotherapy dyad. *Psychotherapy: Theory, Research, Practice, Training*, *36*(2), 123. https://doi.org/10.1037/h0087822

Reneses, B., Munoz, E., & Lopez-Ibor, J. J. (2009). Factors predicting drop-out in community mental health centres. *World Psychiatry*, *8*, 173–177. https://doi.org/10.1002/j.2051-5545.2009.tb00246.x

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

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clinical Psychology Review*, *32*(4), 329-342. https://doi.org/10.1016/j.cpr.2012.02.004

Rodgers, M., Asaria, M., Walker, S., McMillan, D., Lucock, M., Harden, M., & Eastwood, A. (2012). The clinical effectiveness and cost-effectiveness of low-intensity psychological interventions for the secondary prevention of relapse after depression: a systematic review. *Health Technology Assessment*, *16*(28), 1366-5278. https://doi.org/10.3310/hta16280

Royal College of Psychiatrists (2013). *Report of the second round of the National Audit of Psychological Therapies (NAPT).* London: Healthcare Quality Improvement Partnership.

*Ruwaard, J., Lange, A., Schrieken, B., Dolan, C. V., & Emmelkamp, P. (2012). The effectiveness of online cognitive behavioral treatment in routine clinical practice. *PLOS One*, *7*(7), e40089. https://doi.org/10.1371/journal.pone.0040089

Santana, L., Fontenelle, J. M., Yücel, M., & Fontenelle, L. F. (2013). Rates and correlates of nonadherence to treatment in obsessive-compulsive disorder. *Journal of Psychiatric Practice*, *19*(1), 42-53. https://doi.org/10.1097/01.pra.0000426326.49396.97

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Services Research*, *17*(1), 1-13. https://doi.org/10.1186/s12913-017-2031-8

Shafran, R., Myles-Hooton, P., Bennett, S., & Öst, L. G. (2021). The concept and definition of low intensity cognitive behaviour therapy. *Behaviour Research and Therapy*, *138*, 103803. https://doi.org/10.1016/j.brat.2021.103803

Sharf, J., Primavera, L. H., & Diener, M. J. (2010). Drop-out and therapeutic alliance: a meta-analysis of adult individual psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *47*(4), 637. https://doi.org/10.1037/a0021175

Simon, G. E., Ding, V., Hubbard, R., Fishman, P., Ludman, E., Morales, L., ... & Savarino, J. (2012). Early drop-out from psychotherapy for depression with group-and network-model therapists. *Administration and Policy in Mental Health and Mental Health Services Research*, *39*(6), 440-447. https://doi.org/10.1007/s10488-011-0364-x

Simmonds-Buckley, M., Kellett, S., & Waller, G. (2019). Acceptability and efficacy of group behavioral activation for depression among adults: a meta-analysis. *Behavior Therapy*, *50*(5), 864-885. https://doi.org/10.1016/j.beth.2019.01.003

Sinyor, M., Rezmovitz, J., & Zaretsky, A. (2016). Screen all for depression. *British Medical Journal*, *352*. https://doi.org/10.1136/bmj.i1617

Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., & Staiger, P. K. (2017). Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review. *Clinical Psychology Review*, *57*, 141-163. https://doi.org/10.1016/j.cpr.2017.09.002

Spauls, J. (2021). *Bolstering Youth Mental Health and Wellbeing: An Exploration on Acceptance and Commitment Therapy, Possible Selves and Structured Activity* (Doctoral dissertation, University of East Anglia).

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, *43*(2), 476-493. https://doi.org/10.1093/ije/dyu038

Stuart, E. A., Bradshaw, C. P., & Leaf, P. J. (2015). Assessing the generalizability of randomized trial results to target populations. *Prevention Science*, *16*(3), 475-485. https://doi.org/10.1007/s11121-014-0513-z

Swain, J., Hancock, K., Hainsworth, C., & Bowman, J. (2013). Acceptance and Commitment Therapy in the treatment of anxiety: A systematic review. *Clinical Psychology Review*, *8*, 965–978. https://doi.org/10.1016/j.cpr.2013.07.002

Swift, J. K., & Greenberg, R. P. (2014). A treatment by disorder meta-analysis of drop-out from psychotherapy. *Journal of Psychotherapy Integration*, *24*(3), 193. https://doi.org/10.1037/a0037512

Swift, J. K., & Greenberg, R. P. (2015). *Provide role induction*. American Psychological Association.

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: a meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*(4), 547. https://doi.org/10.1037/a0028226

Swift, J. K., & Greenberg, R. P. (2015). *Premature termination in psychotherapy: Strategies for engaging clients and improving outcomes*. American Psychological Association.

Swift, J. K., Greenberg, R. P., Tompkins, K. A., & Parkin, S. R. (2017). Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: A meta-analysis of head-to-head comparisons. *Psychotherapy, 54*(1), 47–57. https://doi.org/10.1037/pst0000104

*Thimm, J. C., & Antonsen, L. (2014). Effectiveness of cognitive behavioral group therapy for depression in routine practice. *BMC Psychiatry*, *14*(1), 1-9. https://doi.org/10.1186/s12888-014-0292-x

Wang, N. (2018). How to conduct a meta-analysis of proportions in R: a comprehensive tutorial. *John Jay College of Criminal Justice*.

*Ward, S., Turner, A., A Hambridge, J., A Halpin, S., E Valentine, M., L Sweetapple, A., & Hackett, M. L. (2016). Group cognitive behavioural therapy for stroke survivors with depression and their carers. *Topics in Stroke Rehabilitation*, *23*(5), 358-365. https://doi.org/10.1080/10749357.2016.1143706

*Voderholzer, U., Beintner, I., Backes, B., Esguerra, E., & Hessler-Kaufmann, J. B. (2021). Implementing Videoconference CBT for Depression in Routine Outpatient Care: Outcome, Working Alliance, and Influence of Patients' Technology Commitment. *Verhaltenstherapie*, *31*(3), 238-247. https://doi.org/10.1159/000513643

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy drop-out. *Professional Psychology: Research and Practice*, *24*(2), 190. https://doi.org/10.1037/0735-7028.24.2.190

Zimmermann, D., Rubel, J., Page, A. C., & Lutz, W. (2017). Therapist effects on and predictors of non-consensual drop-out in psychotherapy. *Clinical Psychology & Psychotherapy*, *24*(2), 312-321. https://doi.org/10.1002/cpp.2022

Appendix A

Population, Intervention, Comparison and Outcome (PICO) Tool Table

Item	Detail
Population (P)	Adults (16+ years) accessing routine care anywhere in the world
Intervention (I)	Low-intensity depression interventions (as defined by review)
Comparison (C)	Any context, no control groups
Outcome (O)	Drop-out rate data as defined by review

Appendix B

Psychotherapy Outcome Study Methodology

Rating Form (POSMRF; Öst, 2008) – Blank Copy

Note: If not enough information is given regarding a specific item a rating of 0 is given. Items marked with an asterisk (*) were omitted from the final rating form.

1. Clarity of sample description

0 Poor. Vague description of sample (e.g. only mentioned whether patients were diagnosed with the disorder).

1 Fair. Fair description of sample (e.g. mentioned inclusion/exclusion criteria, demographics, etc.).

2 Good. Good description of sample (e.g. mentioned inclusion/exclusion criteria, demographics and the prevalence of comorbid disorders).

2. Severity/chronicity of the disorder

0 Poor. Severity/chronicity was not reported and/or subsyndromal patients were included in the sample.

1 Fair. All patients met the criteria for the disorder. Sample includes acute (o1 yr) and/or low severity.

2 Good. Sample consisted entirely of chronic (41 yr) patients of at least moderate severity.

3. Representativeness of the sample

0 Poor. Sample is very different from patients seeking treatment for the disorder (e.g. there are excessively strict exclusion criteria).

1 Fair. Sample is somewhat representative of patients seeking treatment for the Disorder (e.g. patients were only excluded if they met criteria for other major disorders).

2 Good. Sample is very representative of patients seeking treatment for the disorder (e.g. authors made efforts to ensure representativeness of sample).

4. **Reliability of the diagnosis in question**

0 Poor. The diagnostic process was not reported, or not assessed with structured interviews by a trained interviewer.

1 Fair. The diagnosis was assessed with structured interview by a trained interviewer.

2 Good. The diagnosis was assessed with structured interview by a trained interviewer and adequate inter-rater reliability was demonstrated (e.g. kappa coefficient).

5. Specificity of outcome measures

0 Poor. Very broad outcome measures, not specific to the disorder (e.g. SCL-90R total score).

1 Fair. Moderately specific outcome measures.

2 Good. Specific outcome measures, such as a measure for each symptom cluster.

6. Reliability and validity of outcome measures

0 Poor. Measures have unknown psychometric properties, or properties that fail to meet current standards of acceptability.

1 Fair. Some, but not all measures have known or adequate psychometric properties.

2 Good. All measures have good psychometric properties. The outcome measures are the best available for the authors' purpose.

7. Use of blind evaluators

0 Poor. Blind assessor was not used (e.g. assessor was the therapist, assessor was not blind to treatment condition, or the authors do not specify).

1 Fair. Blind assessor was used, but no checks were used to assess the blind.

2 Good. Blind assessor was used in correct fashion. Checks were used to assess whether the assessor was aware of treatment condition.

8. Assessor training

0 Poor. Assessor training and accuracy are not specified, or are unacceptable.

1 Fair. Minimum criterion for assessor training is specified (e.g. assessor has had specific training in the use of the outcome measure), but accuracy is not monitored or reported.

2 Good. Minimum criterion of assessor training is specified. Inter-rater reliability was checked, and/or assessment procedures were calibrated during the study to prevent evaluator drift.

9. Assignment to treatment

0 Poor. Biased assignment, e.g. patients selected their own therapy or were assigned in another non-random fashion, or there is only one group.

1 Fair. Random or stratified assignment. There may be some systematic bias but not enough to pose a serious threat to internal validity. There may be therapist by treatment confounds. N may be too small to protect against bias.

2 Good. Random or stratified assignment, and patients are randomly assigned to therapists within condition. When theoretically different treatments are used, each treatment is provided by a large enough number of different therapists. N is large enough to protect against bias.

10. Design

0 Poor. Active treatment vs. WLC, or briefly described TAU.

1 Fair. Active treatment vs. TAU with good description, or placebo condition.

2 Good. Active treatment vs. another previously empirically documented active treatment.

11. Power analysis

0 Poor. No power analysis was made prior to the initiation of the study.

1 Fair. A power analysis based on an estimated effect size was used.

2 Good. A data-informed power analysis was made and the sample size was decided accordingly.

12. Assessment points

0 Poor. Only pre- and post-treatment, or pre- and follow-up.

1 Fair. Pre-, post-, and follow-up o1 year.

2 Good. Pre-, post-, and follow-up X1 year.

13. Manualized, replicable, specific treatment programs

0 Poor. Description of treatment procedure is unclear, and treatment is not based on a publicly available, detailed treatment manual. Patients may be receiving multiple forms of treatment at once in an uncontrolled manner.

1 Fair. Treatment is not designed for the disorder, or description of the treatment is generally clear and based on a publicly available, detailed treatment manual, but there are some ambiguities about the procedure. Patients may have received additional forms of treatment, but this is balanced between groups or otherwise controlled.

2 Good. Treatment is designed for the disorder. A detailed treatment manual is available, and/or treatment is explained in sufficient detail for replication. No ambiguities about the treatment procedure. Patients receive only the treatment in question.

14. Number of therapists

0 Poor. Only one therapist, i.e. complete confounding between therapy and therapist.

1 Fair. At least two therapists, but the effect of therapist on outcome is not analyzed.

2 Good. Three, or more therapists, and the effect of therapist on outcome is analyzed.

15. Therapist training/experience

0 Poor. Very limited clinical experience of the treatment and/or disorder (e.g. students).

1 Fair. Some clinical experience of the treatment and/or disorder.

2 Good. Long clinical experience of the treatment and the disorder (e.g. practicing therapists).

16. Checks for treatment adherence

0 Poor. No checks were made to assure that the intervention was consistent with protocol.

1 Fair. Some checks were made (e.g. assessed a proportion of therapy tapes).

2 Good. Frequent checks were made (e.g. weekly supervision of each session using a detailed rating form).

17. Checks for therapist competence

0 Poor. No checks were made to assure that the intervention was delivered competently.

1 Fair. Some checks were made (e.g. assessed a proportion of therapy tapes).

2 Good. Frequent checks were made (e.g. weekly supervision of each session using a detailed rating form).

18. Control of concomitant treatments (e.g. medications)

0 Poor. No attempt to control for concomitant treatments, or no information about concomitant treatments provided. Patients may have been receiving other forms of treatment in addition to the study treatment.

1 Fair. Asked patients to keep medications stable and/or to discontinue other psychological therapies during the treatment.

2 Good. Ensured that patients did not receive any other treatments (medical or psychological) during the study.

19. Handling of attrition

0 Poor. Proportions of attrition are not described, or described but no dropout analysis is performed.

1 Fair. Proportions of attrition are described, and dropout analysis or intent-to-treat analysis is performed.

2 Good. No attrition, or proportions of attrition are described, dropout analysis is performed, and results are presented as intent-to-treat analysis.

20. Statistical analyses and presentation of results

0 Poor. Inadequate statistical methods are used and/or data are not fully presented.

1 Fair. Adequate statistical methods are used but data are not fully presented.

2 Good. Adequate statistical methods are used and data are presented with M and SD.

21. Clinical significance

0 Poor. No presentation of clinical significance was done.

1 Fair. An arbitrary criterion for clinical significance was used and the conditions were compared regarding percent clinically improved.

2 Good. Jacobson's criteria for clinical significance were used and presented for a selection (or all) of the outcome measures, and conditions were compared regarding percent clinically improved.

***22. Equality of therapy hours (for non-WLC designs only)**

0 Poor. Conditions differ markedly (X20% difference in therapy hours).

1 Fair. Conditions differ somewhat (10–19% difference in therapy hours).

2 Good. Conditions do not differ (o10% difference in therapy hours).

Appendix C

Study Exclusion Reasons with Excluded Study Titles (n=49)

Randomised control trials (*k=16*)
Guided Internet-Based Cognitive Behavioral Therapy for Depression: Implementation Cost-Effectiveness Study
HIV patient and provider feedback on a telehealth collaborative care for depression intervention
Cognitive behavioral therapy in depressed cardiac surgery patients: role of ejection fraction
Using bibliotherapy to assist people to recover from depression in Thailand: Relationship between resilience, depression and psychological distress
Brief cognitive behavioral therapy for depression among patients with alcohol dependence in Thailand
Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data
A psychoeducational intervention (SWEEP) for depressed women with diabetes
Cost-effectiveness of cognitive behaviour therapy versus talking and usual care for depressed older people in primary care
Computerised cognitive behavioural therapy for the treatment of depression in people with multiple sclerosis: external pilot trial
Incremental benefit and cost of telephone care management and telephone psychotherapy for depression in primary care
Cognitive-behavioural therapy v. usual care in recurrent depression
Cost-effectiveness of treatments for major depression in primary care practice
The 'usual care' of major depression in primary care practice
Treating depressed primary care patients improves their physical, mental, and social functioning
Technology-facilitated depression care management among predominantly Latino diabetes patients within a public safety net care system: Comparative effectiveness trial design
The effects of psychosocial day care program on clinical symptoms and quality of life of persons with depression: a prospective study

Systematic Review (*k=2*)
Treatment of depression in cancer patients
Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health: A Report of the EFPA E-Health Taskforce

Conceptual framework (*k=1*)
The design of Partners in Care: evaluating the cost-effectiveness of improving care for depression in primary care

Unable to gain access (*k=2*)
CanDirect: Effectiveness of a Telephone-Supported Depression Self-Care Intervention for Cancer Survivors
Depression program found more effective but more costly than usual care

Participants chose to take part (*k=1*)
Outcome of cognitive behaviour therapy for minor depression in routine practice

No drop-out data (*k=5*)
Usual care psychotherapy for depression in a large managed behavioral health organization
The clinical effectiveness of cognitive behaviour therapy: outcome for a large sample of adults treated in routine practice
Uptake of Web-based clinical resources from the MacArthur Initiative on Depression and Primary Care
Predictors of Poor Response to Depression Treatment in Primary Care
Clinical effectiveness of cognitive behavioral therapy for depression in routine care: A propensity score based comparison between randomized controlled trials and clinical practice

Depression drop-out not separated from other diagnosis (*k=7*)
The implementation of computerized cognitive behavioural therapies in a service user-led, third sector self help clinic.
The effectiveness of computerized cognitive behavioural therapy in routine care.
A Pilot Study of a Cognitive-Behavioral Treatment for Anxiety and Depression in Patients With Parkinson Disease
Taking computerized CBT beyond primary care
Process and outcome of a non-guided self-help manual for anxiety and depression in primary care: A pilot study
An Online Mental Health and Wellness Intervention Supplementing Standard Care of Depression and Anxiety
Examining an internet-delivered intervention for anxiety and depression when delivered as a part of routine care for university students: A phase IV trial

High-Intensity Treatment (*k=7*)
Group cognitive-behavioral therapy for clients with major depression in residential substance abuse treatment

The clinical effectiveness of evidence-based interventions for depression: a pragmatic trial in routine practice
Targeted prescription of cognitive-behavioral therapy versus person-centered counseling for depression using a machine learning approach
Sudden gains in routine care cognitive behavioral therapy for depression: A replication with extensions
Effectiveness of a postnatal psychological treatment for women who had screened positive for depression
In-home intervention for depressive symptoms with low-income mothers of infants and toddlers in the United States
Telephone counseling as an adjunct to antidepressant treatment in the primary care system. A pilot study

Protocol for prospective study (*k=1*)
The therapist's role in the implementation of internet-based cognitive behavioural therapy for patients with depression: Study protocol

Depression not main presentation (*k=7*)
Evaluation of a pilot innovative cognitive-behavioral therapy-based psychoeducation group treatment for functional non-epileptic attacks
Effectiveness of six-week psychoeducation program on adherence of patients with bipolar affective disorder
Improving access to psychological therapies (IAPT) for people with bipolar disorder: Summary of outcomes from the IAPT demonstration site
A Telehealth Intervention for Veterans on Antiviral Treatment for the Hepatitis C Virus
The effect of nursing self-care educational intervention on depression in women with breast cancer undergoing post-mastectomy chemotherapy: A quasi-experimental study
Depression burden, self-help interventions, and side effect experience in women receiving treatment for breast cancer
Cost-effectiveness of a psychoeducational relapse prevention program for depression in primary care

Appendix D

Risk of Bias (RoB) Assessment Summary Table

Study	Ruwaard et al. 2012	Hedman et al. 2014	Nordgreen et al. 2019	Thimm & Antonsen 2014	Jakobsen et al. 2017	Baumel et al. 2018	Mathieson et al. 2018	Ward et al. 2016	Voderholzer et al. 2021	Cavanagh et al. 2011	Learmonth et al. 2008
Clarity of sample description	Fair	Good	Fair	Good	Good	Fair	Fair	Fair	Good	Fair	Fair
Severity/ chronicity of disorder	Fair	Fair	Fair	Fair	Poor	Good	Poor	Poor	Poor	Fair	Poor
Representat-iveness of sample	Good	Fair	Fair	Good	Fair	Good	Fair	Fair	Fair	Fair	Good
Reliability of diagnosis in question	Fair	Fair	Fair	Good	Fair	Good	Fair	Poor	Poor	Fair	Good
Specificity of outcome measures	Good	Good	Good	Good	Good	Good	Fair	Good	Good	Good	Good
Reliability and validity of outcome measures	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good
Use of blind evaluators	Poor	Poor	Poor	Poor	Poor	Fair	Poor	Poor	Poor	Poor	Poor
Assessor training	Poor	Fair	Fair	Poor	Poor	Poor	Fair	Poor	Poor	Poor	Poor
Assignment to treatment	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor
Design	Poor	Poor	Poor	Poor	Poor	Fair	Poor	Poor	Poor	Poor	Poor

Power Analysis	Poor	Poor	Poor	Poor	Poor	Good	Poor	Poor	Poor	Poor	Poor
Assessment points	Good	Fair	Fair	Fair	Fair	Poor	Poor	Fair	Poor	Poor	Poor
Manualised, replicable, specific treatment programs	Good	Good	Good	Good	Good	Fair	Fair	Good	Poor	Fair	Fair
Number of therapists	Good	Fair	Fair	Fair	Fair	Poor	Fair	Fair	Good	Poor	Fair
Therapist training/experience	Good	Good	Good	Fair	Poor	Poor	Good	Good	Fair	Fair	Good
Checks for treatment adherence	Fair	Poor	Poor	Poor	Poor	Poor	Fair	Poor	Poor	Poor	Poor
Checks for treatment competence	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Fair	Poor
Control of concomitant treatments (e.g. medications)	Fair	Fair	Fair	Poor	Fair	Poor	Poor	Poor	Poor	Poor	Poor
Handling of attrition	Fair	Good	Good	Fair	Fair	Poor	Good	Fair	Fair	Poor	Fair
Statistical analysis and presentation of results	Good	Good	Good	Good	Good	Poor	Good	Good	Good	Good	Good

Clinical significance	Good	Poor	Good	Good	Fair	Poor	Poor	Fair	Good	Fair	Good
Equality of therapy hours	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Overall score and Rating	24/44 "Well above average"	21/44 "Well above average"	20/44 "Above Average"	21/44 "Well above average"	17/44 "Below average"	16/44 "Below average"	16/44 "Below average"	16/44 "Below average"	15/44 "Below average"	13/44 "Well below average"	18/44 "Below average"

Note: Poor = 0, Fair = 1, Good = 2.

