Defining the Assessment and Treatment Competencies to Deliver Low-Intensity Cognitive Behavior Therapy: A Multi-Center Validation Study

KELLETT, S., SIMMONDS-BUCKLEY, M., LIMON, E., HAGUE, J., HUGHES, L., STRIDE, C. and MILLINGS, Abigail

Citation:

Copyright and re-use policy

Defining the Assessment and Treatment Competencies to Deliver Low Intensity Cognitive

Behavior Therapy: A Multi-Centre Validation Study

Stephen Kellett*

University of Sheffield, UK, and

Sheffield Health and Social Care NHS Foundation Trust, UK

Mel Simmonds-Buckley

University of Sheffield, UK

Emma Limon

University of Sheffield, UK

Jennie Hague

University of Sheffield, UK

Lucy Hughes

University of Sheffield, UK

Chris Stride

University of Sheffield, UK

&

Abigail Millings

University of Sheffield, UK

* Correspondence to Stephen Kellett, Clinical Psychology Unit, Cathedral Court, University

of Sheffield, S1 1HD, UK; email: s.kellett@sheffield.ac.uk Tel: 0044 222 6537

ABSTRACT

Despite the vastly increased dissemination of the low intensity (LI) version of cognitive behavior therapy (CBT) for the treatment of anxiety and depression, no valid and reliable indices of the LI-CBT clinical competencies currently exist. This research therefore sought to develop and evaluate two measures; the low intensity assessment competency scale (LIAC) and the low intensity treatment competency scale (LITC). Inductive and deductive methods were used to construct the competency scales and detailed rating manuals were prepared. Two studies were then completed. The first study used a quantitative, fully-crossed design and the second a multi-centre, quantitative longitudinal design. In study one, novice, qualified and expert LI-CBT practitioners rated an LI-CBT assessment session (using the LIAC) and an LI-CBT treatment session (using the LITC). Study two used the LIAC and LITC across four training sites to analyse the competencies of LI-CBT practitioners over time, across raters and in relation to the actor/patients feedback concerning helpfulness, the alliance and willingness to return. Both the LIAC and LITC were found to be single factor scales with good internal, test-retest reliability and reasonable inter-rater reliability. Both measures were sensitive to measuring change in clinical competence. The LIAC had good concurrent, criterion, discriminant and predictive validity, whilst the LITC had good concurrent, criterion and predictive validity, but limited discriminant validity. A score of 18 accurately delineated a minimum level of competence in LI-CBT assessment and treatment practice, with incompetent practice associated with patient disengagement. These observational ratings scales can contribute to the clinical governance of the burgeoning use of LI-CBT interventions for anxiety and depression in routine services and also in the methods of controlled studies.

Defining the Assessment and Treatment Competencies to Deliver Low Intensity Cognitive

Behavior Therapy: A Multi-Centre Validation Study

Demand for talking therapies for anxiety and depression increases year on year (Pollecoff, 2016). Services have responded by changing organisational systems, workforces and treatment approaches to increase access to evidenced-based interventions (Firth, Barkham & Kellett, 2015). The most significant change in the mental health workforce has been the development of non-professional psychological practitioner roles that require less advanced training than traditional professional therapists (Fairburn & Patel, 2014). In stepped-care service designs, patients with mild to moderate anxiety and depression are initially offered effective, brief, low intensity (LI) guided self-help interventions delivered by these non-professional psychological practitioners (Bower & Gilbody, 2005). The clinical method of LI work is grounded in cognitive behavioral theory and is therefore termed low intensity cognitive behaviour therapy (LI-CBT; Ali et al, 2014). The evidence base for low intensity interventions covers over 30 reviews and meta-analyses (Delgadillo, 2018). This includes, for example, evaluations of bibliotherapy (Cuipers, 1997), technology assisted self-help (Karyotaki et al, 2017), guided self-help (Farrand & Woodford, 2013) and guided self-help to prevent the onset of depression (Cuijpers, Muñoz, Clarke & Lewinsohn, 2009). Increasing access to evidenced-based interventions in low/middle-income countries particularly emphasises the role played by LI-CBT (Singla, Kohrt, Murray, Anand, Chorpita, & Patel, 2017).