This page is intentionally blank

Section Two: Research Report

Developing and evaluating a measure of treatment integrity for the delivery of group psychoeducation for common mental health problems

Abstract

Objectives

Despite the importance of reliably assessing the manner in which low intensity (LI) psychological interventions are delivered, there is currently no measure of treatment integrity for group LI psychoeducational interventions. This study aimed to develop and evaluate a group LI psychoeducational intervention treatment integrity measure for use in routine service settings and research trials.

Design

A psychometric validation design was used to assess the reliability and validity of an 9-item observational measure of treatment integrity; the group psychoeducation treatment integrity measure (GPTIM).

Method

A literature search was used to develop the GPTIM, constituent items and a detailed scoring manual. Firstly, expert rater opinions of the GPTIM determined content validity. Secondly, recordings of group psychoeducational sessions (n=10) sessions delivered in routine practice were made, which were then rated by psychological wellbeing practitioners (n=8) using the GPTIM. A subset of patients (n=11) receiving the intervention also rated the sessions, alongside the collection of outcome data and service utilization indices.

Results

The GPTIM was found to be a single factor scale with excellent content validity, good to excellent construct validity and reasonable predictive validity. The GPTIM had excellent internal consistency and good test-retest reliability, but had poor inter-rater reliability.

Conclusions

The GPTIM has promising, but not complete, psychometric properties. With further testing, the GPTIM has the potential to contribute to the clinical governance of group psychoeducational interventions, training of LI practitioners and research trial methods.

Practitioner Points and Limitations

- The clinical governance of group psychoeducation is an important, but neglected, topic.
- The GPTIM demonstrated an array of acceptable psychometric properties and is supported by a detailed manual, indicating its potential future use in routine services and training contexts.
- Aspects of the GPTIM and rating process need fine-tuning to further test certain components of its psychometric evidence base (i.e. inter-rater reliability).
- The main study limitations were the small sample size (in relation to facilitators, patients and session recordings) and the relatively brief rater training.

Keywords

Treatment Integrity; Competency; Adherence; Differentiation; Psychoeducation.

Introduction

In England, the amount of people experiencing common mental health problems (CMHPs), such as anxiety and depression, increased by 20% between 1994 and 2014 (McManus et al., 2016). This translates to around 1 in 6 adults (16+ years old) having experienced a CMHP in the last week (Baker, 2020). The impact of CMHPs are often ubiquitous and can include considerable social and occupational impairments (APA, 2013). Due to the impact of CMHPs and their increased prevalence, both anxiety and depression are recognised as major ongoing public health problems (Bastiampillai et al., 2019) and the necessity of access to evidence-based psychological therapies has increased (Pollecoff, 2016). Accordingly, to provide a framework for the effective delivery of interventions, the National Institute for Health and Care Excellence (NICE) provides guidelines that recommend evidence-based psychological interventions within England (NICE, 2011).

In response to the prevalence of CMHPs within the population and the need to deliver evidence-based interventions, the UK Government funded the Improving Access to Psychological Therapies (IAPT) programme to implement NICE guidelines nationally (Clarke et al., 2018). IAPT services are based on a stepped care model, which is an organising framework that provides patients with progressively intensive effective psychological interventions through a care pathway (Bower & Gilbody, 2005; Firth et al., 2015). Stepped care has been shown to produce better outcomes than traditional care (Firth et al., 2014). The model therefore provides patients with low-intensity (LI) interventions for mild-to-moderate anxiety and depression at step-2 or high-intensity (HI) interventions at step 3 dependent on need, severity, risk and responsivity to the initial LI intervention (Boyd et al., 2019).

The LI interventions at step 2 are brief (≤8 sessions), psychoeducational, guided self-help (GSH) in nature, based on the principles of cognitive behavioural therapy (CBT)

and delivered via numerous mediums (via telephone, computerised CBT, in large groups or in a one-to-one format; Wakefield et al., 2021). Within IAPT services, GSH is delivered by psychological well-being practitioners (PWPs) in a one-to-one or group format (CSIP, 2008). PWPs deliver psychological interventions more as a 'coach' rather than as a traditional therapist (Turpin, 2010). The training of PWPs development and learning is competency driven and mostly involves them completing observation based clinical exams assessing their fidelity to manualised LI approaches (Richards & Whyte, 2009).

To meet service demand, IAPT services are increasingly using psychoeducational groups to deliver GSH to a greater proportion of patients (Burns et al., 2016). Such groups are either delivered in a didactic teacher-student lecture large group style (e.g. *Stress Control;* White & Keenan, 1990) or via smaller group interactive workshops (Wykes, 2013). Notably, large psychoeducational groups have been found to be effective interventions for the treatment of depression (Cuijpers, Muñoz, Clarke, & Lewinsohn, 2009; De Souza Tursi et al., 2013) and anxiety (Delgadillo et al., 2016; Houghton & Saxon, 2007). Psychoeducational groups therefore embody the "low-contact/high-volume" approach underpinning all LI work (Clark et al., 2009). Thus, the large psychoeducational group format provides an empirically supported, accessible and brief intervention that upholds the ongoing objectives of the IAPT programme.

Nevertheless, Delgadillo et al. (2016) found significant outcome variability between large didactic psychoeducational *Stress Control* groups (White & Keenan, 1990). These findings indicate that, despite *Stress Control* groups being manualised, moderator variables appear to be present that influence patient outcomes and these may include the manner in which groups are delivered. Treatment integrity (TI; interchangeably referred to as treatment fidelity), which is an index of whether or not a treatment is delivered as intended (Yeaton & Sechrest, 1981), has been noted as pivotal to ensuring the effective implementation of evidence-based interventions (Landsverk, 2013). TI is comprised of

three elements: treatment adherence, therapist competency and treatment differentiation (Perepletchikova et al., 2007). Adherence refers to the extent to which an intervention is delivered consistently with the treatment manual (Perepletchikova & Kazdin, 2005), competence refers to the skill with which an intervention is implemented (Sharpless & Barber, 2009) and differentiation refers to the extent treatments differ from one another as defined by the treatment manual (Southam-Gerow & McLeod, 2013).

Despite the importance of TI, the link between integrity and outcome has previously been unclear (Muse & McManus, 2013). Nonetheless, the assessment of TI remains a critical component of implementing evidence-based interventions, as it strengthens confidence in client outcomes, whilst also improving intervention replicability (Boyle et al., 2020; McCay et al., 2016). Waltz et al. (1993) posit that without the reliable assessment of TI, it is difficult to draw conclusions about the effectiveness of a specific treatment in routine practice. Moreover, a recent meta-analysis found a positive significant association between TI and therapeutic outcomes (whereas adherence and competency alone did not; Power et al., 2022). Recommendations from this meta-analysis suggest that services should aim to assess TI as part of therapist training, during routine session delivery and within clinical supervision.

In IAPT services, however, the use of TI measures has not happened. Instead, competency measures (which the literature has developed more frequently) are more regularly employed for training and clinical governance purposes. For example, to facilitate the delivery of one-to-one HI CBT treatments, IAPT services implement the Cognitive Therapy Scale Revised (CTS-R; Blackburn et al., 2001). More recently, competency measures of assessment for and treatment of anxiety and depression have also been developed and evaluated (Kellett et al., 2020). Despite a large number of patients accessing LI-CBT interventions (e.g., 1,010,000 began a course of IAPT treatment, with 555,000 finishing treatment in 2017-18) and a high proportion of these accessing

psychoeducational groups, no valid and/or reliable tool for rating the manner in which group psychoeducation is delivered currently exists. When considering the widespread use of psychoeducational groups and their increased uptake by IAPT services (Burns, Kellett & Donohoe, 2016), a noticeable gap for a LI group TI facilitator measure is apparent. A valid and reliable group facilitator TI measure could ensure appropriate evidence-based delivery of groups in routine practice (Noble et al., 2020), better case management supervision (Shepherd & Rosairo, 2008) and be used to improve the internal validity of research trials (Dinger et al., 2015). Such a measure would also allow IAPT services to effectively support PWP training and development, which could ensure trainees are more likely to deliver treatments correctly and in a skilled manner irrespective of experience (Delgadillo et al. 2016). Finally, such a measure could help to further analyse the association between facilitator TI and patient outcomes (adding to the Webb et al. (2010) and Power et al. (2022) reviews).

In response to the absence of a measure assessing the competencies of LI group psychoeducation facilitators, Noble et al. (2020) completed a Delphi study that produced 36 CBT group-based facilitator competencies. Final competencies were based around four subcategories: group setup, group content, group process and group closure. As the skills required for facilitating LI group interventions are likely specialised (Barlow, 2012), the items from this study provided the foundation from which an integrity measure could be developed and evaluated. Considering the findings of the Power et al. (2022) meta-analysis, it was decided to expand the realms of the measure to also include adherence and differentiation items. This study therefore aimed to develop a TI measure for the facilitators of CBT-based psychoeducational groups (*Stage One*). Once developed, the study then endeavored to test the reliability and validity of the measure (*Stage Two*). The aims of the study were to: (1) develop a TI measure and accompanying manual to assess the TI of facilitators delivering LI-CBT psychoeducational groups and (2) test the validity

(predictive, construct and content) and reliability (inter-rater, test-retest reliability and internal consistency) of the measure in a cross-over design.

Ethics

The study was ethically reviewed and granted approval (Appendix A; IRAS reference number: 293765). The University of Sheffield also granted ethical approval (URMS number: 170907). Group facilitators consented to their sessions being recorded and rated, whilst raters also consented. Similarly, patients completing evaluations of facilitators consented to their involvement. All participants could withdraw at any point. Information and consent sheets are displayed in Appendix B.

Stage One: Method

Development of the Framework, Measures and Manual

The initial stage of the study aimed to develop a measure and a manual centered around the TI components related to the facilitation of LI psychoeducational groups. The group psychoeducation TI measure (GPTIM) (also referred to as the *expert measure*; Appendix C) and the patient GPTIM (also referred to as the *patient measure*; Appendix C) were then developed using items selected from a recent Noble et al., (2020) Delphi study. Additionally, the cognitive behavioural maintenance-adherence scale (CBMT-AS; Weck, et al., 2011), which evaluates therapists adherence to the treatment manual for depression, was also referred to. Key texts (Bennett-Levy et al., 2010; Brown, 2018) were also perused to examine crucial skills for facilitators within psychoeducational groups, along with the UCL CBT interventions competency framework (Roth & Pilling, 2007). Finally, the LI-CBT adherence measure was consulted (Richards & Whyte, 2009). This approach,

whereby various relevant sources were used to inform the item generation process, was employed to ensure the measure was comprehensive and inclusive. A process of truncation was then performed to remove duplicates, whereby each of the identified items was discussed at length, before nine final items were decided upon. Out of these nine items, six were facilitator competencies and three were facilitator adherence and differentiation items. The GPTIM was designed to be underpinned by the behaviour change model (COM-B model; Michie et al., 2011), which is a widely adopted framework that provides the conceptual basis of LI-CBT assessment and treatment work (UCL, 2015).

The patient version of the GPTIM was developed to parallel the items included within the expert GPTIM. Although patients would be unable to evaluate some of the more technical and nuanced items of the expert measure (i.e. differentiating CBT from other approaches), there were some items that could be reworded and appropriately appraised by a service user of a psychoeducational group intervention (setting clear aims for the group and homework setting, for example). After discussions and a further period of development, the research team created a nine-item measure that resembled the expert measure.

Each GPTIM item was rated using the Dreyfus (1989) 7-point rating system, as this is a recognised approach that can capture performances ranging from novice to expert. The patient GPTIM adopted a more accessible 5-point Likert scale which provided patients the opportunity to rate how much they agree with a statement (from strongly disagree to strongly agree). Item ratings for the GPTIM and the patient GPTIM were presented in ascending order without any overlap. The final score of both measures are calculated as a total sum of their individual item ratings, with a scoring range between 0-54.

The 18-page manual was created to assist expert raters in consistently completing the GPTIM (Appendix D). In order to create consistency between raters, the manual comprehensively described each point on the rating scale. The GPTIM manual provides raters with the theory underpinning the measure, rating guidelines and an item introduction. TI items also covered the course of the group sessions, accounting for the start (i.e. agenda setting), middle (i.e. psychoeducational content) and end of sessions (i.e. homework setting).

During the final stages of the measure development, the GPTIM and its accompanying manual were used independently by all five members of the research team to rate a sample *Stress Control* (White & Keenan, 1990) session. Intra-class correlation coefficients (ICCs) were produced to assess the levels of consistency in the ratings between the research team on the GPTIM (Koo & Li, 2016). Overall, the ICC score was .83 (95% CI .56 to .96), which is recognised as a 'good' level of agreement (Koo & Li, 2016). For the ratings that were not fully agreed, joint agreement was achieved before ratings were finalised. Once finalised, some minor amendments were made to the manual to achieve greater item clarity. Subsequently, the measure was deemed ready for the next stage of development.

Participants

As part of the development of the GPTIM, six senior PWPs working in IAPT services were recruited to evaluate the relevance of the final nine TI items. A convenience sampling technique was employed to recruit raters for this stage of the study, to ensure that those holding senior positions, who also had extensive experience of delivering psychoeducational groups (delivered ≥10 groups each), were consulted.

Design

Stage one of the study employed a quantitative design to assess the degree of agreement regarding the relevance of the expert measure items using the content validity index (CVI; Lynn, 1986). Six raters were recruited for the procedure as this is endorsed by relevant research (Yusoff, 2019).

Procedure

Once the CVI raters consented to their involvement, the GPTIM and manual were sent to them electronically. After having familiarized themselves with the manual and measure, each rater rated the relevance of the proposed final items. The relevance of these items was rated on a 4-point ordinal scale (1=not relevant, 2=somewhat relevant, 3=quite relevant, 4=highly relevant), which is widely recommended due to its avoidance of a neutral midpoint (Lynn, 1986; Waltz & Bausell, 1981). Ratings were then sent to the research team, along with some demographic information regarding the raters (age, gender, years qualified and number of psychoeducational groups facilitated in their career). A copy of the CVI questionnaire accessed by the experts is included in Appendix E.

Data Analysis

To produce a CVI score for the GPTIM, the Polit & Beck (2006) guidelines were followed to produce item level (I-CVI) and scale level (S-CVI) scores. The I-CVI was calculated by dividing the number of raters who gave an item a relevance rating of 3 or 4 with the number of overall raters (I-CVI = (agreed items)/(number of experts)). The S-CVI was calculated using two methods. The S-CVI/UA was calculated by adding the proportion of items on the scale that achieved a relevance score of 3 or 4 by all experts (S-CVI/UA = (sum of UA scores)/(number of measure items)). The S-CVI/Av was calculated by

averaging the proportion of items rated relevant across experts (S-CVI/Av = (sum of proportion scores)/number of experts)). According to Lynn (1986), acquiring convergent scores for items over 0.67 is considered acceptable, whereas a CVI rating higher than 0.90 is deemed excellent (Polit & Beck, 2006).

Stage One: Results

Content Validity

The CVI of the expert measure was calculated to assess the relevance of the proposed final GPTIM items. Raters had a mean age of 42.8 years (*SD=8.09; range=35-59*) and were 66% female (n=4). They were qualified for a mean of 9.7 years (*SD=4.71; range=3-18*) and had delivered a mean of 60 LI psychoeducational groups (*SD=63.77; range=30-200*). The I-CVI ranged from 0.83-1.0, with a mean score of 0.96 across all items ('excellent'). The S-CVI/UA score was 0.78 ('acceptable'). Table 1 reports the full CVI results. It was therefore concluded that the I-CVI, S-CVI/Ave and S-CVI/UA had satisfactory content validity. Accordingly, the second stage of the research (where the measure trialed and evaluated) could be conducted.

Table 1

Content Validity Index (CVI) Scores and Calculations

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	No. in agreement	Item CVI
Psychoeducational Approach	X	X	X	X	X	X	6	1.00
Well matched psychoeducation	X	X	X	X	X	X	6	1.00
Underpinned by CBT	X	X	X		X	X	5	0.83
Agenda	X	X	X	X	X	X	6	1.00
Pacing	X	X	X	X	X	X	6	1.00
Engaging and enthusiastic	X	X	X	X	X	X	6	1.00

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	No. in agreement	Item CVI
Clear and accurate communication	X	X	X	X	X	X	6	1.00
Change methods	X	X	X	X	X	X	6	1.00
Homework	X	X	X		X	X	5	0.83
Proportion Relevant	1.00	1.00	1.00	0.78	1.00	1.00	S-CVI-Ave	0.96
							Mean I-CVI	0.96
							S-CVI/UA	0.78

Stage Two: Method

The second stage of the study aimed to test various psychometric variables of the GPTIM, including construct and predictive validity and internal, test-retest and inter-rater reliabilities.

Setting

The second stage of the study was conducted in collaboration with an IAPT service in the North of England. The service implements a stepped care model, providing evidence-based psychological interventions for adults (18+ years of age) with CMHPs. As part of this model, the service offered both LI (step 2) and HI (step 3) interventions, with the LI input consisting of 1:1 and group interventions (including *Managing Stress (*also known as *Stress Control)*). The current study only recorded and rated *Managing Stress* sessions.

Participants; Inclusion and Exclusion Criteria and Associated Demographics

Two groups of participants were involved in the second stage of the study: patients and expert raters. Patients (n=78) attending two separate *Managing Stress* psychoeducation groups were recruited. As part of their participation, their demographic, group outcome and attendance data were accessed. The mean age of all patients was 33 years (*SD=11.6 years, range=18-64* years). Over half were female (63%; n=49) and 83% of the total sample categorized their ethnicity as 'white' (n=65; with 6% selecting 'Asian or Asian British', 5% selecting 'another ethnic group', 4% selecting 'Black or Black British' and 1% who 'did not state' their ethnicity). Just over half (51%; n=40) attended the groups due to 'generalised anxiety disorder', with 42% (n=33) attending due to a 'depressive episode'. Remaining patients attended due to 'panic disorder' (n=2), 'recurrent depressive disorder' (n=1) and 'adjustment disorder' (n=2). Demographic data is outlined in Table 2.

A subsection of the patients consented to completing the patient GPTIM regarding their sessions. Participants could only be included if they were 18+ years old, met the criteria for an anxiety disorder, depressive disorder, co-morbid anxiety and depression disorder or other similar diagnosis (as assessed by the GAD7 or PHQ9) and could read and write in English. Patients were excluded if they were engaging in other psychological support within or outside the service or required an interpreter. Patients with an additional diagnosis (such as social phobia or post-traumatic stress disorder) that would be better treated using HI approaches were excluded.

Alongside the patients, a group of expert raters (n = 8) were recruited to provide GPTIM ratings. A convenience sampling procedure was utilised to recruit raters, who had put themselves forward in service meetings and following a recruitment email. The rater sample consisted of both PWPs (3) and senior PWPs (5), all of whom had facilitated ≥3 psychoeducational groups in line with the service and model protocols. The expert raters were a mean age of 43 years (*SD=10.3, range=28-60*), had delivered a mean of 61.4

psychoeducational groups (*SD =79.1, range=6-250*) and had been qualified a mean of 6 years (*SD=4.1, range=2-13).* All expert raters were also currently providing supervision to other PWPs and had delivered supervision for a mean of 3 years (*SD=3, range=2 months-7 years).* Six of the expert rater group were female (75%).

Table 2

Patient Demographic Data (n=78)

Characteristic	Number	Overall Percentage of Group
Age in years: mean (*SD*) [range]	33 (11.6) [18-64]	
Gender		
Female	49	63%
Male	29	37%
Ethnicity		
White	65	83.3%
Asian/Asian British	5	6.4%
Other	4	5.1%
Black/Black British	3	3.8%
Did not state	1	1.3%
Presentation		
Generalised anxiety disorder	40	51.2%
Depressive episode	33	42.3%
Panic disorder	2	2.6%
Adjustment disorder	1	1.3%
Recurrent depressive disorder	2	2.6%

Procedure

Before and during the rating process, group psychoeducational sessions (n=10) were recorded to generate a collection of sessions where a pair of facilitators had delivered a LI psychoeducational group. Recordings were made across two separate didactic *Managing Stress* groups, resulting in the involvement of two separate PWP facilitator pairs. All PWPs who delivered the groups were qualified, which involved the completion of a 1-year post-graduate certificate in LI interventions accredited by the British

Psychological Society (BPS). Three out of the four PWP facilitator sample were female and facilitators had a mean age of 32 years (*SD*=6.30) years. Mean years post-qualification PWP experience was 2.5 years (*SD*=1.11). Each facilitator also received weekly case management supervision. All group sessions were delivered online. Therefore, facilitators were required to record their own sessions and securely transfer the recording to the researchers before dissemination to the raters.

Prior to the rating process, all expert raters received a pre-recorded two-hour training session (presentation available in Appendix F) that introduced the research rationale, the GPTIM manual and the measure. The training provided raters with the opportunity to trial the GPTIM alongside the manual, by rating a sample *Stress Control* session. Their ratings could then be compared to a benchmark score that was agreed by the research team (detailed within *Stage One* of the study). Once they self-reported their completion of the training and submitted their sample session scores, raters were then signed off. After training completion, raters completed an evaluation questionnaire regarding their level of agreement with six statements to gauge their impressions of the training (i.e. “The training workshop increased my knowledge of what TI, adherence, differentiation and competency are” etc.; Appendix G).

Following this, four *Managing Stress* sessions were securely sent to each of the expert raters. Each rater was given two recordings from one pair of facilitators, along with a second pair of recordings for the second pair of facilitators. Each rater submitted their ratings by completing the measure on Qualtrics, but also had access to a PDF version. This process resulted in the completion of 32 ratings (the power analysis detailing this decision is included in the *data analysis* subsection).

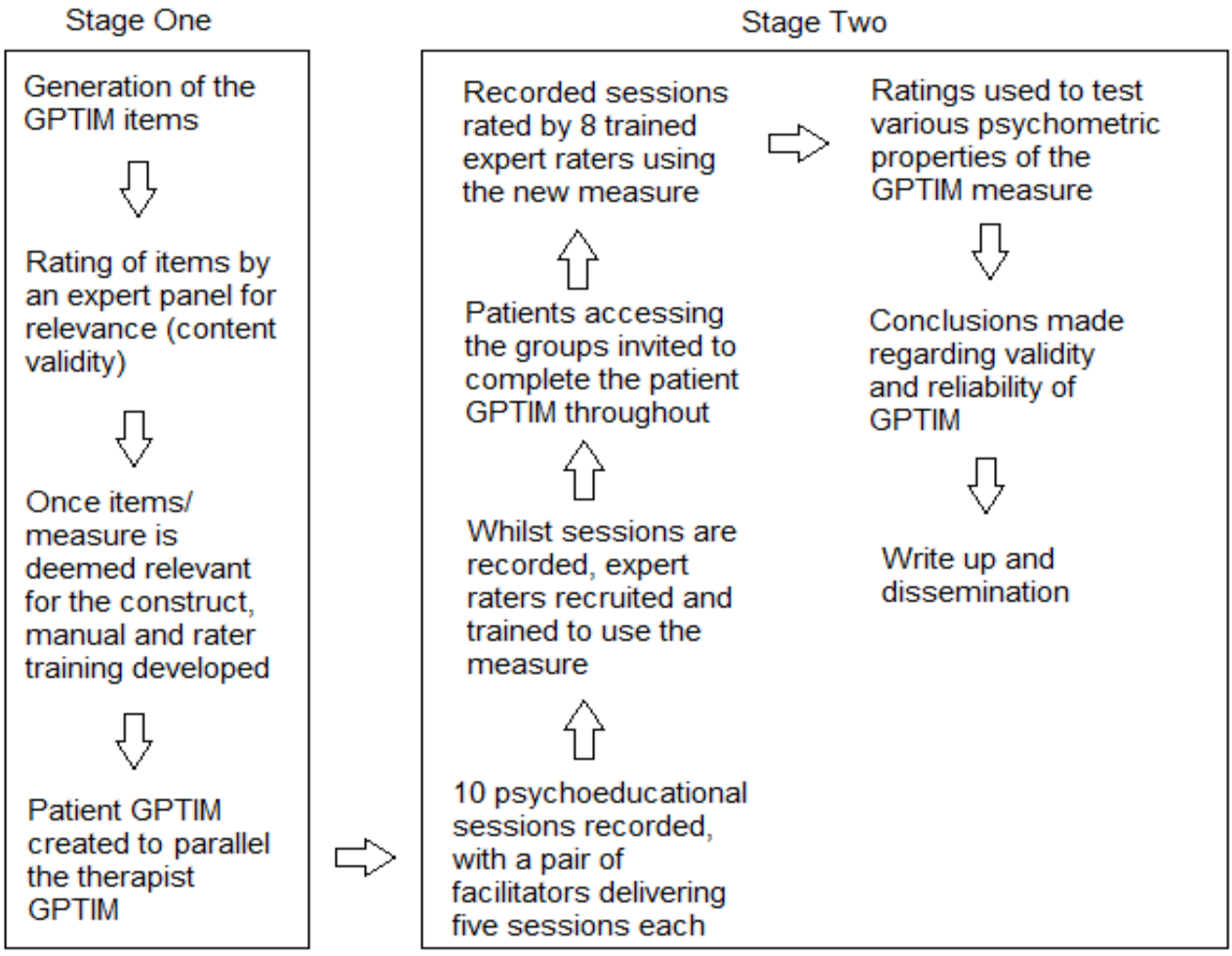
Throughout the group sessions, patients attending psychoeducational groups were asked to complete the patient GPTIM. Prior to group sessions, patients were sent a link so

they could complete the measure after the session on Qualtrics. Patients were asked to do this for all recorded group sessions. Prior to the commencement of the group sessions, patients were asked for their consent to be involved. As part of this process, they were all sent the information and consent sheets for their perusal. The information sheet informed the patients of their right to withdraw, what personal data we would be accessing and more (Appendix B; all patients consented). Following the completion of groups, patient group outcome data were retrieved from the Iaptus database. Iaptus is an electronic patient record that supports psychological services. Data that was accessed from the Iaptus included outcome scores, attendance, ethnicity, presentation/diagnosis, gender and age. The outcome scores included the Patient Health Questionnaire–9 (PHQ-9; Kroenke et al., 2001; Appendix H) and the General Anxiety Disorder–7 (GAD-7; Spitzer et al., 2006; Appendix H). From these, an overall change score was calculated, which was computed by taking away the final available outcome score from the first outcome score (known as the Last Observation Carried Forward (LOCF) approach, which is used as standard in IAPT services. Attendance was calculated in relation to the proportion of patients who dropped out of therapy. In this instance, drop-out was defined as an occasion when a patient attends <2 sessions (which is related to the IAPT manual definition of a sufficient dose of therapy (National Collaborating Centre for Mental health, 2018)). Additionally, two further definitions of drop-out were utilised to assess the impact on study findings. The additional definitions were: 1) patients who dropped out prior to attending all five group sessions, and 2) patients who attended less than 50% of sessions.

Figure 1 depicts the procedure for the whole study. This includes the process of therapist GPTIM and patient GPTIM development, along with the psychometric testing of the therapist GPTIM.

Figure 1

Complete Procedure of GPTIM Development and Trialling (including stage one and two)



Materials

Each expert rater watched and rated four one hour-long recorded group psychoeducational sessions. They used the manual and GPTIM to rate the TI of the pair of facilitators and electronically submitted their ratings via Qualtrics. Whilst accessing iaptus, scores for both the PHQ-9; (Kroenke et al., 2001) and the GAD-7 (Spitzer et al., 2006) were gathered. The PHQ-9 has high internal consistency (Beard et al., 2016), whilst demonstrating responsiveness to change, reliability and discriminant validity (Cameron et al., 2008). The GAD-7 can be used for a multitude of anxiety presentations (Spitzer et al.,

2006) and it is a reliability and valid as an anxiety measurement tool (Johnson et al., 2019; Löwe et al., 2008).

Data Analysis

Construct validity was investigated by calculating the correlations (using Pearson's correlation coefficient (*r*)) between the GPTIM scores and the patient GPTIM scores. Data were interpreted in relation to the cut-off ranges established by Portney and Watkins (2009), which deems scores of $p<0.25$ as small, p 0.25-0.50 as moderate, p 0.50-0.75 as good and $p>0.75$ as excellent.

To evaluate the predictive validity of the measure, Pearson's correlation coefficients (*r*) assessed the relationships between: 1) patient outcomes (using change scores) and GPTIM scores, and 2) patient drop-out rates and GPTIM scores. Patient outcomes were measured in relation to their GAD-7 and PHQ-9 scores (rated prior to the group and at every session attended). When analysing the association between patient drop-outs and GPTIM scores, only patients who received at least two sessions were included. This is because the IAPT literature outlines that once a patient have attended ≥2 sessions, they have engaged in the psychotherapeutic intervention (Clark, 2011, Gyani et al., 2013; IAPT Manual (National Collaborating Centre for Mental health, 2018)). Thus, drop-out was defined as those patients who attended <2 group sessions.

The test-retest reliability of the GPTIM was investigated by calculating the Pearson's correlation coefficients (*r*) to assess the relationship between the rater's first and second ratings of the same facilitator pair. This was assessed to ensure that the scale yields consistent results over a given time period (Walsh & Betz, 2001). The relationships were interpreted in accordance with the criteria outlined by Cicchetti (1994), where scores

of 0.40 to 0.59 are ‘fair’, 0.60 to 0.74 are deemed ‘good’ and scores of ≥0.75 are considered ‘excellent’.

Inter-rater reliability was also assessed through the completion of an ICC (Koo & Li, 2016). A one-way mixed effects ICC was used, where reliability was based on average ratings and absolute agreement (Hallgren, 2012; Shrout & Fleiss, 1979). ICC reliability levels were interpreted between the following ranges: <0.5 (poor), between 0.50-0.75 (moderate), 0.75-0.90 (good), and >0.90 (excellent) (Koo & Li, 2016). Prior to its use, a power analysis was conducted to calculate the amount of ratings needed to adequately utilise an ICC. In line with the literature (Walter et al., 1998) and a well-established online sample calculator (Arifin, 2018), the recommended number of ratings for the study was 32 (with ten sessions being rated by three raters each). The calculator was set with an alpha significance level of 0.05 and a power rating of 80%.

The internal consistency was investigated, as this relates to the estimate of how consistently individuals respond to items within a measure (Vaske et al., 2017). This was calculated by completing an exploratory factor analysis (EFA), producing Cronbach’s alpha scores and exploring descriptive statistics. Prior to conducting the EFA, preliminary tests assessed data suitability. Initially, correlations between all items were checked to ensure correlations of ≥0.3 were achieved (Critobal et al., 2007). The Kaiser-Meyer Olkin (KMO) measure and Bartlett’s Test of Sphericity were also conducted. The KMO measure rates scores between 0.00-0.49 as ‘unacceptable’, 0.50-0.59 as ‘miserable’, 0.60-0.69 as ‘mediocre’, 0.70-0.79 as ‘middling’, 0.80-0.89 as ‘meritorious’ and 0.90-1.00 as ‘marvelous’ (Dodge, 2008). Bartlett’s Test of Sphericity deems that if its score is lower than the chosen significance level, then the data set is suitable (Bartlett, 1950). In this instance, the level of significance was set at $p<0.05$. An EFA was used as it enables one to decide how many factors are necessary to explain a structure (Petkov et al., 2010). Resultantly, an EFA with principal axis factoring with a direct oblimin (oblique) rotation was used. Cronbach’s alpha

of the GPTIM was determined to indicate the internal consistency across all items (Cronbach, 1951). It is reported that scores above 0.70 are considered ‘acceptable’, with scores between 0.80 and 0.89 deemed ‘very good’ and scores above 0.90 judged as ‘excellent’ (Cortina, 1993). Item-total correlations and Guttman split-half reliabilities were also calculated. Item-correlation scores above the cut-off point of .30 are considered ‘acceptable’ (Field, 2013). Finally, descriptive statistics were calculated to examine the means, standard deviations, skewness and kurtosis of the GPTIM item data. These were assessed to check the normality of the data (Hopkins & Weeks, 1990). SPSS (IBM, 2015) was used to conduct all statistical analyses.

Stage Two: Results

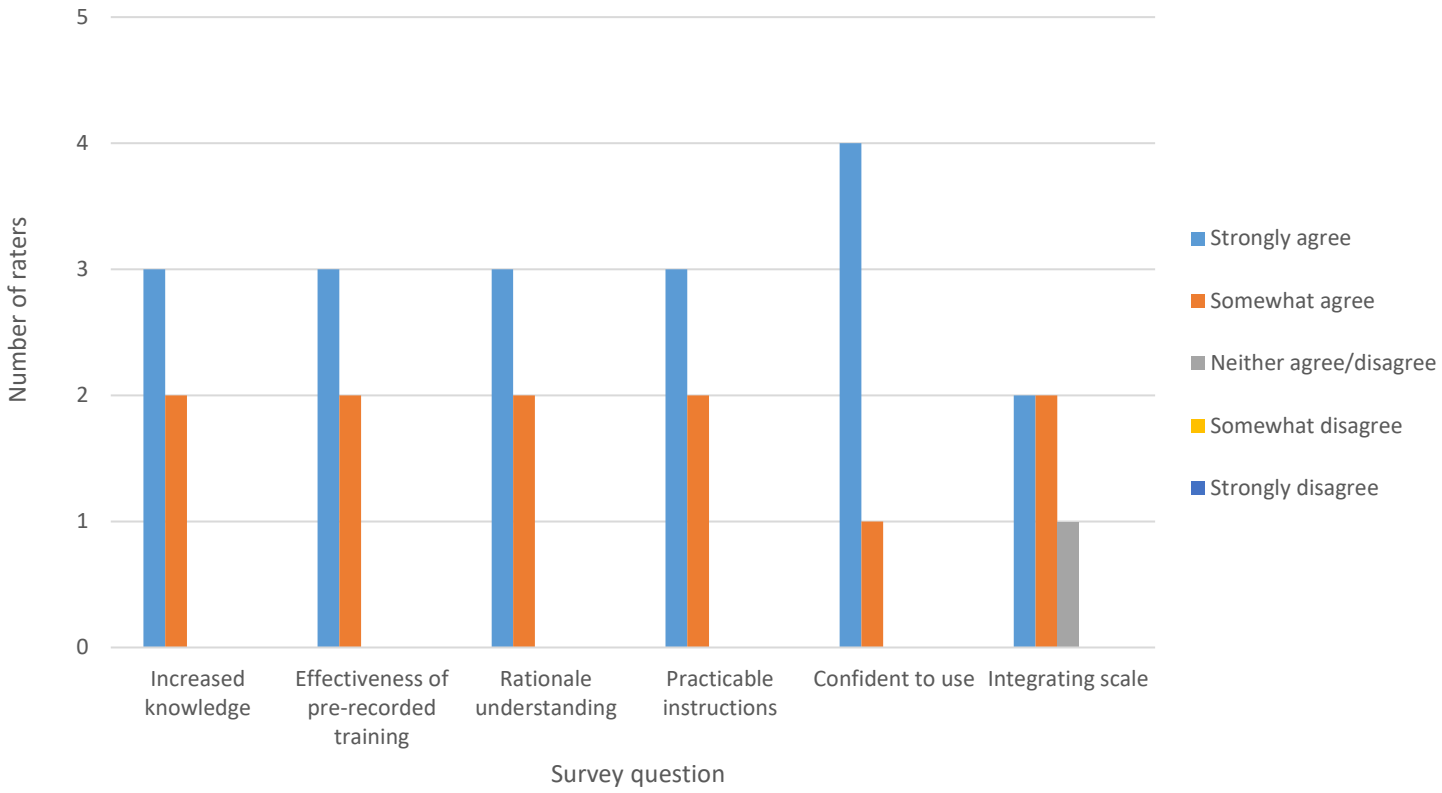
Training Evaluation

To ensure that the training provided raters with an effective understanding of the measure, a six-question training evaluation was created (Appendix G). Each question asked the raters how much they agreed with the statements and they were required to select an option from a 5-point Likert scale, ranging from strongly agree to strongly disagree. Out of the eight raters, five completed the evaluation. Overall, they found the training a positive experience and valued its usefulness. Results are displayed in Graph 1. All five agreed that the training either strongly or somewhat increased their knowledge of what TI, adherence, differentiation and competency are. They also all agreed that the use of pre-recorded training was effective and that the workshop increased their understanding of why a TI measure is useful to implement. All raters who completed the evaluation agreed that the instructions provided regarding the implementation of the scale were practicable and felt confident about using the scale in their clinical practice. Finally, four

agreed that integrating the scale in their clinical practice would be helpful, with one rater selecting a neutral position.

Graph 1

Training Evaluation Rater Responses (N=5)



Construct Validity

The mean measure full-scale therapist GPTIM scores significantly positively correlated with mean patient GPTIM scores, with higher mean measure scores relating to higher mean evaluations ratings (r= .68, p=.03). This construct validity score was considered ‘good’ (Portney and Watkins, 2009). Additionally, despite only being a small sample size and therefore not achieving significance (n=11), specific mean rater measure session scores had a very strong positive correlation with specific mean patient evaluation

session scores (*r*= .98*, p*=.114). According to the cut-offs, this construct validity score is considered ‘excellent’.

Predictive Validity

The therapist GPTIM mean full-scale scores (N=8) significantly positively correlated with mean change improvement scores on both the PHQ9 (*r*= *.68, p*=.03) and GAD7 (*r*= 0.68*, p*=.03). That is, higher mean TI scores were associated with higher mean improvement in PHQ9 and GAD7 scores. Furthermore, the mean measure full-scale GPTIM scores were also found to have significant negative correlation with drop-out rate (*r*= - *.68, p=.03*). To further investigate the GPTIM predictive validity, further testing was done on those patients who dropped out prior to attending all 5 group sessions (*r*= *.68, p=.03*) and those patients who attended less than 50% of sessions (*r*= *.68, p=.03*). Thus, higher mean GPTIM scores were associated with higher dropout rates. In terms of the relationship between average patient GPTIM scores and the dropout rate, this was significantly negatively correlated (*r*= *-1.00, p=.000*). So, the higher the patient GPTIM scores, the lower the drop-out rate. Eight patients rated 11 sessions overall and therefore the sample size was small. Predictive validity findings are displayed in Table 3.

Table 3

Group Psychoeducation TI Measure (GPTIM) Predictive Validity Findings for Drop-Out Definitions and Outcome Scores (n=11)

	Therapist GPTM Score	Patient GPTIM Score	Overall Mean Percentage Rate
Drop-out: Attended ≥ 2 sessions	*r= - .68, p=.03*	*r= -1.00, p=.000*	17.95%

	Therapist GPTM Score	Patient GPTIM Score	Overall Mean Percentage Rate
Drop-out: Attended < 5 sessions	*r= .68, p=.03*	*r=1.00, p=.000*	73.1%
Drop-out: Attended < 50% of sessions	*r= .68, p=.03*	*r=1.00, p=.000*	29.5%
PHQ-9	*r= .68, p=*.03	N/A	N/A
GAD-7	*r= .68, p=*.03	N/A	N/A

Notes. PHQ-9: patient health questionnaire – 9; GAD-7: generalised anxiety disorder – 7; r: pearson correlation; *p:* significance level.

Test-retest Reliability

For the full-scale therapist GPTIM measure, a significant strong positive correlation was found (*r*=.668*, p=.005*). For individual items, a significant correlation (*p*=<0.05) was found for all measure items ranging from .640 (*p*=.008; 'psychoeducational approach') to .847 (*p*=.000; *pacing*). In accordance with the criteria outlined by Cicchetti (1994), seven items demonstrated 'good' test-retest reliability (0.60 to 0.74) and two items exhibited excellent test-retest reliability (≥0.75). Overall, the GPTIM was found to have 'good' test-retest reliability. Findings are displayed in Table 4.

Table 4

Group Psychoeducation TI Measure (GPTIM) Items and Total Test-Retest Pearson Correlation Scores (n=32)

GPTIM Item	Pearson Correlation (*r*)	Significance (*p*)
The facilitators were clearly using a psychoeducational approach	.847***	.000
The psychoeducational information delivered was well matched to the needs of the group	.694***	.003
The session was underpinned by cognitive behavioural theory (and not another theory)	.835***	.000
The facilitators shared and then abided by an agenda	.695***	.003

GPTIM Item	Pearson Correlation (*r*)	Significance (*p*)
The facilitators paced the session appropriately	.640**	.008
The facilitators presented the materials in an engaging and enthusiastic manner	.686**	.003
The facilitators clearly & accurately communicated the psychoeducational information	.748***	.001
The facilitators presented change methods with clarity	.712***	.002
The facilitators provided guidance on the content of between session work ('homework')	.738***	.001
Total score	.668**	.005

Notes. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold.