Bennett-Levy, Richards and Farrand (2010) established the defining conceptual and clinical features of LI-CBT. LI-CBT is based on the behavior change model (Michie, van Stralen & West, 2011) and uses highly structured psychoeducational workbooks to deliver treatment. Due to the high levels of practitioner guidance evident during LI-CBT, the role has been likened to a 'coach' rather than a therapist (Turpin, Clark, Duffy & Hope, 2009). High intensity CBT with anxiety and depression is underpinned by a competency framework (Roth & Pilling, 2007) that has a raft

of associated generic (e.g. the Cognitive Therapy Scale-Revised; Blackburn et al. 2001) and

disorder-specific competency measures (e.g. Cognitive Therapy Competence Scale-Social

Phobia; Consbruch, Clark, & Stangier, 2012). This body of work operationalising and measuring

high intensity CBT competency has unfortunately not been mirrored in LI-CBT.

The Improving Access to Psychological Therapy (IAPT) initiative in the United Kingdom

(UK) is an example of scaling up of clinical guidelines for anxiety and depression for national

delivery (Clark, 2011). LI-CBT psychological practitioners in IAPT are trained to assess for and

deliver psychoeducational guided self-help interventions for mild-to-moderate anxiety and

depression (University College London, 2015). IAPT has radically transformed the topography

of psychological therapy services, with LI-CBT practitioners being a significant component of

the workforce (Green, Barkham, Kellett & Saxon, 2014). LI-CBT interventions appeal to service

managers and commissioners (Delgadillo et al., 2018), due to being brief (i.e. 6-8, 35-minute

treatment sessions), low-cost (i.e. delivered by non-professionals), flexible (i.e. delivered one-to-

one and also in large groups) and accessible (i.e. low patient burden, as treatment is also

delivered via the internet, email or telephone).

Many patients are being assessed for and treated with LI-CBT interventions (e.g. 1,440,000

referrals were made to IAPT in 2017-18; 1,010,000 began a course of treatment, with 555,000

finishing treatment). Due to the high numbers of patients being assessed and treated with LI-

CBT, defining and measuring LI-CBT competencies plays an important role in terms of ensuring

patient safety (Ginzburg et al., 2012). Equally importantly, the lack of valid and reliable measures

of clinical competency restricts the delivery of evidence-based psychological therapies in routine

services (Kohrt et al., 2015). Fairburn & Cooper (2011) would conceptualize LI-CBT treatment

competency in a routine service delivery context, as the practitioner's ability to adhere to the

structure of the psychoeducational workbook, with treatment quality being the expertise with

which the guided self-help is facilitated. Meta-analytic evidence of the association between

competency and outcome varies. When therapies are combined there is little association (Webb et al. 2010), but when limited to solely CBT studies, there is a small significant total effect (r=0.24; Zarafonitis-Müller et al. 2014). When disorder-specific protocols are used to guide intervention, there is a clear association between competency and outcome (Ginzburg et al. 2012).

Whilst an LI-CBT adherence measure has been previously developed, this was never validated (Richards & Whyte, 2009). The decision to develop a new LI-CBT competency measure (i.e. rather than validate the extant adherence measure) was because this would better support the case management supervision of LI-CBT practitioners (i.e. by focusing on skilfulness of the guided self-help delivered, rather than whether guided self-help was delivered). The LI-CBT role requires close case management supervision due to the relative lack of knowledge and experience of the psychological practitioners compared to CBT (UCL, 2015). Observational measures for use by LI-CBT supervisors were therefore developed to support clinical governance efforts in routine services and because observational measures of clinical competency are considered 'gold standard' (Fairburn & Cooper, 2011). Self-assessment of competency is prone to marked bias (Brosan, Reynolds & Moore, 2008), as practitioners with less experience particularly tend to over-rate their abilities (Kruger & Dunning, 1999). There have been previous calls for competency measures to support the work and development of LI-CBT practitioners (Burns, Kellett & Donohoe, 2015).

The rationale for the present study was that the development and validation of observational LI-CBT competency measures would enable (a) the sound governance of the large numbers of patients screened for LI-CBT in clinical services, (b) accurate monitoring of competency with which LI-CBT interventions are delivered and (c) support for the methods of LI-CBT clinical trials. Here, two studies are described that have developed and then tested the reliability and validity of two LI-CBT competency measures; the low intensity assessment competency scale (LIAC) and the low intensity treatment competency scale (LITC).