Inter-rater Reliability

Table 5 details the inter-rater reliability for total therapist GPTIM scores and item scores. Poor inter-rater reliability was found for the full-scale therapist GPTIM (ICC= -.13; 95% CI -2.20 to .70). One of the nine measure items had moderate rating, which was the 'engaging and enthusiastic' item (ICC = .50; 95% CI -.42 to .86). Two items had poor positive inter-rater reliability scores, including the 'clear and accurate communication' (ICC = .35; 95% CI -.85 to .82) and 'change methods' (ICC = .12; 95% -1.51 to .76) items. Seven measure items generated negative values (ICC range -.13 to -.97). A post-hoc inter-rater reliability analysis was conducted to investigate the separate adherence and differentiation and competency GPTIM items. Further testing found that poor inter-rater reliability was found for the total adherence and differentiation GPTIM scores (ICC= -.46; 95% CI -3.15 to .60) and the total competency GPTIM scores (ICC= .16; 95% CI -1.39 to .77). This suggests that consistency between raters was more evident for the competency items in comparison to the adherence items.

Table 5

Inter-rater Reliability for Group Psychoeducation TI Measure (GPTIM) Items and Total (n=32)

GPTIM Item	ICC	95% CI
The facilitators were clearly using a psychoeducational approach	-.97	-4.58 to .46
The psychoeducational information delivered was well matched to the needs of the group	-.38	-2.90 to .63
The session was underpinned by cognitive behavioural theory (and not another theory)	-.65	-3.68 to .55
The facilitators shared and then abided by an agenda	-.44	-3.10 to .61
The facilitators paced the session appropriately	-.81	-4.14 to .51
The facilitators presented the materials in an engaging and enthusiastic manner	.50	-.42 to .86
The facilitators clearly & accurately communicated the psychoeducational information	.35	-.85 to .82
The facilitators presented change methods with clarity	.12	-1.51 to .76
The facilitators provided guidance on the content of between session work ('homework')	-.72	-3.88 to .53
Adherence/Differentiation average (items 1-3)	-.46	-3.15 to .60
Competency average (items 4-9)	.16	-1.39 to .77
Total	-.13	-2.20 to .70

Notes. ICC: inter-class correlation coefficient; CI: confidence interval.

Internal Consistency

The data were initially screened for univariate outliers. As no values were identified as outliers, none of the data were recoded. In all, according to the requirements, the minimum amount of data for the completion of a factor analysis was satisfied (de Winter et al., 2009).

Prior to the completion of a factor analysis, multiple tests were adopted to examine the factorability of the measure items. Initially, it could be observed that all of the items correlated at least .3 with all other items (>.3 = good correlation; Critobal et al., 2007). Following this, the KMO measure of sampling accuracy was calculated as .85, which is above the recommended value of .6 (Samuels, 2017) and is considered 'meritorious' (Dodge, 2008). Bartlett's Test of Sphericity yielded a significant score (X^2 [36]=188.955, p = 0.00). Consequently, no sphericity concerns were found for the measure. In light of these indicators, a factor analysis was deemed appropriate for all items.

To distinguish and calculate composite scores for the factors underlying the therapist GPTIM, a principal components analysis was completed. Eigen values (see Table 6) outlined that the first factor accounted for 61.71% of the variance. The subsequent second, third and fourth factors explained 10.89%, 7.62% and 5.70%, respectively. Remaining factors had eigen values that accounted for between 4.66% and 1.30% of variance. Results led to the selection of a unidimensional factor solution that explained 61.71% of the variance. This was opted for due to the "leveling off" of the scree plot eigen values after the first factor and the insignificance of the second and successive factors (see Appendix I).

None of the nine GPTIM items were disregarded from the measure, as they all contributed towards the primary factor and met the minimum criteria of having a primary factor loading of ≥.4 (see Table 7). All items ranged between .72 and .86 and the loadings appeared to be equal across the adherence, differentiation and competency items. Considerable alpha increases were not observed by eliminating particular items.

Internal consistency for each of the therapist GPTIM items was assessed using Cronbach's alpha (see Table 8). Results demonstrated excellent item alpha values (*range*=.90 to .92). The therapist GPTIM had excellent internal consistency when testing

for the full-scale (α=.92) and further testing revealed that the measure also presented with an excellent Guttman split-half coefficient (*r*SHG=.90).

The descriptive statistics of the item data, including its means, standard deviations, skewness and kurtosis were assessed to examine if it was normally distributed (displayed in Table 9). Most of the data were negatively skewed. In all, one item was deemed symmetrical, six items were moderately skewed and two were highly skewed (Bulmer, 1979). With regards to kurtosis, the data were found to be platykurtic, which translates to a lower and broader central peak with longer tails compared to a normal distribution (Westfall, 2014). Visual examination of histograms depicting the data distribution also confirmed the negatively skewed data. Frequently, negatively skewed data can occur when an upper limit is present that cannot be exceeded (Robertson & Allison, 2012), as is the case for the GPTIM.

Overall, these analysis indicated that one distinct factor was underlying the raters responses to the therapist GPTIM and that this principal factor was strongly internally consistent.

Table 6

Total Variance Explained (n=32)

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.554	61.707	61.707	5.554	61.707	61.707
2	.980	10.894	72.601			
3	.686	7.620	80.22			
4	.513	5.702	85.923			

		Initial Eigenvalues		Extraction Sums of Squared Loadings
5	.419	4.656	90.580	
6	.333	3.698	94.278	
7	.222	2.472	96.750	
8	.176	1.953	98.802	
9	.117	1.298	100.000	

Table 7

Factor Loadings based on an Exploratory Factor Analysis (EFA) for 9 items of the Group Psychoeducation TI Measure (GPTIM) (n=32)

GPTIM Item	Loading
The facilitators were clearly using a psychoeducational approach	.85
The psychoeducational information delivered was well matched to the needs of the group	.76
The session was underpinned by cognitive behavioural theory (and not another theory)	.76
The facilitators shared and then abided by an agenda	,78
The facilitators paced the session appropriately	.74
The facilitators presented the materials in an engaging and enthusiastic manner	.72
The facilitators clearly & accurately communicated the psychoeducational information	.86
The facilitators presented change methods with clarity	.84
The facilitators provided guidance on the content of between session work ('homework')	.73

Table 8

Internal Consistency of the Group Psychoeducation TI Measure (GPTIM) Items (n=32)

GPTIM Item	Item-Total (if deleted)	Cronbach Alpha (if deleted)
The facilitators were clearly using a psychoeducational approach	.80	.91
The psychoeducational information delivered was well matched to the needs of the group	.68	.91
The session was underpinned by cognitive behavioural theory (and not another theory)	.69	.91
The facilitators shared and then abided by an agenda	.72	.91
The facilitators paced the session appropriately	.67	.91
The facilitators presented the materials in an engaging and enthusiastic manner	.65	.92
The facilitators clearly & accurately communicated the psychoeducational information	.82	.90
The facilitators presented change methods with clarity	.79	.90
The facilitators provided guidance on the content of between session work ('homework')	.66	.91

Table 9

Descriptive Statistics for the Group Psychoeducation TI Measure (GPTIM) Items (n=32)

GPTIM Items	*M (SD)*	Skewness	Kurtosis
The facilitators were clearly using a psychoeducational approach	5.36 (.783)	-.762	-.915
The psychoeducational information delivered was well matched to the needs of the group	5.03 (1.015)	-1.209	1.435
The session was underpinned by cognitive behavioural theory (and not another theory)	5.30 (.918)	-1.180	.525
The facilitators shared and then abided by an agenda	5.09 (.843)	-.513	-.521
The facilitators paced the session appropriately	4.67 (.957)	-.623	.605
The facilitators presented the materials in an engaging and enthusiastic manner	4.58 (1.119)	-.629	-.015
The facilitators clearly & accurately communicated the psychoeducational information	4.88 (1.083)	-.689	-.068

The facilitators presented change methods with clarity	4.82 (.983)	-.664	.578
The facilitators provided guidance on the content of between session work ('homework')	4.73 (.839)	-.444	-.101
All items	4.94 (.274)	.369	-1.120

Notes. M: mean; SD: standard deviation.

Summary of Results

Various calculations deemed the therapist GPTIM content validity 'excellent'. The measure had 'good' to 'excellent' construct validity when comparing the GPTIM scores with the patient GPTIM scores. Reasonable predictive validity for outcomes was displayed, but mixed results were found for drop-out dependent upon the definition.

For the reliability measures, 'excellent' internal consistency was demonstrated with a unidimensional factor solution explaining the majority of variance. 'Good' test re-test reliability was evident for the whole measure, with 'good' to 'excellent' scores demonstrated across items. 'Poor' inter-rater reliability was evidenced for most items, the full-scale scores and item sub-groups (adherence/differentiation and competency); although moderate inter-rater reliability was found for one item.

Discussion

The current study has been the first of its kind to develop and then psychometrically evaluate a LI group facilitator TI measure. The development of the 9-item therapist GPTIM and accompanying manual was based on initial Delphi evidence. Primarily, the study aimed to improve the clinical governance of group psychoeducational interventions in IAPT services due to the frequency with which these interventions are delivered for patients with mild-to-moderate anxiety and depression in stepped-care. The psychometric

evaluation of the measure showed that the GPTIM demonstrated ‘excellent’ content validity, ‘good’ to ‘excellent’ construct validity, ‘excellent’ internal consistency, ‘good’ test-retest reliability, and reasonable predictive validity, but had poor inter-rater reliability. The poor inter-rater reliability was indexed through the poor ICCs in the fully crossed design. Faherty et al. (2020) showed in a fully crossed design that the make-up of pairings in terms of personality affected the level of agreement found for the clinical skills being rated. As such factors were not assessed in the current study, this may have accounted for the poor ICCs. It is worth noting that inter-rater reliability is an important component of any measure of TI. The solution to this maybe the testing of composite ratings in future studies (i.e., a rating based on the average score of paired raters of equal training and experience) as this may create improved accuracy. However, in busy clinical services the additional time needed to arrive at a consensus rating may not be warranted or possible. Overall, although these findings do not fully support the immediate use of the GPTIM within clinical settings, the development of this novel, accessible and somewhat psychometrically robust measure, sets the foundation for its further development and use.

A strong component of the measure was its grounding in the behaviour change model (Michie et al., 2011), ensuring that it was theoretically underpinned by the principles of LI-CBT (UCL, 2015). Additionally, comprehensive LI-CBT frameworks underpinned the measure items, further supporting its underlying principles. The measure presented with ‘excellent’ internal consistency, which relates to the stability with which individual raters score the items; indicating the presence of a primary construct. Measures demonstrating internal consistencies of this ilk are likely suitable for use within research trials (Charter, 2008; Nunally & Bernstein, 1994). Although higher levels of internal consistency may indicate item redundancy (where the same item is phrased in different ways; Hulin et al., 2001), these current findings are similar to individual HI and LI-CBT competency measure results (i.e. CTS, CTS-R, CTCS-SP, LIAC and LITC (Blackburn et al., 2001; Ginzburg et

al., 2012; Kazantzis et al., 2018; Kellett et al., 2021). Higher internal consistencies within such measures may be indicative of the overlap present between items, due to the unifying quality of CBT frameworks (Muse et al., 2017).

The ICC scores and their highly variable CI indicated poor levels of inter-rater reliability, which is symptomatic of inconsistent agreement between raters (Bobak et al., 2018). A large number of the ICC scores were negative values, which can occur when there is very low between-subject variation (where variability within the group is greater than across groups; Shrout & Fleiss, 1979). Such negative ICC scores should be interpreted with caution (Bartko, 1976; Wu et al., 2012). Bartko (1976) suggested that if one is averaging the ICC scores and it is suspected that negative values are due to sampling error, negative scores can be reset to '0'. In this instance, following this process results in a mean ICC of 0.11, which remains within the 'poor' range. It is reasonable to suggest that sampling errors may have been present, due to the small number of subjects, small number of raters and/or a lack of subject variability (Portney & Watkins, 2009; Lee et al., 2012). Although the rater sample pool was selected based upon guidelines (Koo & Lee, 2016), the study appeared to lack session variability of integrity (which would likely affect ICC ratings), as the mean GPTIM score of all sessions was fairly high (*M=44.7/54, SD=6.25, range=27-54/54).* Another consideration was the fact that the measure required raters to score the group facilitators as a pair, rather than as individuals. Current TI and competency measures that are regularly used all rate individual practitioners. Thus, rating facilitator pairs would have been unfamiliar for raters, irrespective of their experience levels and may have introduced variability.

Another area that may be partially responsible for the ICC results is the delivery of brief and remote rater training. Importantly, training was remote due to the COVID-19 pandemic. Although, despite the study aiming to provide rater training that would result in the measure being pragmatic for services (i.e. able to be completed remotely, at a time to

suit each individual clinician without affecting clinic days), similar research has provided training anywhere from half to a full-day (Gordon, 2006; McCay et al., 2016; Muse et al., 2017; Pierson et al., 2007). Therefore, although less intensive rater training is feasible (Kellett et al., 2021), additional training may have resulted in higher inter-rater reliability. In line with this, a higher ICC scofre was observed for the 'engaging and enthusiastic item' (moderate), which is an element that requires little to no rater experience or training to identify. The training was also pre-recorded and raters self-reported their completion of the workshop with no adherence checks. Educational literature notes the negative impact of pre-recorded lectures on attainment (Le, 2022) and demonstrates that learning is heavily dependent upon motivation levels for this format (Islam et al., 2020). In addition, the self-reported completion of training may be unreliable, whilst remote training eliminates the opportunity for raters to agree on a joint consensus regarding sample ratings (McCay et al., 2016). Resultantly, it may have been preferential to offer rater training in a face-to-face format (similarly to the evaluation of most measures) or via a live online workshop. As the effective training of raters using a TI measure is pivotal (Perpletchikova, 2011), the above considerations should be heavily considered when developing the GPTIM in the future.

The measure demonstrated predictive validity, with higher TI ratings strongly correlating with improvements in patients' GAD-7 and PHQ-9 scores. These findings are meaningful, as they support the conclusions of Power et al. (2022) associating TI with outcomes. Historically, the literature has presented conflicting findings and results can differ due to a number of factors (i.e. patient characteristics, treatment, therapists etc; Perpletchikova & Kazdin, 2005). Mixed predictive validity results were found for the association between dropout and TI, which was dependent on dropout definition. Due to this, it would be misguided to make any conclusions regarding the measures' ability to forecast drop-out outcomes. In the wider literature, common predictors of patient drop-out, such as a poor alliance, dissatisfaction with therapeutic change and distrust (Barratt et al.,

2008; Swift & Greenberg, 2014), could be heavily related to the facilitators TI. As a result, although conclusive findings were not discovered, the relationship with service engagement and clinical outcome should be further investigation.

Limitations

All rated sessions were gathered from one service and it would have been judicious to have expanded this across various services for generalisation purposes. The sample size (raters, patients and group treatment sessions) was also relatively small, which impacted the strength of statistical tests and results. For example, low rater numbers can result in lower inter-rater reliability scores, which is indicative of the tendency for ICCs to be inflated with more raters (Karterud et al., 2013). The construct validity analysis was also impacted by low sample size, as it was reliant upon patient GPTIM completion. Overall, these were completed by a limited number of patients and were resultantly unrepresentative of the whole sample. Importantly, this appears to suggest that the level of burden placed upon patients by asking them to complete an additional measure was not feasible. When creating the patient GPTIM, despite carefully attempting to parallel the expert GPTIM, a CVI process was not conducted. Patients were also absent from the patient GPTIM development process, which could have negatively impacted the included items and their wording. Consequently, one should interpret results related to this measure with caution. Another limitation was related to the pool of raters, as this was chosen with an arbitrary cut-off regarding group facilitation experience. This meant that experience levels, past the minimum standard, were not controlled for. It should also be highlighted that the findings and conclusions are limited to LI-CBT *Stress Control* group interventions (White, 1990) that are delivered to people with mild-to-moderate anxiety and/or depression. Therefore, any generalisations outside of this format or population may be

invalid. Finally, as patients are likely unaware of the correct content for group psychoeducation sessions, it may have been an oversight to request that they score certain items (some of which would rely on a basic understanding of facilitator expectations) that contribute to an overall TI score. However, the inclusion of this measure was warranted, as patients could rate the majority of items relating to TI accurately, without needing to be trained or briefed on what TI is.

Clinical & Research Implications

In relation to the clinical implications, this study has established that the therapist GPTIM is a measure with reasonable psychometric properties. Thus, within the real world, this means that services are closer to measuring the TI of facilitators delivering group psychoeducation (a concept closely linked to patient outcome (Power et al., 2022)). Despite this, due to the requirement for further research to fine-tune the GPTIM (due to the inter-rater reliability results), utilisation of the measure would not be indicated at this point. Services should therefore opt to continue using other reliable and valid measures prior to the full validation of the GPTIM.

Future studies could alter the raters training, which could involve an increase in training time (as shown in Kühne et al., 2020 and Perpletchikova et al., 2007) and/or changes in training formats (i.e. purely face to face, live video or a mix of the two). Further recommendations include refinement of the GPTIM integrity rating descriptions (Fairburn & Cooper, 2011) and additional clarity regarding the joint rating of facilitators. In addition, due to the didactic nature of psychoeducation sessions, the measure could be further developed using simulated psychoeducational sessions, which may be less labour intensive and affordable than real-life examples (Gizburg et al., 2012). Finally, raters ranging in experience should complete the measure, as the reliability of competency

ratings has been found to be influenced by rater experience (Muse & McManus, 2013). This may result in practical suggestions regarding the measure (i.e., should a rating disparity be observed, novice raters needing support from experienced practitioners; Muse et al., 2017; Kellett et al., 2021).

Conclusions

Despite the evident popularity of LI psychoeducation groups within primary mental health care, services have not previously had a measure to rate the TI of facilitators. Resultantly, the current study developed and evaluated a novel LI group psychoeducation TI measure (GPTIM), and the measure was demonstrated to have an acceptable range of reliability and validity indices. However, the GPTIM had poor inter-rater reliability, possibly due to rater training, methodological issues and a limited sample. To fine-tune the measure, future research should look to improve the training of raters, adapt the GPTIM manual and increase the number of rated sessions. Once the measure is psychometrically sound, it could then add to the clinical governance of LI psychoeducational groups and contribute research trial methods.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Arifin, W. A. (2018). A web-sample size calculator for reliability studies. *Education in Medical Journal, 10*(3). https://doi.org/10.21315/eimj2018.10.3.8

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x

Bartko, J.J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin, 83*(5), 762-765.

Bastiampillai, T. J., Allison, S., Harford, P., Perry, S. W., & Wong, M. L. (2019). Has the UK Improving Access to Psychological Therapies programme and rising antidepressant use had a public health impact?. *The Lancet Psychiatry*, *6*(3), e8-e9. https://doi.org/10.1016/S2215-0366(19)30040-9.

Beard, C., Hsu, K. J., Rifkin, L. S., Busch, A. B., & Björgvinsson, T. (2016). Validation of the PHQ-9 in a psychiatric sample. *Journal of Affective Disorders*, *193*, 267-273. https://doi.org/10.1016/j.jad.2015.12.075

Bennett-Levy, J., Richards, D. A., & Farrand, P. (2010). Low intensity CBT interventions: a revolution in mental health care. In Bennett-Levy, J., Richards, D. A., Farrand, P., Christensen, H., Griffiths, K. M., Kavanagh, D. J., & Proudfoot, J. (Eds.), *Oxford Guide to Low Intensity CBT Interventions*, (pp. 3-18). Oxford University Press. doi.org/10.1093/med:psych/9780199590117.003.0001

Bobak, C., Barr, P. & O'Malley, A. (2018). Estimation of an inter-rater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement scales. *BMC Medical Research Methodology, 18*, 93. https://doi.org/10.1186/s12874-018-0550-6

Boyd, L., Baker, E., & Reilly, J. (2019). Impact of a progressive stepped care approach in an improving access to psychological therapies service: An observational study. *PLOS One*, *14*(4), e0214715. https://doi.org/10.1371/journal.pone.0214715

Boyle, K., Deisenhofer, A. K., Rubel, J. A., Bennemann, B., Weinmann-Lutz, B., & Lutz, W. (2020). Assessing TI in personalized CBT: The inventory of therapeutic interventions and skills. *Cognitive Behaviour Therapy*, *49*(3), 210-227. https://doi.org/10.1080/16506073.2019.1625945

Brown, N. W. (2018). *Conducting effective and productive psychoeducational and therapy groups: A guide for beginning group leaders*. Taylor & Francis.

Burns, P., Kellett, S., & Donohoe, G. (2016). “Stress Control” as a large group psychoeducational intervention at Step 2 of IAPT services: Acceptability of the approach and moderators of effectiveness. *Behavioural and Cognitive Psychotherapy*, *44*(4), 431-443. https://doi.org/10.1017/S1352465815000491

Bulmer, M. G. (1979). *Principles of statistics*. Courier Corporation.

Cameron, I. M., Crawford, J. R., Lawton, K., & Reid, I. C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice*, *58*(546), 32-36. https://doi.org/10.3399/bjgp08X263794

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284. https://doi.org/10.1037/1040-3590.6.4.284

Charter, R. A. (2008). Statistical approaches to achieving sufficiently high test score reliabilities for research purposes. *The Journal of General Psychology*, *135*(3), 241-251. https://doi.org/10.3200/GENP.135.3.241-251

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98. https://doi.org/10.1037/0021-9010.78.1.98

Clark, D., Layard, R., Smithies, R., Richards, D. R., Suckling, R., & Wright, W. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy, 47*, 910–920 https://doi.org/10.1016/j.brat.2009.07.010

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International Review of Psychiatry*, *23*(4), 318-327. https://doi.org/10.3109/09540261.2011.606803

Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet*, *391*(10121), 679-686. https://doi.org/10.1016/S0140-6736(17)32133-5

Cristobal, E., Flavian, C., & Guinaliu, M. (2007). Perceived e-service quality: Measurement validity and effects on consumer satisfaction and web site loyalty. *Managing Service Quality*, *17*, 317–340. https://doi.org/10.1108/ 09604520710744326

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

Cuijpers, P., Muñoz, R. F., Clarke, G. N., & Lewinsohn, P. M. (2009). Psychoeducational treatment and prevention of depression: the “Coping with Depression” course thirty years later. *Clinical Psychology Review*, *29*(5), 449-458. https://doi.org/10.1016/j.cpr.2009.04.005

de Winter, J. C., Dodou, D. I. M. I. T. R. A., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147-181. https://doi.org/10.1080/00273170902794206

Dinger, U., Zilcha-Mano, S., Dillon, J., & Barber, J. P. (2014). Therapist adherence and competence in psychotherapy research. *The Encyclopedia of Clinical Psychology*, 1-5. https://doi.org/10.1002/9781118625392.wbecp340

Dodge, Y. (2008). *The concise encyclopedia of statistics*. Springer Science & Business Media.

Dreyfus, H. L. (1989). The Dreyfus Model of Skill Acquisition. In J. Burke (Eds.), *Competency based education and training.* Falmer Press

Faherty A, Counihan T, Kropmans T, & Finn Y. (2020). Inter-rater reliability in clinical assessments: do examiner pairings influence candidate ratings? *BMC Medical Education*, 20(1), 1-9. https://doi.org/10.1186/s12909-020-02009-4

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Firth, N., Barkham, M., & Kellett, S. (2015). The clinical effectiveness of stepped care systems for depression in working age adults: a systematic review. *Journal of Affective Disorders*, *170*, 119-130. https://doi.org/10.1016/j.jad.2014.08.030

Ginzburg, D. M., Bohn, C., Höfling, 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

Gordon, P. K. (2006). A comparison of two versions of the Cognitive Therapy Scale. *Behavioural and Cognitive Psychotherapy*, *35*(3), 343-353. https://doi.org/10.1017/S1352465806003390

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy*, *51*(9), 597-606. https://doi.org/10.1016/j.brat.2013.06.004

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23. https://doi.org/10.20982/tqmp.08.1.p023

Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement*, *50*(4), 717-729. https://doi.org/10.1177/0013164490504001

Hulin, C., Netemeyer, R., & Cudeck, R. (2001). Can a Reliability Coefficient Be Too High. *Journal of Consumer Psychology*, *10*, 55-58. https://doi.org/10.1207/S15327663JCP1001&2_05

IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.

Islam, M., Kim, D. A., & Kwon, M. (2020). A comparison of two forms of instruction: Pre-recorded video lectures vs. live ZOOM lectures for education in the business management field. *Sustainability*, *12*(19), 8149. https://doi.org/10.3390/su12198149

Johnson, S. U., Ulvenes, P. G., Øktedalen, T., & Hoffart, A. (2019). Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *Frontiers in Psychology, 10*, 1713. https://doi.org/10.3389/fpsyg.2019.01713

Karterud, S., Pedersen, G., Engen, M., Johansen, M. S., Johansson, P. N., Schlüter, C., ... & Bateman, A. W. (2013). The MBT Adherence and Competence Scale (MBT-ACS): development, structure and reliability. *Psychotherapy Research*, *23*(6), 705-717. https://doi.org/10.1080/10503307.2012.708795

Kazantzis, N., Clayton, X., Cronin, T., Farchione, D., Limburg, K., & Dobson, K. (2018). The cognitive therapy scale and cognitive therapy scale-revised as measures of therapist competence in cognitive behavior therapy for depression: Relations with short and long term outcome. *Cognitive Therapy and Research*, *42*(4), 385-397. https://doi.org/10.1007/s10608-018-9919-4

Kellett, S., Simmonds-Buckley, M., Limon, E., Hague, J., Hughes, L., Stride, C., & Millings, A. (2021). Defining the assessment and treatment competencies to deliver low-intensity cognitive behavior therapy: A multi-center validation study. *Behavior Therapy*, *52*(1), 15-27. https://doi.org/10.1016/j.beth.2020.01.006

Kendall, P. C. (1998). Empirically supported psychological therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 3. https://doi.org/10.1037/0022-006X.66.1.3

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kroenke, K., Spitzer, R. L., & Williams, J. B. (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

Kühne, F., Heinze, P. & Weck, F. (2020). Standardized patients in psychotherapy training and clinical supervision: study protocol for a randomized controlled trial. *Trials, 21***,** 276. https://doi.org/10.1186/s13063-020-4172-z

Landsverk, J. (2013). Reflections on TI in a dissemination and implementation framework. *Clinical Psychology: Science and Practice*, *20*(1), 114. https://doi.org/10.1111/cpsp.12027

Le, K. (2022). Pre-recorded lectures, live online lectures, and student academic achievement. *Sustainability*, *14*(5), 2910. https://doi.org/10.3390/su14052910

Lee, K. M., Lee, J., Chung, C. Y., Ahn, S., Sung, K. H., Kim, T. W., ... & Park, M. S. (2012). Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. *Clinics in Orthopedic Surgery*, *4*(2), 149-155. https://doi.org/10.4055/cios.2012.4.2.149

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, *46*(3) 266-274. https://doi.org/10.1097/MLR.0b013e318160d093

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

McCay, E., Carter, C., Aiello, A., Quesnel, S., Howes, C., & Johansson, B. (2016). Toward TI: developing an approach to measure the TI of a dialectical behavior therapy intervention with homeless youth in the community. *Archives of Psychiatric Nursing*, *30*(5), 568-574. https://doi.org/10.1016/j.apnu.2016.04.001

Meadows, J., & Kellett, S. (2017). Development and evaluation of Cognitive Analytic Guided Self-Help (CAT-SH) for use in IAPT services. *Behavioural and Cognitive Psychotherapy*, *45*(3), 266-284. https://doi.org/10.1017/S1352465816000485

Michie, S., Ashford, S., Sniehotta, F., Dombrowski, S., Bishop, A., & French, D. (2011). A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. *Psychology & Health, 26*(11), 1479-1498. https://doi.org/10.1080/08870446.2010.540664

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

Muse, K., McManus, F., Rakovshik, S., & Thwaites, R. (2017). Development and psychometric evaluation of the Assessment of Core CBT Skills (ACCS): An observation-based tool for assessing cognitive behavioral therapy competence. *Psychological Assessment*, *29*(5), 542-555. https://doi.org/10.1037/pas0000372

National Collaborating Centre for Mental Health (2018). *The improving access to psychological therapies manual: appendices and helpful resources.* Retrieved from: england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual-v5.pdf

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Orlinsky, D. E., & Howard, K. I. (1986). The Psychological Interior of Psychotherapy: Explorations with the Therapy Session Reports. In L. S. Greenberg & W. M. Pinsof (Eds.), *Guilford clinical psychology and psychotherapy series. The psychotherapeutic process: A research handbook* (p. 477–501). Guilford Press.

Pierson, H. M., Hayes, S. C., Gifford, E. V., Roget, N., Padilla, M., Bissett, R., & Fisher, G. (2007). An examination of the motivational interviewing treatment integrity code. *Journal of Substance Abuse Treatment*, *32*(1), 11-17. https://doi.org/10.1016/j.jsat.2006.07.001

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice*, *18*(2), 148-153. https://doi.org/10.1111/j.1468-2850.2011.01246.x

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, *12*(4), 365. https://doi.org/10.1093/clipsy.bpi045

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*(6), 829. https://doi.org/10.1037/0022-006X.75.6.829

Petkov, J., Harvey, P., & Battersby, M. (2010). The internal consistency and construct validity of the partners in health scale: validation of a patient rated chronic condition self-management measure. *Quality of Life Research*, *19*(7), 1079-1085. https://doi.org/10.1007/s11136-010-9661-1

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: applications to practice* (Vol. 892). Upper Saddle River, NJ: Pearson/Prentice Hall.

Power, N., Noble, L., Simmonds-Buckley, M., Kellett, S., Stockton, C., Firth, N., & Delgadillo, J. (2022). Associations between treatment adherence-competence-integrity (ACI) and adult psychotherapy outcomes: a systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*.

Richards, D., & Whyte, M. (2009). *Reach Out: National programme student materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions*. Rethink Mental Illness.

Robertson, H. T., & Allison, D. B. (2012). A novel generalized normal distribution for human longevity and other negatively skewed data. *PLOS One*, *7*(5), e37025. https://doi.org/10.1371/journal.pone.0037025

Roth, A., & Pilling, S. (2007). *The Competences Required to Deliver Effective Cognitive and Behavioural Therapy for People with Depression and with Anxiety Disorder*. Department of Health.

Samuels, P. (2017) *Advice on Exploratory Factor Analysis.* Technical Report. Research Gate.

Shafran, R., Myles-Hooton, P., Bennett, S., & Öst, L. G. (2021). The concept and definition of 'low intensity'cognitive behaviour therapy. *Behaviour Research and Therapy*, 103803. https://doi.org/10.1016/j.brat.2021.103803

Schoenwald, S. K., Mehta, T. G., Frazier, S. L., & Shernoff, E. S. (2013). Clinical supervision in effectiveness and implementation research. *Clinical Psychology: Science and Practice*, *20*(1), 44. https://doi.org/10.1111/cpsp.12022

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

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420. https://doi.org/10.1037/0033-2909.86.2.420

Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying TI research for dissemination and implementation science: Introduction to special issue. *Clinical Psychology: Science and Practice*, *20*(1), 1. https://doi.org/10.1111/cpsp.12019

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Swift, J. K., & Greenberg, R. P. (2014). A treatment by disorder meta-analysis of dropout from psychotherapy. *Journal of Psychotherapy Integration*, 24(3), 193. https://doi.org/10.1037/a0037512

Toussaint, A., Hüsing, P., Gumz, A., Wingenfeld, K., Härter, M., Schramm, E., & Löwe, B. (2020). Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). *Journal of Affective Disorders*, *265*, 395-401. https://doi.org/10.1016/j.jad.2020.01.032

University College London (2015). National Curriculum for the Education of Psychological Wellbeing Practitioners. Department of Health.

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's alpha. *Leisure Sciences*, *39*(2), 163-173. https://doi.org/10.1080/01490400.2015.1127189

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (2021). Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *British Journal of Clinical Psychology*, *60*(1), 1-37. https://doi.org/10.1111/bjc.12259

Walsh, W. B., & Betz, N. E. (2001). *Tests and assessment (4th ed.).* Prentice Hall.

Waltz, C. F., & Bausell, R. B. (1981). *Nursing research: Design, statistics, and computer analysis*. FA Davis Company.

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*, *78*(2), 200. https://doi.org/10.1037/a0018912

F. Weck., Kaufmann, Y. M., & Höfling, V. (2017) Competence feedback improves CBT competence in trainee therapists: A randomized controlled pilot study. *Psychotherapy Research*, *27*(4), 501-509. https://doi.org/10.1080/10503307.2015.1132857

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*(4), 620. https://doi.org/10.1037/0022-006X.61.4.620

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*, *78*(2), 200. https://doi.org/10.1037/a0018912

Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *The Journal of Nervous and Mental Disease*, *199*(4), 276-279. https://doi.org/10.1097/NMD.0b013e3182124617

White, J., & Keenan, M. (1990). Stress control: A pilot study of large group therapy for generalized anxiety disorder. *Behavioural and Cognitive Psychotherapy*, *18*(2), 143 146. https://doi.org/10.1017/S0141347300018267

Wu, S., Crespi, C. M., & Wong, W. K. (2012). Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. *Contemporary Clinical Trials, 33*(5), 869-880. https://doi.org/10.1016/j.cct.2012.05.004

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, *49*(2), 156–167. https://doi.org/10.1037//0022-006X.49.2.156

Appendix A

Ethical Approval Documents

Dr Stephen Kellett

Clinical Psychology Unit,

Department of Psychology

Email: approvals@hra.nhs.uk

University of Sheffield, Cathedral Court
HCRW.approvals@wales.nhs.uk

1 Vicar Lane, Sheffield

S1 2LTN/A

08 June 2021

Dear Dr Kellett

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

Study title:	**Evaluating a measure of facilitator competency for the delivery of group psychoeducation for anxiety and depression in the Improving Access to Psychological Therapies Service (IAPT)**
IRAS project ID:	**293765**
Protocol number:	**N/A**
REC reference:	**21/YH/0058**
Sponsor	**The 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 standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, 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 **293765**. Please quote this on all correspondence.

Yours sincerely,

Anna Bannister

Approvals Specialist

Email: approvals@hra.nhs.uk

Copy to: *Dr Stephen Kellett* **List of Documents**

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

Document	*Version*	*Date*
Cover Letter [Cover Letter Response Table]		
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)		19 February 2021
IRAS Application Form [IRAS_Form_22022021]		22 February 2021
IRAS Application Form XML file [IRAS_Form_22022021]		22 February 2021
IRAS Checklist XML [Checklist_22022021]		22 February 2021
Organisation Information Document		
Participant consent form [Participant Consent Form]	2	30 November 2020
Participant consent form [Participant Consent Form]	2	30 November 2020
Participant information sheet (PIS) [Participant Information Sheet Vol 4]	4	28 May 2021
Participant information sheet (PIS) [Practitioner Participant Information Sheet Vol 4]	4	28 May 2021
Research protocol or project proposal [Research Proposal]	2	30 November 2020
Schedule of Events or SoECAT [IRAS Schedule of Events]	2	28 May 2021
Summary CV for Chief Investigator (CI) [Dr Stephen Kellett's CV]		
Summary CV for student [Student/Researcher]	1	02 February 2021
Summary CV for supervisor (student research) [Supervisor]		17 February 2021
Validated questionnaire [IAPT Questionnaire Minimum Data Set]		

IRAS project ID	293765

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
There is only one participating NHS organisation therefore there is only one site type.	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 and the sponsor is not requesting and does not expect any other site agreement to be used.	No application for external funding will be made.	The Chief Investigator will be responsible for all research activities performed at study sites.	Use of identifiable patient records held by an NHS organisation to identify potential participants should be undertaken by a member of the direct care team for the patient, so it would not normally be acceptable for this to be done by staff not employed by that organisation. A Letter of Access (or equivalent) would be expected for any external NHS/research staff undertaking all of the other activities for the study once consent from the participant is in place. The preengagement checks should include a standard DBS check and Occupational Health Clearance.

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 they do not intend to apply for inclusion on the NIHR CRN Portfolio.

NHS
Health Research Authority

Yorkshire & The Humber - South Yorkshire Research Ethics Committee
NHSBT Newcastle Blood Donor Centre
Holland Drive
Newcastle upon Tyne
NE2 4NQ

Telephone: 0207 1048091

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

11 May 2021

Dr Stephen Kellett

Clinical Psychology Unit, Department of Psychology

University of Sheffield, Cathedral Court

1 Vicar Lane, Sheffield

S1 2LT

Dear Dr Kellett

Study title:	**Evaluating a measure of facilitator competency for the delivery of group psychoeducation for anxiety and depression in the Improving Access to Psychological Therapies Service (IAPT)**
REC reference:	**21/YH/0058**
Protocol number:	**N/A**
IRAS project ID:	**293765**

Thank you for your letter of 27 April 2021, responding to the Research Ethics Committee's (REC) request for further information on the above research [and submitting revised documentation].

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a **favourable ethical opinion** for the above research on the basis described in the application form, protocol and supporting documentation [as revised], subject to the conditions specified below.

Good practice principles and responsibilities

The UK Policy Framework for Health and Social Care Research sets out principles of good practice in the management and conduct of health and social care research. It also outlines the responsibilities of individuals and organisations, including those related to the four elements of research transparency:

1. registering research studies
2. reporting results
3. informing participants
4. sharing study data and tissue

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials

All research should be registered in a publicly accessible database and we expect all researchers, research sponsors and others to meet this fundamental best practice standard.

It is a condition of the REC favourable opinion that **all clinical trials are registered** on a publicly accessible database within six weeks of recruiting the first research participant. For this purpose, 'clinical trials' are defined as the first four project categories in IRAS project filter question 2. Failure to register a clinical trial is a breach of these approval conditions, unless a deferral has been agreed by or on behalf of the Research Ethics Committee (see here for more information on requesting a deferral:

https://www.hra.nhs.uk/planning-and-improving-research/research-planning/research-registratio n-research-project-identifiers/

If you have not already included registration details in your IRAS application form, you should notify the REC of the registration details as soon as possible.

Further guidance on registration is available at:
https://www.hra.nhs.uk/planning-and-improving-research/research-planning/transparency-respo nsibilities/

Publication of Your Research Summary

We will publish your research summary for the above study on the research summaries section of our website, together with your contact details, no earlier than three months from the date of this favourable opinion letter.

Should you wish to provide a substitute contact point, make a request to defer, or require further information, please visit:

https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-sum maries/

N.B. If your study is related to COVID-19 we will aim to publish your research summary within 3 days rather than three months.

During this public health emergency, it is vital that everyone can promptly identify all relevant research related to COVID-19 that is taking place globally. If you haven't already done so, please register your study on a public registry as soon as possible and provide the REC with the registration detail, which will be posted alongside other information relating to your project. We are also asking sponsors not to request deferral of publication of research summary for any projects relating to COVID-19. In addition, to facilitate finding and extracting studies related to COVID-19 from public databases, please enter the WHO official acronym for the coronavirus disease (COVID-19) in the full title of your study. Approved COVID-19 studies can be found at:

https://www.hra.nhs.uk/covid-19-research/approved-covid-19-research/

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document “After ethical review – guidance for researchers” gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report
- Reporting results

The latest guidance on these topics can be found at https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/.

Ethical review of research sites

NHS/HSC sites

The favourable opinion applies to all NHS/HSC sites taking part in the study, subject to confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or management permission (in Scotland) being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS/HSC sites

I am pleased to confirm that the favourable opinion applies to any non-NHS/HSC sites listed in the application, subject to site management permission being obtained prior to the start of the study at the site.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Cover Letter [Cover Letter Response Table]		
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)		19 February 2021
IRAS Application Form [IRAS_Form_22022021]		22 February 2021
Participant consent form [Participant Consent Form]	2	30 November 2020
Participant consent form [Participant Consent Form]	2	30 November 2020
Participant information sheet (PIS) [Practitioner Participant Information Sheet]	3	15 March 2021
Participant information sheet (PIS) [Participant Information Sheet]	3	15 March 2021
Research protocol or project proposal [Research Proposal]	2	30 November 2020
Summary CV for Chief Investigator (CI) [Dr Stephen Kellett's CV]		
Summary CV for student [Student/Researcher]	1	02 February 2021
Summary CV for supervisor (student research) [Supervisor]		17 February 2021
Validated questionnaire [IAPT Questionnaire Minimum Data Set]		

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:

http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at:

https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS project ID: 293765 Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely

pp

Dr Max Huxham Chair

Email: southyorks.rec@hra.nhs.uk

Enclosures: "After ethical review – guidance for researchers"

Copy to: Dr Stephen Kellett

Lead Nation England: approvals@hra.nhs.uk

Appendix B

Patient and Rater Participant Documents

PARTICIPANT INFORMATION SHEET

IRAS Project ID: 293765

Project Title: Developing and evaluating a measure of treatment integrity for the delivery of group psychoeducation for anxiety and depression

You are being invited to participate in a doctoral research project that is being completed by Jonah Gosling, Postgraduate Researcher and Trainee Clinical Psychologist, from the University of Sheffield. Before deciding if you would like to be involved, it is important you understand why the research is being done and what it will involve. Please take the time to read the following information and if anything is unclear, please ask questions and we (researcher and the study research team) will provide you with further information.

Purpose of the research

Group sessions are often used by Improving Access to Psychological Therapies (IAPT) services as they are known to reduce common mental health difficulties. However, a rating tool to measure the ability of the group facilitators does not exist. Therefore, this project aims to develop and evaluate a rating tool to assess group facilitator ability. The measure will aim to check facilitators are meeting expected standards. Also, the measure aims to develop IAPT practitioner training and improve supervision.

What would taking part involve?

If you agree to take part, it will not change the standard of care you receive. All individuals attending groups will receive the same treatment and will complete the same weekly service self-report measures. If you agree to take part in the study, you will complete a further facilitator rating scale on a weekly basis via an online link. By agreeing to take part, you consent to us using information from your medical records for the research project. This information will include your initials, NHS number and contact details. People will use this information to do the research or to check your records to make sure that the research is being done properly. Furthermore, you will be consenting to the sessions being recorded so facilitators can be rated by project collaborators. Throughout the duration of the study, all research data (including session recordings) will be stored on the secure University storage system, which will only be accessible by the researchers involved. All information collected will remain confidential and the data will be annoymised. If you choose not to take part, your information will not be passed to the research team.

Do I have to take part?

Participation is voluntary. If you decide to take part, you will be given this information sheet to keep and will need to fill in a consent form and demographic information sheet. You can withdraw your information at any time without it affecting your treatment.

What are the possible benefits of taking part?

Regardless of whether you take part or not, by attending group sessions you will receive evidenced-based treatment. By allowing your data to be included in the study, it will enable us to

develop a measure that will contribute to the improvement and further investigations of groups and their facilitators, which could help future service users.

What are the possible risks and disadvantages of taking part?
There will be no change to the treatment or care you receive, so there are no perceived risks or disadvantages of taking part in the research. All data and recordings will be kept confidential and will be analysed anonymously.

Supporting information

How will my information be kept confidential?
The researcher will be responsible for the data and will only use information that is needed for the research study. We will let very few people know your name or contact details, and only if they really need it for this study. Everyone involved in this study will keep your data safe and secure. We will also follow all privacy rules. All data will be stored securely on an electronic password protected folder only accessible by the researcher. To anonymise the data, participant identification numbers will be used instead of names. The data will also be analysed collectively so you will not be identifiable within any reports or publications. Any personal data will be securely stored for 3 years before it is destroyed. Also, anonymised research data will be stored on the secure University repository for 10 years.

What if I decide I don't want to be involved in the study anymore?
If you decide you do not want to be involved anymore, you can request that your data not be passed to the research team by informing the lead facilitator or the researcher (details below). Any data that has been collected previously and passed to the research team will then be removed and destroyed.

What if I need extra support in a session?
The service policy means that you will be able to make the course facilitator aware of your need for extra support and they will provide this if required. Furthermore, should you disclose the need for additional wellbeing support, your named practitioner will be able to contact you and provide this.

What if I have any concerns, want to complain or something goes wrong?
If you have any concerns regarding the research, we ask that you contact either the named researcher or IAPT lead (details outlined below). Complaints regarding the research should be forwarded to Amrit Sinha, Research Support and Data Protection Officer, the University of Sheffield (contactable at a.sinha@sheffield.ac.uk or 0114 222 6650). In the event that something does go wrong and you are harmed in the study, public liability insurance is available via the University of Sheffield. In these circumstances you may also have grounds for legal action against Sheffield Health & Social Care NHS Foundation Trust, but you may have to pay your legal costs.

What if I don't complete one or more of the rating scales?
If you fail to fully complete the rating scale via the online link, a reminder will be sent. However, should you fail to complete the rating scale thereafter, you will be withdrawn from the project.

Where can I find out more about how my information is used?
You can find out more about how your information will be used at https://www.hra.nhs.uk/information-about-patients/. Should you require any further details regarding how your information will be used, please contact the researcher (details below) or Amrit Sinha, Research Support and Data Protection Officer, the University of Sheffield (contactable at a.sinha@sheffield.ac.uk or 0114 222 6650).

What will happen to the results of this study?
The results will make up part of the researchers' Clinical Psychology Doctorate thesis and will be reported back to the service. If you would like to know more about the outcome of the study, please contact the researcher using the details below to receive outcome updates.

Ethical consent for this study has been obtained from the NHS Research Ethics Committee.
If you would like any further information about the study, please contact the researcher or IAPT lead using the contact details below.

Researcher
Name: Jonah Gosling
Address: Psychology Department
University of Sheffield
Cathedral Court, 1 Vicar Lane
Sheffield, S1 2LT
Email: jgosling3@sheffield.ac.uk

IAPT Lead
Name: Jodie Millington
Address: Sheffield IAPT
St George's Health Centre
Winter Street, Sheffield,
S3 7ND
Telephone: 0114 226 4380

Version 3.0: 15th March 2021

PARTICIPANT/PATIENT CONSENT FORM

IRAS Project ID: 293765

Participant ID Number:

Title of Project: Developing and evaluating a measure of treatment integrity for the delivery of group psychoeducation for anxiety and depression

Researcher: Jonah Gosling (Jgosling3@sheffield.ac.uk)

Please initial box

1. I confirm that I have read and understood the information sheet dated 15th March 2021 (version 3.0) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw my data up to a specific date (tbc) without giving any reason, without my medical care or legal rights being affected. ☐

3. I give permission for members of the research team to have access to my demographic information, clinical data, previous episodes of IAPT care and self-report outcome measures completed as part of symptom monitoring for group therapy. I understand that my name will not be linked with the research materials, and that I will not be identified or identifiable in the results of the research. ☐

4. I understand that the information collected from the study may be anonymously used in other future research, may be shared anonymously with other researchers and may be used to support other research. ☐

5. I agree to the recording of video group sessions and understand that these recordings are for the purposes of rating group facilitators only. I understand I will not be visible in these recordings. ☐

6. I agree to take part in the above research. ☐

Name of person taking consent	Date	Signature
Name of participant	Date	Signature

PRACTITIONER/RATER PARTICIPANT INFORMATION SHEET

IRAS Project ID: 293765

Project title: Developing and evaluating a measure of facilitator treatment integrity for the delivery of group psychoeducation for anxiety and depression

You are being invited to participate in a doctoral research project that is being completed by Jonah Gosling, Postgraduate Researcher and Trainee Clinical Psychologist, from the University of Sheffield. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information carefully and discuss with others if you wish. Please ask questions if there is anything that is not clear or if you would like more information.

Purpose of the research

Group sessions are often used and are increasingly being used by Improving Access to Psychological Therapies (IAPT) services as they have been found to be effective for reducing common mental health difficulties. However, a rating tool to measure the treatment integrity (the competency (skill) and adherence (fidelity to the therapeutic approach) with which the facilitator delivers the material) of group facilitators does not exist. Therefore, this project aims to develop and evaluate such a tool that can be used to rate whether facilitators are meeting the expected standards of services when delivering groups, ensuring they are providing patients with an effective and evidence-based intervention. The measure will also aim to develop IAPT practitioner training and improve supervision. We are also looking to investigate if better group facilitators result in improved outcomes, lower dropout rates and improved attendance.

What would taking part involve?

If you agree to take part in the study, you will continue to deliver group sessions and your role will remain the same. By agreeing to take part, you are consenting to the recording of group sessions that you deliver. Recorded group sessions will then be rated by senior PWPs, using the developed facilitator treatment integrity measure. In addition, patients who consent to participating will also complete a treatment evaluation questionnaire. By agreeing to take part, you will consent to your information being included in our analysis of facilitator treatment integrity and its relationship to patient outcomes. In addition, for analysis purposes and for the purposes of the study, some demographic data will be taken. All research data (including session recordings) will be stored on the secure University storage system, which will only be accessible by the researchers involved. All information collected by the research team will remain confidential and the data will be anonymous. If you choose not to take part, your information will not be passed to the research team.

Do I have to take part?

Participation in this research is voluntary. If you decide to take part, you will be given this information sheet to keep and will be requested to fill in a consent form and demographic information sheet prior to the first group session. You can decide to withdraw at any time without it being viewed negatively.

What are the possible benefits of taking part?

By consenting to your involvement in the study it will enable us to develop an integrity measure which will contribute to the improvement and further investigations of groups and their facilitators.

What are the possible risks and disadvantages of taking part?

There will be no change in your role or the treatment which you deliver to patients, so there are no perceived risks or disadvantages of taking part in the research. All data and recordings will be kept confidential and analysed anonymously.

Supporting information

How will my information be kept confidential?

The researcher will be responsible for the data and recordings. They will only use information that is needed for the research study and will let very few people know your name or contact details, and only if they really need it for this study. All data and recordings will be stored on University premises in a locked cabinet or in an electronic password protected folder only accessible by the researcher. The data will be analysed collectively so you will not be identified or identifiable within any reports or publications. In line with research requirements, any personal data will be securely stored for 3 years before it is destroyed. Also, anonymised research data will be stored on the secure University repository for 10 years.

What if I have any concerns, want to complain or something goes wrong?

If you have any concerns regarding the research, we ask that you contact either the named researcher or IAPT lead (details outlined below). Complaints regarding the research should be forwarded to Amrit Sinha, Research Support Officer, the University of Sheffield (contactable at a.sinha@sheffield.ac.uk or 0114 222 6650). In the event that something does go wrong and you are harmed in the study, public liability insurance is available via the University of Sheffield. In these circumstances you may also have grounds for legal action against Sheffield Health & Social Care NHS Foundation Trust, but you may have to pay your legal costs

What if I decide I don't want to be involved in the study anymore?

If you decide you do not want to be involved anymore, you will be required to contact IAPT Lead facilitator or the researcher using the details below. Any data that has been collected previously and passed to the research team will then be removed and destroyed.

Where can I find out more about how my information is used?

You can find out more about how your information will be used at https://www.hra.nhs.uk/information-about-patients/. Should you require any further details regarding how your information will be used, please contact the researcher (details below) or Amrit Sinha, Research Support and Data Protection Officer, the University of Sheffield (contactable at a.sinha@sheffield.ac.uk or 0114 222 6650).

What will happen to the results of this study?

The results will make up part of the researchers Clinical Psychology Doctorate thesis and be used to report back to the IAPT service to help inform treatment and service policy in order to improve group treatments for patients. If you would like to know more about the outcome of the study, please contact the researcher using the details below to receive outcome updates.

Ethical consent for this study has been obtained from the NHS Research Ethics Committee.

If you would like any further information about the study, please contact the researcher or IAPT Lead using the contact details below.

Researcher

Name: Jonah Gosling
Address: Psychology Department
University of Sheffield
Cathedral Court, 1 Vicar Lane
Sheffield, S1 2LT
Email: jgosling3@sheffield.ac.uk

IAPT Lead

Name: Jodie Millington
Address: Sheffield IAPT
St George's Health Centre
Winter Street, Sheffield,
S3 7ND
Telephone: 0114 226 4380

PRACTITIONER/RATER PARTICIPANT CONSENT FORM

IRAS Project ID: 293765
Participant ID Number:

Title of Project: Developing and evaluating a measure of treatment integrity for the delivery of group psychoeducation for anxiety and depression

Researcher: Jonah Gosling (Jgosling3@sheffield.ac.uk)

Please initial box

1. I confirm that I have read and understood the information sheet dated November 2020 (version 2.0) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw my data at any time without giving any reason, without my legal rights being affected. ☐

3. I give permission for members of the research team to record me facilitating group sessions as part of the IAPT service. I give permission for members of the research team to have access to the session videos and consent to them being rated anonymously. I understand that my name will not be linked with the research materials, and that I will not be identified or identifiable in the results of the research. ☐

4. I understand that the information collected from me will be used to support other research in the future, and may be shared anonymously with other researchers. ☐

5. I agree to take part in the above research. ☐

Name of Participant	**Date**	**Signature**
Name of Person taking consent	**Date**	**Signature**

Appendix C

Blank Expert and Patient Treatment Integrity Measures

Psychoeducational Treatment Integrity Measure (GPTIM)

Was the right style of psychoeducational content used?							
	0	**1**	**2**	**3**	**4**	**5**	**6**
The facilitators were clearly using a psychoeducational approach							
The psychoeducational information delivered was well matched to the needs of the group							
The session was underpinned by cognitive behavioural theory (and not another theory)							
	Not at all	**Infrequently**	**Somewhat**	**Quite a bit**	**Considerably**	**Extensively**	

How skillfully was the psychoeducational content delivered?							
	0	1	2	3	4	5	6
The facilitators shared and then abided by an agenda							
The facilitators paced the session appropriately							
The facilitators presented the materials in an engaging and enthusiastic manner							
	Incompetent	Novice	Advanced beginner	Competent	Proficient	Expert	

How skillfully was the psychoeducational content delivered?							
	0	1	2	3	4	5	6
The facilitators clearly & accurately communicated the psychoeducational information							
The facilitators presented change methods with clarity							
The facilitators provided guidance on the content of between session work ('homework')							
	Incompetent		**Novice**	**Advanced beginner**	**Competent**	**Proficient**	**Expert**

Overall TI Score (add the scores for all the items above together)	**/54**

Patient Facilitator Evaluation

(Patient Group Psychoeducation Treatment Integrity Measure (GPTIM))

Rate the facilitators from today's group session by selecting a number nearest to the description that best fits your experience. Ensure that you select an individual number and rate all areas.

1. The facilitators set clear aims for this session						
Strongly disagree			**Neither agree nor disagree**			**Strongly agree**

2. The facilitators approach meant the session flowed smoothly						
Strongly disagree			**Neither agree nor disagree**			**Strongly agree**

3. The facilitators explained the session content well						
Strongly disagree			**Neither agree nor disagree**			**Strongly agree**

4. The facilitators delivered the session with confidence, warmth and enthusiasm						
Strongly disagree			**Neither agree nor disagree**			**Strongly agree**

5. The facilitators made me feel that the psychological information was related to my difficulties						
Strongly disagree			**Neither agree nor disagree**			**Strongly agree**

6. The facilitators talked about psychological information in a jargon-free and confident way						
Strongly agree			**Neither agree nor disagree**			**Strongly disagree**

7. The facilitators provided a coping method in the session which I can use and the rationale for this was explained						
Strongly agree			**Neither agree nor disagree**			**Strongly disagree**

8. The facilitators described the new coping methods clearly, taking clarifying questions and answering them sufficiently, when needed						
Strongly agree			**Neither agree nor disagree**			**Strongly disagree**

9. The facilitators explained and set a homework task that was linked to the session						
Strongly agree			**Neither agree nor disagree**			**Strongly disagree**

Appendix D

Treatment Integrity Measure Manual

Group Psychoeducation Treatment Integrity Measure (GPTIM) Manual

Introduction

Practitioners delivering low intensity group psychoeducational interventions offer treatment for patients with mild-moderate depression and anxiety disorders. When delivering group

psychoeducational content, practitioners employ a didactic style, which results in the practitioner becoming a ‘teacher’ and the service user becoming a ‘learner’. Self-help materials based on cognitive behavioural theory and principles provide the focus for treatment. The guided self-help clinical method emphasizes the skill of practitioners in utilizing psychoeducational materials and helping patients use them effectively to self-manage their symptoms. The competence of the practitioner in delivering low intensity treatment is crucial to ensure the progress/safety of the patient. Competence refers to the skill and appropriateness with which techniques/methods are delivered. The adherence of the practitioner is also crucial, as it refers to the extent they are delivering techniques/methods that are consistent with the therapy model and/or protocol. Measurement of the competency and adherence components combine to assess a practitioner’s overall TI.

Treatment Using the COM-B as a Theoretical Guide

In relation to the low intensity cognitive behavioral approach, consideration of behaviour change theory is pivotal. It is imperative that practitioners are able to consider the ways behaviour change underpins the low intensity method and apply this knowledge within treatment. The integrative model of behaviour and behaviour change for low intensity cognitive behavioural practitioners is the COM-B model (Michie et al, 2014). The model conceptualises the patient’s problem behaviour as resulting from the interaction of three components: (a) capability to perform behaviour change (b) the opportunity to carry out necessary behaviour change and (c) the motivation for behaviour change. Therefore when treating patients using low intensity treatment methods, the COM-B model can be used to inform, guide and influence PWP treatment delivery. Practitioners should utilise the COM-B model to inform and influence the gathering and synthesis of information to aid clinical decision-making and treatment planning. The manner in which this can be achieved is set out below:

Capability

Considerations about the patient’s capability to engage in behavior change should be built into the treatment plan. The practitioner should show evidence of providing low intensity materials, exercises, interventions and techniques that enable the patients to change their behavior/reasoning/executive functioning. The practitioner should aim to facilitate the patients in developing a good understanding of their common mental health problems and also of the mechanisms that must be targeted to create change to promote recovery.

Opportunity

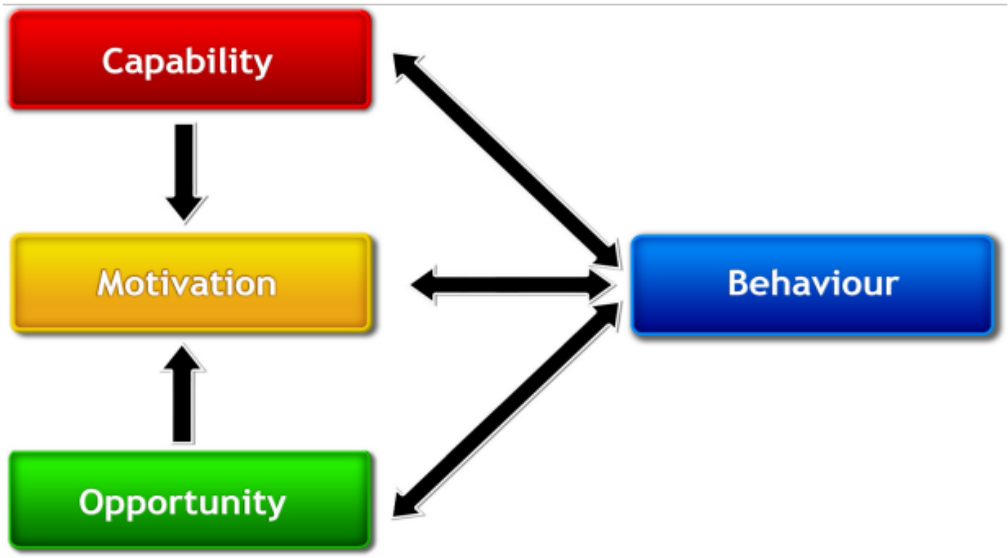
Consideration of the patient’s opportunity to engage in behavior change should be integrated into the treatment plan. Practitioners should focus on supporting the patient to change factors in their

environment (or their response to their environment) that would facilitate symptom change or lead to a reduction of the impact of their anxiety or depression.

Motivation

Practitioners should focus on addressing any issues with avoidance to enable effective engagement with self-management strategies i.e. in behavioral activation the focus would be on enhancing access to positive reinforcers. Alternatively, practitioners should aid patients in reduced cognitive/behavioral/emotional avoidance strategies that maintain their problem.

Raters should note the extent to which the practitioner applies the COM-B model to capitalise on opportunities to facilitate change. Application of the model is recognised as important as applying the framework assists practitioners to deliver in session interventions or between session work sensitively to patients.



Low-Intensity Psychoeducational Group TI Measure Manual

This manual outlines a scale for measuring the level of TI of low intensity cognitive behavioural practitioners during group treatment sessions. The scale contains 9 items which will enable raters to examine a range of adherence (3) and competency (6) components:

- Psychoeducational approach
- Psychoeducational needs
- Cognitive behavioural underpinning
- Agenda setting
- Effective use of time
- Engagement
- Psychoeducational communication

- Change methods presentation
- Between-session work guidance

The low intensity group psychoeducational integrity measure is a rating scale to be used by supervisors, trainers and managers to assess practitioner's performance in treatment sessions. Practitioners can make use of the self to self-rate sessions to enhance reflections and development.

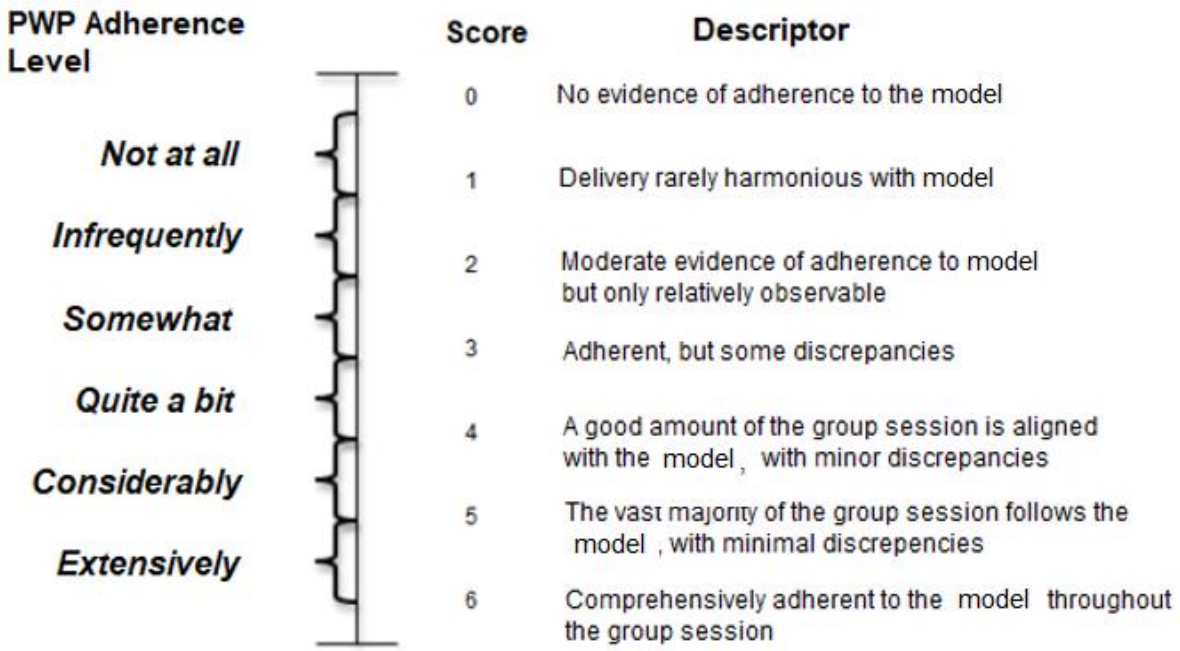
Scoring

The low-Intensity psychoeducational group TI measure uses two scoring scales to rate the facilitator competency of 6 areas and the facilitator adherence of 3 areas. Both scales utilise a 7-point Likert scale (0-6). For all items, raters must evaluate the extent to which an item is present/ (adherence/differentiation) and the skill with which the item is delivered (competency). The rating scales for both adherence and competence are defined in detail below. Furthermore, each items rating scale scores have been assigned specific descriptors to aid the rating process.

Adherence/Differentiation Scale

As outlined, the adherence/differentiation scale (below) provides a rating format to evaluate the extent to which a practitioner follows an item's protocol. The higher ratings reflect behaviours that are more consistently in line with the expected protocol. Whereas, the middle range scores reflect a practitioners inconsistent and variable fidelity to the protocol. The lowest scores indicate the complete or partial absence of the practitioner following the protocol.

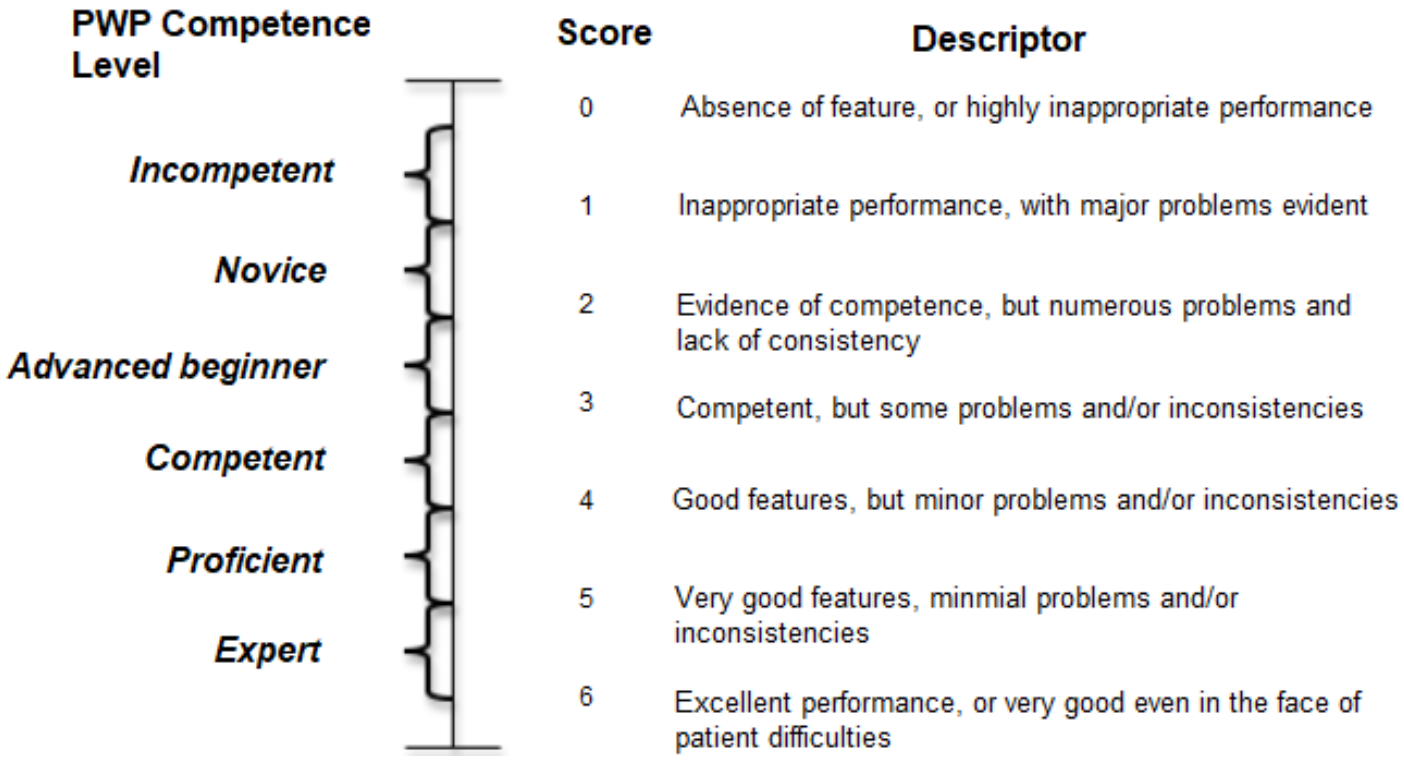
For this rating scale, the starting point for practitioners is "0". The rater should assign a number greater than "0" only if they observe examples of the item protocol behaviour. When applying the measure to managing stress sessions, raters should score facilitators according to their adherence to a cognitive behavioural model.



Competency Scale

The competency rating scale (below) provides a clear format by which to rate how skilled and appropriately the practitioner delivers the item in question. The higher ratings reflect a practitioner who displays a thoroughly adept delivery of the item, across the span of the group session. The middle range scores are reflective of practitioners who are observed to be adequately competent,

but problems and inconsistencies are present at varying levels. The lowest scores reflect either the inappropriate delivery of the item or its complete absence. The highest score is often characterised by the application of competencies in the face of patient difficulties. However, it is possible to score a 6 in the absence of patient difficulties should the rater feel this provides the most accurate rating of the practitioners competence. For this rating scale, the starting point for each item should be “3”, as the raters should start by assuming that the practitioner will perform at a competent/average level.



Scoring Guidelines

In order to establish a consistent and reliable approach between raters, a set of clear scoring guidelines have been created which all raters should abide by. These guidelines should be followed when rating all of the items within the measure.

- All ratings should be based on the **performance of both facilitators as a pair**
- All items refer to the facilitator’s behaviour and therefore raters should consider what the therapist **actually does in the session** not what they might have intended to do
- As raters must make intricate distinctions among therapist behaviours, it is essential that the raters **listen and observe the session without distraction**

- Raters must also ensure they **rate what actually occurred**, not what they think ought to have occurred from their perspective. For example, ratings on one item should not have a bearing on others, raters liking/disliking of the practitioner should be irrelevant and how skilled the therapist believes the practitioner to be should be insignificant

- Raters must **use the rater's manual during each rating**, as this will prevent rater drift and ensure that the process is more reliable and consistent

- Raters should only rate by selecting **whole numbers** and **must rate every item**

- All tapes and recordings are **confidential** and must be handled as if they were medical records. All ratings should be completed in an appropriate place away from individuals not involved in the study

Adherence/Differentiation Items:

Was the right style of psychoeducational content used?

Items:

Psychoeducational approach

Psychoeducational needs

Cognitive behavioural underpinning

Psychoeducational Approach

Psychoeducational groups are recognised as low-intensity interventions due to their concentration on providing mental health information. Whilst delivering groups, facilitators have minimal contact with patients (i.e. in a group didactic format over limited sessions) and present the information at a low-intensity level that is not overcomplicated. Facilitators also must ensure that a psychoeducational approach is clearly observable. Typically, this can take the form of providing information regarding psychological concepts (i.e. stress) and introducing strategies to alleviate the introduced psychological concepts (i.e. worry time). Although the inclusion of these is dependent on the session number, hallmarks of this approach should be noticeable throughout.

Psychoeducational Approach
- Are the facilitators clearly delivering the group using a psychoeducational approach? - Do the facilitators provide the patients with mental health information through their delivery of the content? - Are the typical hallmarks of a psychoeducational approach evident (i.e. introduction of psychological concepts, rationale regarding why change may be helpful, introduction of a change strategy)?

Psychoeducational Approach Scale

Rating	Descriptor
0	No evidence of a psychoeducational approach used.
1	Sparse evidence of psychoeducational approach being used. Possibly only limited or some mental health information given, for example.
2	Facilitators evidence moderate use of the psychoeducational approach. This is only relatively observable, however.
3	Practitioners are observed using the psychoeducational approach, with some hallmarks noticeably present. Some discrepancies (i.e. not always clear or noticeable).
4	As above, but with a good amount of the session using a psychoeducational approach. Only minor discrepancies evident.
5	Large majority of session is based on the use of a psychoeducational approach, with only minimal discrepancies evident. Evidence of psychoeducational mental health information twinned with possible strategies (if suitable to session number).
6	As above, but no discrepancies and psychoeducational approach is comprehensively applied throughout the group session.

Psychoeducational Needs

The facilitators of psychoeducational groups should ensure that the information they deliver is well-matched to the needs of the group. Importantly, those who attend psychoeducational groups are referred so they can access low-intensity interventions. Therefore, the content should be reflective of this. Facilitators should also deliver content which is relevant to the generalised needs of patient's within psychoeducational groups, which typically relates to experiences of stress, depression and/or anxiety. Information that appears to cover more complex or unrelated difficulties will consequently not match the group aims. Facilitators should also ensure that earlier sessions focus on information giving, whilst later sessions concentrate on enabling change.

Item Rater Questions
- Do the facilitators deliver psychoeducational information at a low-intensity level? - Is the information they deliver overcomplicated or pitched incorrectly? - Do the facilitators deliver psychoeducational information that is related to the needs of the group (i.e. associated to stress, depression and/or anxiety)? - Is the content delivered in line with the stage the group is at on the course (i.e. information giving for earlier sessions, change methods for later sessions)?

Psychoeducational Needs Scale

Rating	Descriptor
0	No evidence of facilitators matching psychoeducational information to group needs. Possibly inappropriate information delivered too early/too late.
1	Sparse evidence of practitioners matching psychoeducational information to the needs of the group. Limited attempts made, session may glaze over such concepts and possibly inaccurate if covered.
2	Practitioners evidence moderate matching of psychoeducational information to the needs of the group. This is only relatively observable however and is somewhat accurate (i.e. slightly relates to stress, depression and/or anxiety).
3	Practitioners match a reasonable amount of the psychoeducational information to the needs of the group but some discrepancies are evident (i.e. may be overcomplicated, pitched in the wrong session and/or at the wrong level).
4	Good amount of the session matching the psychoeducational information to the needs of the group. Only minor discrepancies evident.
5	Practitioners match large majority of psychoeducational information to the needs of the group (mostly relating to stress, depression and/or anxiety). Only minimal discrepancies evident.
6	As above, but no discrepancies and facilitators comprehensively matched psychoeducational information to the needs of the group throughout the session. Appropriate content delivered for the stage of the group.

Cognitive Behavioural Underpinning

An important component of psychoeducational groups is that they are heavily underpinned by cognitive behavioural theory. Therefore, group facilitators must ensure that they deliver content related to this theory. Facilitators must also ensure that they do not deliver content from other theories in conjunction with or instead of cognitive behavioural concepts and/or theory. Typically, the theory is supported by the use of the five areas model, which formulates an individual's difficult situation in relation to their thoughts, physical feelings, behaviours and mood. Whilst presenting, the facilitators may make reference to this model or areas of the model to reinforce the cognitive behavioural underpinning of the information covered. Facilitators may also make reference to aspects of cognitive behavioural theory which are not as explicit (i.e. referring to the inter-related nature of thoughts and behaviours etc.).

Item Rater Questions
- Do the facilitators discuss psychoeducational concepts that are underpinned by cognitive behavioural theory? - Do the practitioners inappropriately attempt to introduce or reference other theory? - Are the concepts that are covered underpinned by cognitive behavioural theory? - Do the facilitators reference the five areas model (thoughts, feelings, behaviours, mood, trigger) in relation to the content covered? - Do the facilitators make reference to aspects of cognitive behavioural theory in passing or in a less explicit manner?

Cognitive Behavioural Underpinning Scale

Rating	Descriptor
0	No evidence of a cognitive behavioural underpinning.
1	Sparse evidence of practitioners underpinning the group session with cognitive behavioural theory. Some examples may be observable, but this may be limited to glazing over certain concepts and not covering them fully.
2	Practitioners evidence moderate underpinning of session with cognitive behavioural theory. These are relatively observable.
3	Practitioners underpin the session with cognitive behavioural theory but some discrepancies are apparent (i.e. partial use of another model, partial coverage of cognitive behavioural theory information).
4	As above, but with a good amount of the session underpinned by cognitive behavioural theory with only minor discrepancies evident.
5	Practitioners underpin the large majority of the session with cognitive behavioural theory. Most of the content appears related to the theory and only minimal discrepancies are evident (i.e. subtle misinformation or omission).
6	As above, but no discrepancies and cognitive behavioural theory comprehensively underpins the whole group session.

Competence Items:

How skillfully was the psychoeducational content delivered?

Items:

Agenda

Effective use of time

Engagement

Psychoeducational communication

Change methods presentation

Between-session work guidance

Agenda

Low-intensity practitioners set an agenda to ensure that the key topics and information that need to be covered are done so in an efficient and time-ordered way. The agenda should aim to cover new concepts and/or concepts from the previous session, such as homework assignments. Facilitators should aim to demonstrate their ability to set, utilise and communicate a clear and structured agenda (with 2-4 items). The items included within the agenda must be appropriate, whilst also providing enough and not too much content for the allocated group time. This should result in a fluent and well-paced session overall. Importantly, practitioners will need to appropriately follow the agenda without shifting between different topics too quickly or slowly.

Item Rater Questions
- Was the practitioner fluent and well-paced in their adherence to the agenda? - Did the practitioner communicate the agenda clearly and succinctly? - Were the items included within the agenda appropriate? - Was the agenda followed in a clear and logical way? - Was time allocated efficiently to each of the items?

Agenda Scale

Rating	Descriptor
0	No focus of the session agreed or provided.
1	Ineffective agenda setting as key information omitted (i.e. failure to plan to discuss between session tasks). Very vague agenda. No fluency.
2	Framing provided, but vague and numerous problems evident with important information missing (i.e. failure to plan to discuss between session tasks). Lacks fluency and patchy adherence to agenda.
3	Competent and effective agenda set, but somewhat lacking in fluency. Key appropriate standing items of low-intensity sessions planned and outlined. Adherence to agenda a little inconsistent.
4	Clear agenda outlined and fluently delivered. Key standing items of low-intensity sessions planned and outlined. Good subsequent adherence.
5	As above with very good fluency in agenda setting and consistent adherence throughout. Key standing items of low-intensity sessions planned and outlined.
6	As above, but with excellent features. Possibly even in the face of patient difficulties.

Effective Use of Time

Low-intensity practitioners must ensure that they utilise the time available for the group session in accordance with the set agenda. Therefore, adequate control of the session must be exhibited. Facilitators should ensure that they do not rush pivotal aspects of the session. They should also display their ability to proficiently use the time allocated for the session and be seen to pace the session well (possibly referencing the time etc.). Effective pacing involves facilitating the flow of the group session through discreet start, middle and end phases. Facilitators must ensure that the pace of the session is appropriate for the group. Sessions should not go over the allotted time (unless due to unforeseen circumstances) and must neither be too fast or too slow. When switching between topics and/or agenda items, this should not be done too quickly, so as to ensure that the group is given enough time to understand the material adequately.

Item Rater Questions
- **Does the session flow smoothly between discreet phases (start, middle and end)?** - **Was the time allocated to each part of the session appropriate for the items? May involve ensuring not too much time and/or not enough is allocated to any items.** - **Did the pacing seem appropriate for the group/low-intensity style?** - **Does the session go over the allotted time?**

Rating	Descriptor
0	No evidence of effective use of time or pacing evident.
1	Only slight attempts at effectively using time in the session. Does not flow between phases. Major problems evident.
2	Some evident attempts to effectively use time in the session and some flow apparent between phases evident, but this is inappropriate and inconsistent.
3	Facilitators make appropriate attempts to use time effectively by moving between discreet phases, allocating sufficient time for each agenda item and more. However, this is done inconsistently and some problems are evident.
4	Good use of time with appropriate pacing of the session according to agenda and evidence of other features (i.e. flowing smoothly) which are more consistent than not. Only minor problems or inconsistencies evident.
5	Very good and consistent use of time within the session according to the agenda. Evidence of smooth transitions between phases, references to time, possible overview of timings stated. Minimal problems and/or inconsistencies.
6	As above but whilst moving very smoothly between session phases, with some reference to the time and/or agenda. Considered an excellent display of effective use of time.

Engagement

Facilitators should present materials in an engaging and enthusiastic manner to maintain the engagement of patients to aid their connection to the materials being presented. If facilitators are unable to facilitate their engagement, patients may become disillusioned with the process which may impact the intervention effectiveness. Furthermore, as the low-intensity format is defined as guided self-help, it is imperative that patients leave sessions feeling empowered and enthusiastic about employing the concepts independently. Facilitators should present with an enthusiastic manner, displaying a warm, empathic and compassionate approach. This may result in them referencing how difficult some experiences can be. Information should be pitched at an accessible level, which may involve explaining concepts in a range of ways. Facilitators should also capture the attention of the patients and present with confidence and/or flair.

Item Rater Questions
- **Do the facilitators present the materials with confidence and/or flair?** - **Is there evidence of a warm, empathic and/or compassionate presenting style?** - **Does the practitioner present the content with enthusiasm?** - **Does the practitioner check regarding understanding?** - **Is the content paraphrased (if required) for the group and/or explained in a range of accessible ways?**

Engagement Scale

Rating	Descriptor
0	No attempt or evidence of practitioner presenting materials in an engaging way.
1	Very little attempt made by the practitioner to engage the group. Major problems evident. May appear unenthusiastic or uninterested.
2	Some inconsistent evidence of practitioner engaging the group. Problems evident in engaging patients with warmth, empathy and/or enthusiasm.
3	Apparent engagement of group, but inconsistent and some problems apparent. Some evidence of engaging patients warmth, empathy and/or enthusiasm.
4	As above, with good engagement of group, but minor problems and/or inconsistencies apparent. Possibly presents as confident, with some flair in delivery.
5	Very good engagement of group. Minimal inconsistencies apparent. Confidence, enthusiasm and flair apparent. Most information accessible. Evidence of paraphrasing or presenting information in different ways.
6	As above, but considered to be an excellent performance. Practitioner presents with engaging features throughout. May even manage to engage group in the face of patient difficulties. Confident throughout and flair is evident. Excellent engagement overall.

Psychoeducational Communication

Throughout group sessions, facilitators will be required to deliver and be knowledgeable regarding psychoeducational information that are typically delivered as part of a cognitive behavioural low-intensity intervention. When communicating key psychoeducational concepts, they should always be clear and accurate. Clarity refers to the coherence and simplicity with which the facilitators deliver information to the patients. On the other hand, accuracy refers to the precision with which the practitioner delivers information to the patients. Some sessions will introduce psychoeducational content and others may refer to content that has been covered previously. Whilst communicating these concepts, facilitators should ensure that they relay information confidently, are clear and engaging an empathic regarding the possible impact of the information for patients.

Item Rater Questions
- **Are concepts communicated with confidence?** - **Is the information that is delivered done accurately?** - **Does the practitioner show empathy when communicating concepts?** - **Is the practitioner clear and engaging in their delivery of the concepts?** - **Are patients given a chance to check their understanding?**

Psychoeducational Communication Scale

Rating	Descriptor
0	No evidence of appropriate psychoeducational communication.
1	Inappropriate performance. Observable lack of confidence in communicating concepts and a lack of empathy. Major problems evident.
2	Facilitators use some appropriate language (showing vague evidence of confidence and use of empathy), but numerous problems are apparent and features lack consistency.
3	Competent psychoeducational communication, but only somewhat confident in delivery and might only sometimes present information with clarity or present empathically. Some problems and inconsistencies apparent.
4	Clear communication of psychoeducational concepts with good features, but minor problems or inconsistencies evident.
5	Facilitators communicate psychoeducational content with confidence and display empathy. Communication clear and engaging. Patients given opportunity to check info. Only minimal problems or inconsistencies apparent.
6	As above, but considered an excellent performance throughout. May also be delivered in the presence of patient difficulties. No problems or inconsistencies.

Change Methods Presentation

Facilitators of psychoeducational groups are required to include the clear and accurate presentation of change methods (i.e. thought diaries, behavioural activation planners etc.). As group sessions are based on self-help principles, such methods must be presented effectively, so patients can independently apply the techniques. Thus, there needs to be a change method apparent in the session that would help or does help the patient to self-manage their difficulties more. If applicable, the practitioner should also evidence revisiting and reiterating change methods. Practitioners should ensure that if they present change methods, this must be done with a rationale and must be communicated clearly and accurately. They must also consider the obstacles that the patients may come across when applying the methods, to ensure that the independent application of the techniques is more likely to be effective.

Item Rater Questions
- Did the practitioners present the methods in a clear manner? - Does the practitioner explain the rationale regarding the change methods? - Do the change methods seem linked to the content? - Do they discuss possible obstacles related to the change methods? - If required, are methods further broken down for patients to understand?

Change Methods Scale

Rating	Descriptor
0	No evidence of within-sessions change methods being set or being explored.
1	Inappropriate setting up of and/or discussion of within-sessions change methods. Lack of rationale regarding methods discussed and they do not seem linked to the material. Obstacles not discussed and methods not broken down.
2	Attempts made to appropriately explore change methods, but inappropriate elements present (not clear, not engaging etc.). Some evidence of exploring a rationale and/or linking the task to sessions but may be problems with this.
3	Competent setting up of and/or discussion regarding change methods. Elements of competent delivery with regards to task rationale, session link, clarity, time etc. However, problems/inconsistencies with performance evident.
4	Good setting up of and/or discussion regarding change methods. As above, with good features and minor problems/inconsistencies.
5	Very good presentation of change methods in a clear and engaging manner. Rationale and link explored. Obstacles explored and methods breakdown covered (if required). Minimal problems evident.
6	As above, but an excellent performance and no problems/inconsistencies evident throughout. May have been delivered even in the face of patient difficulties.

Between-session Work Guidance

Patient progress is reliant upon facilitators setting tasks to be completed outside of the group environment. The tasks should ideally involve the testing of a hypothesis, the incorporation of new perspectives, and may suggest some aspect of behaviour change. Practitioners must always explain the rationale behind the homework and this should be linked to the content. Guidance around these tasks may involve preparation, where practitioners explore the rationale behind their use and how they can be applied. Practitioners may provide guidance on tasks that have been applied and may need some refining. Practitioners should clearly set up between-session tasks; delivering a rationale, along with a straight forward explanation around its application. They must also consider the obstacles that patients may come across to ensure they fully support the independent application of the task. Tasks should also be broken down effectively so they are manageable and accessible for the patients.

Item Rater Questions
- Were between-session task obstacles discussed? - Did the facilitators effectively break down the tasks so they are manageable? - Were they presented in a clear manner? - Is there a clear rationale discussed?

Between-session Work Guidance Scale

Rating	Descriptor
0	No evidence of between-session tasks being set and/or being explored.
1	Inappropriate setting up of and/or discussion of between-session tasks. Possible lack of rationale and lack of clear explanation.
2	Attempts made to appropriately set up homework task, but inappropriate elements. Some evidence of exploring a rationale may be problems with this.
3	Competent setting up of and/or discussion regarding between-session tasks. Elements of competent delivery with regards to task rationale, explanation, obstacles and session link. However, problems/inconsistencies evident.
4	Good setting up of and/or discussion regarding between-session tasks. As above, but with good features and minor problems/inconsistencies.
5	Very good setting up of and/or discussion of between-session tasks. Evidence of exploration of task rationale, a clear/engaging explanation around its application and obstacles. Minimal problems evident.
6	As above, but an excellent performance and no problems/inconsistencies evident throughout. May have been delivered even in the face of patient difficulties.

Appendix E

Content Validity Index Survey

Content Validity Index (CVI)

CVI Information Please rate the following items based on the degree of relevance you believe each item has in relation to the measured domain. For this block of questions (Q1-Q3), the items refer to the adherence/differentiation of the facilitators to the group psychoeducational format. Ensure you refer to the measure manual for specific descriptions for how each item is operationalised.

Q1 The facilitators were clearly using a psychoeducational approach

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q2 The psychoeducational information delivered was well matched to the needs of the group

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q3 The session was underpinned by cognitive behavioural theory (and not another theory)

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Please rate the following items based on the degree of relevance you believe each item has in relation to the measured domain. For this block of questions (Q4-Q9), the items refer to the competency of facilitators

when delivering the group psychoeducational format. Ensure you refer to the measure manual for specific descriptions for how each item is operationalised.

Q4 The facilitators shared and then abided by an agenda

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q5 The facilitators paced the session appropriately

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q6 The facilitators presented the materials in an engaging and enthusiastic manner

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q7 The facilitators accurately communicated psychoeducational information

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q8 The facilitators presented change methods with clarity

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q9 The facilitators provided guidance on the content of between session work ("homework")

- ○ Not relevant (1)
- ○ Somewhat relevant (2)
- ○ Quite relevant (3)
- ○ Highly relevant (4)

Q10 What is your age?

_____

Q11 What is your gender?

_____

Q12 How many years have you been qualified as a PWP?

_____

Q13 How many psychoeducational groups have you completed in your career (estimate if unsure)?

_____

Appendix F

Measure Training Presentation

Group Psychoeducation TI Measure

CONTRIBUTORS

Jonah Gosling, Stephen Kellett, Mel Simmonds-Buckley, Katarzyna Olenkiewicz-Martyniszyn and Dan Duffy

Background

- IAPT services are using psychoeducation groups at an increasing rate
- However, outcome research shows differences in patient outcomes despite the groups being manualised
- Currently, unlike individual LI sessions (i.e. CTRS, LIAC, LITC), group sessions **do not have a measure** to assess the extent facilitators deliver the model as intended or the level of skill with which they deliver the information
- Assessing these factors is important for the safety and wellbeing of the client, to ensure we are delivering evidence based practice, to enable us to train PWPs to a high level, to support the ongoing training and supervision of PWPs and to ensure that all the practitioners are operating at the same standards for audit and governance means.

Importance of TI

- 'TI' measures a composite of adherence and competence
- Competency and adherence have been found to be significant factors (at differing levels) in positively influencing patient outcomes
- Thus, developing a scale that measures TI may be more useful than measuring competency or adherence alone
- ***Adherence*** is related to the to the extent the practitioner is delivering techniques/methods that are consistent with the therapy model and/or protocol
- ***Competency*** refers to the skill and appropriateness with which techniques/methods are delivered

Why Develop the Scale?

- **No current** competency, adherence or integrity **measure** available for group psychoeducation
- A measure is needed as part of the **PWP training**
- The model acknowledges and is **underpinned by the COM-B model**
- Can be used in OSCEs/simulations as **a standardized measure of integrity**
- Can be used in the services to **support supervision and audit**

The Scale

- Creation and development of the 'Low-Intensity Psychoeducational Group TI Measure'
- The scale requires the rater to rate 3 adherence items and 6 competency items to create an overall TI score
- The scale must be used in conjunction with the attached manual to ensure scoring is conducted accurately
- When completing the scale, the rater must rate both facilitators as a pair throughout the group session

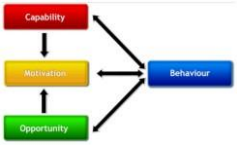
Low-Intensity Psychoeducational Group TI Measure

Behaviour Change & COM-B

- COM-B Model (Michie et al., 2014)

a – ***Capability*** to perform behaviour change

b – ***Opportunity*** to carry out necessary behaviour change c – ***Motivation*** for behaviour change



Behaviour Change & COM-B

- **CAPABILITY**

Considerations about the patient's capability to engage in BC should be built into the treatment plan. The practitioner should evidence providing LI materials, exercises, interventions and techniques that enable the patient to change their behavior/reasoning/executive functioning. The practitioner should aim to facilitate the patient in developing a good understanding of their common mental health problems and also of the mechanisms that must be targeted to create change to promote recovery.

- **OPPORTUNITY**

Consideration of the patient's opportunity to engage in BC should be integrated into the treatment plan. Practitioners should focus on supporting the patient to change factors in their environment (or their response to their environment) that would facilitate symptom change or lead to a reduction of the impact of their anxiety or depression.

- **MOTIVATION**

Practitioner should focus on addressing any issues with patterns of avoidance with a view to enable effective engagement with self-management strategies for example in behavioral activation the focus would be on enhancing access to positive reinforces. Alternatively practitioners should aid patients in reduced cognitive/behavioral/emotional avoidance strategies that maintain their problem.

Group Session Adherence/Differentiation Scoring

PWP Adherence Level	Score	Descriptor
Not at all	0	No evidence of adherence to the model
Infrequently	1	Delivery rarely harmonious with model
Somewhat	2	Moderate evidence of adherence to model but only relatively observable
Quite a bit	3	Adherent, but some discrepancies
Considerably	4	A good amount of the group session is aligned with the model , with minor discrepancies
Extensively	5	The vast majority of the group session follows the model , with minimal discrepencies
	6	Comprehensively adherent to the model throughout the group session

Group Session Competency Scoring

PWP Competence Level	Score	Descriptor
Incompetent	0	Absence of feature, or highly inappropriate performance
Novice	1	Inappropriate performance, with major problems evident
Advanced beginner	2	Evidence of competence, but numerous problems and lack of consistency
Competent	3	Competent, but some problems and/or inconsistencies
Proficient	4	Good features, but minor problems and/or inconsistencies
Expert	5	Very good features, minmal problems and/or inconsistencies
	6	Excellent performance, or very good even in the face of patient difficulties

Adherence and Competency Items

Competency Items Adherence Items

- Psychoeducationalapproach
- Psychoeducationalinfo well matchedto group needs
- Underpinned by CBT

Competency Items Adherence Items

- Agenda setting
- Effective use of time
- Engagement
- Psychoed communication
- Change methods communication
- Between-session workguidance

TI

Adherence/Differentiation Items

Psychoeducational Approach

Are the facilitators clearly delivering the group using a psychoeducational approach?

Do the facilitators provide the patients with mental health information through their delivery of the content?

Are the typical hallmarks of a psychoeducational approach evident (i.e. introduction of psychological concepts, rationale regarding why change may be helpful, introduction of a change

Psychoeducational Needs

- Do the facilitators deliver psychoeducational information at a low-intensity level?
- Is the information they deliver overcomplicated or pitched incorrectly?
- Do the facilitators deliver psychoeducational information that is related to the needs of the group (i.e. associated to stress, depression and/or anxiety)?
- Is the content delivered in line with the stage the group is at on the course (i.e. information giving for earlier sessions, change methods for later sessions)?

Adherence/Differentiation Items

Cognitive Behavioural Underpinning

- Do the facilitators discuss psychoeducational concepts that are underpinned by
- cognitive behavioural theory?
- Do the practitioners inappropriately attempt to introduce or reference other theory?
- Are the concepts that are covered underpinned by cognitive behavioural theory?
- Do the facilitators reference the five areas model (thoughts, feelings, behaviours, mood, trigger) in relation to the content covered?
- Do the facilitators make reference to aspects of cognitive

Competency Items

Agenda Setting

- Are the facilitators fluent and well- pa in their adherence to the agenda?
- Is the agenda communicated clearly a succinctly?
- Are the agenda items appropriate?

Effective Use of Time

- Does the session flow smoothly between discreet phases (start, middle and end)?
- Was the time allocated to each part of the session appropriate for the items? May involve ensuring not too much time and/or not enough is allocated to any items.

Competency Items

Engagement

- Do the facilitators present the materials with confidence and/or flair?
- Is there evidence of a warm, empathic and/or compassionate presenting style?
- Does the practitioner present the content with enthusiasm?
- Does the practitioner check regarding understanding?
- Is the content paraphrased (if required) for the group and/or explained in a range of accessible

Psychoeducational Communication

- concepts communicated with onfidence?
- he information that is delivered one accurately?
- es the practitioner show mpathy whencommunicating concepts?

Competency Items

Change Methods Communication

- Did the practitioners present the methods in a clear manner?
- Does the practitioner explain the rationale regarding the change methods?
- Do the change methods seem link to the content?
- Do they discuss possible obstacles related to the change methods?
- If required, are methods further broken down for patients to

Between-session Work Guidance

- Were between-session tasks obstacles discussed?
- Did the facilitators effectively break down the tasks so they are

Rating Examples

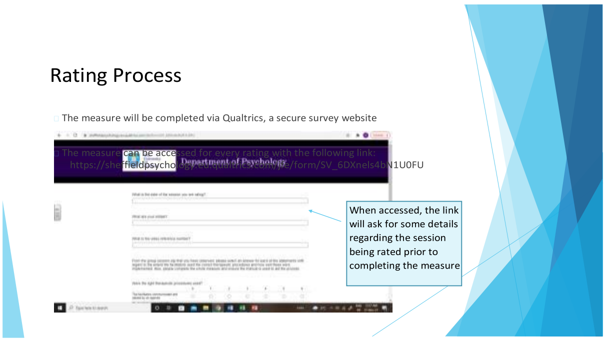
Psychoeducational Approach Scale

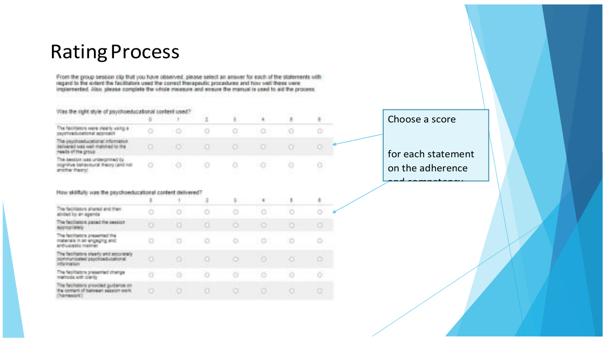
Rating	Descriptor
0	No evidence of a psychoeducational approach used.
1	Sparse evidence of psychoeducational approach being used. Possibly only limited or some mental health information given, for example.
2	Facilitators evidence moderate use of the psychoeducational approach. This is only relatively observable, however.
3	Practitioners are observed using the psychoeducational approach, with some hallmarks noticeably present. Some discrepancies (i.e. not always clear or noticeable)
4	As above, but with a good amount of the session using a psychoeducational approach. Only minor discrepancies evident.
5	Large majority of session is based on the use of a psychoeducational approach, with only minimal discrepancies evident. Evidence of psychoeducational mental health information twinned with possible strategies (if suitable to session number).
6	As above, but no discrepancies and psychoeducational approach is comprehensively applied throughout the group session.

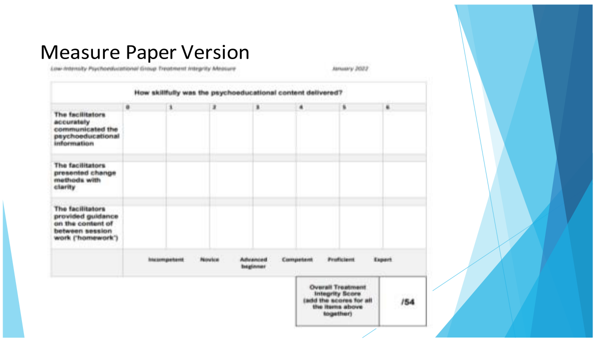
The Study and Your Role

- The study is part of Jonah Gosling's DClinPsych Thesis
- Sheffield IAPT is a trailblazer as they are the only service involved
- Sheffield IAPT will also be the first service to utilise the measure following its development
- Aim to make the measure available for the NHS to use nationally across IAPT services

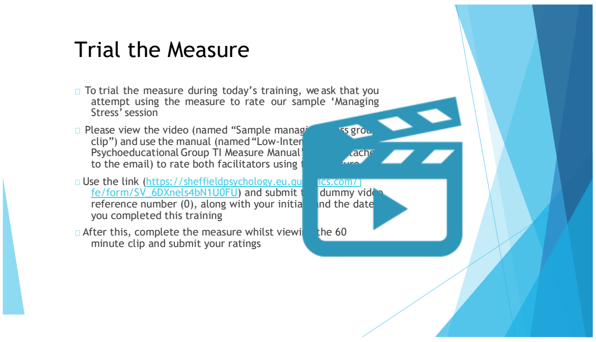
- We are asking you to rate 3-4 managing stress sessions using the measure
- These will be sent to you electronically
- Ratings will be used to assess the psychometric properties of the measure


Rating Process
The measure will be completed via Qualtrics, a secure survey website
When accessed, the link will ask for some details regarding the session being rated prior to completing the measure


Rating Process
Choose a score
for each statement on the adherence


Measure Paper Version
Low Intensity Psychoeducational Group Treatment Integrity Measure
January 2022
How skilfully was the psychoeducational content delivered?
0
1
2
3
4
5
6
The facilitators accurately communicated the psychoeducational information
The facilitators presented change methods with clarity
The facilitators provided guidance on the content of between session work ("homework")
Incompetent
Novice
Advanced beginner
Competent
Proficient
Expert
Overall Treatment Integrity Score (add the scores for all the items above together)
/54


Thanks for your time!
If you have any questions please email us at:
jgosling3@sheffield.ac.uk


Trial the Measure
To trial the measure during today's training, we ask that you attempt using the measure to rate our sample 'Managing Stress' session
After this, complete the measure whilst viewi the 60 minute clip and submit your ratings


Trial the Measure

Appendix G

Rater Measure Training Evaluation Form

Training Workshop

Evaluation Form

Please indicate your level of agreement with the following statements;

Q1 The training workshop increased my knowledge of what TI, adherence and competency are.

- ○ Strongly agree (1)
- ○ Somewhat agree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat disagree (4)
- ○ Strongly disagree (5)

Q2 Accessing the training workshop in a pre-recorded format was an effective medium to learn about and practice using the TI scale.

- ○ Strongly agree (11)
- ○ Somewhat agree (12)
- ○ Neither agree nor disagree (13)
- ○ Somewhat disagree (14)
- ○ Strongly disagree (15)

Q3 The training workshop increased my understanding of why a TI scale is useful to implement when delivering interventions.

- ○ Strongly agree (1)
- ○ Somewhat agree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat disagree (4)
- ○ Strongly disagree (5)

Q4 The training workshop instructions regarding the implementation of the TI scale were practicable.

- ○ Strongly agree (1)
- ○ Somewhat agree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat disagree (4)
- ○ Strongly disagree (5)

Q5 After the training workshop I feel confident about using the TI scale in my clinical practice as a facilitator or as a supervisor.

- ○ Strongly agree (1)
- ○ Somewhat agree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat disagree (4)
- ○ Strongly disagree (5)

Q6 I think integrating the TI scale into my clinical practice (as a facilitator or supervisor) would be helpful to the implementation of IAPT CBT groups.

- ○ Strongly agree (1)
- ○ Somewhat agree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat disagree (4)
- ○ Strongly disagree (5)

Q7 Please add any additional comments if required:

_____

Appendix H

Patient Health Questionnaire-9 (PHQ-9)

PATIENT HEALTH QUESTIONNAIRE-9 (PHQ-9)

Over the last 2 weeks, how often have you been bothered by any of the following problems? *(Use "✓" to indicate your answer)*	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

FOR OFFICE CODING *0* **+** _____ **+** _____ **+** _____

=Total Score: _____

Generalised Anxiety Disorder-7 (GAD-7)

GAD-7 Anxiety

Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3

Column totals _____ + _____ + _____ + _____ =

Total score _____

If you checked any problems, how difficult have they made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult
☐	☐	☐	☐

Source: Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD-PHQ). The PHQ was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke, and colleagues. For research information, contact Dr. Spitzer at ris8@columbia.edu. PRIME-MD® is a trademark of Pfizer Inc.

Appendix I

Factor Analysis Scree Plot