LI-CBT COMPETENCIES

Method

Development of the LI competency framework, measures and manuals

Ethical approval for the project was granted (ref: 006168). The low intensity assessment competency measure (LIAC) and the low intensity treatment competency measure (LITC) were designed to reflect a blend of global competencies (i.e. the abilities to collaborate and promote a strong alliance; Southam-Gerow & McLeod, 2013) and also limited-domain competences (e.g. skilfulness in delivering LI-CBT change methods; Barber, Sharpless, Klostermann, & McCarthy, 2007). The behaviour change model provides the theoretical basis of LI-CBT assessment and treatment work (UCL, 2015) and the LIAC and LITC measures were grounded in this framework. LIAC and LITC were designed solely for assessing the clinical competency of LI-CBT work with mild-to-moderate anxiety and depression presentations.

The process for item generation for each scale followed the McCoach, Gable and Madura (2013) guidelines consisting of, (a) a literature review of existing CBT competency measures and manuals, (b) confirmation that there were no existing competency measures adequately serving the same purpose, (c) description of the LI-CBT assessment and treatment competency domains and providing preliminary conceptual definitions and finally, (d) specifying the potential dimensions of the assessment and treatment competency domains. Once these LI-CBT assessment and treatment competency domains were delineated, an item pool for each measure was generated via a combination of deductive and inductive methods (Hinkin, 1995). The deductive methods were supported via a literature review of extant CBT competency measures and manuals (Limon, 2017). The literature review concerned the psychometric evidence base of existing measures of CBT competency and a search of the manuals for common items. The Richards and Whyte (2009) LI-CBT adherence measure was also accessed. The inductive methods comprised potential items being considered and evaluated in five successive focus groups containing the same three LI-CBT trainers. The Dreyfus (1989) 7-point competency rating system was adopted so on each scale, competencies were measured in an ordinal manner (i.e., in

LI-CBT COMPETENCIES

an ascending order without any overlap, with each point on the rating scale being well described in the associated manuals).

Five LI-CBT practitioners employed in IAPT services then used the content validity index (CVI; Lynn, 1986) to assess and rate the relevance of the proposed final items of the LIAC and LITC prior to factor analysis. Convergent scores for items over .67 are considered acceptable (Lynn, 1986), with ratings higher than .9 showing excellent content validity (Polit & Beck, 2006). The CVI results showed complete agreement across the five raters for all competency items included in the final versions of the LIAC and LITC. The LIAC is supported by a 16-page rating manual that describes in detail effective conduct of an LI-CBT assessment session; 1) *introducing the assessment session;* 2) *establishing engagement;* 3) *interpersonal skills;* 4) *information gathering relevant to the problem;* 5) *information giving relevant to the problem* and 6) *shared planning and decision making*. The overall purpose of an assessment session is to ascertain the presenting problem, assess risk, consider suitability for an LI-CBT intervention and then allocate to that intervention or to 'signpost' to other more appropriate services. The LITC is supported by a 28-page rating manual that contains summaries of the competencies needed to deliver the core LI-CBT interventions, and each are described in terms of stage of treatment and thus allow competency rating to occur across early, middle and late LI-CBT treatment sessions. The LITC comprised summary competency items relating to the conduct of a LI-CBT treatment session; 1) *focussing the treatment session,* 2) *maintaining engagement,* 3) *interpersonal skills,* 4) *information gathering specific to change,* 5) *delivery of a low intensity change method* and 6) *homework planning*. The overall purpose of treatment sessions is delivery of an LI-CBT intervention.

Study One: Method

Aims and Hypotheses

The aim of this study was to test the construct validity, predictive validity, internal and interrater reliabilities of the measures in a cross-over design. The study used three groups of LI-

LI-CBT COMPETENCIES

CBT practitioners that varied in level of experience (i.e. novice, qualified, expert LI-CBT

psychological practitioners). Specifically, it was hypothesized that the predictive validity of the

measures would be established by the three groups differing in their ratings of LI-CBT

competencies, with novice practitioners scoring highest in their ratings, due to the influence of

cognitive bias (Kruger & Dunning, 1999).

Participants

The total sample was N=162. The expert-rater group consisted of n=24 LI-CBT trainers

(71% female, with a mean age of 35 and over five-years' experience of teaching LI-CBT) from

n=4 Universities. The qualified group was n=59 qualified LI-CBT practitioners drawn from n=13

IAPT stepped care services (all trained to PG Certificate level, 71% female, with a mean age of

38 and minimum of 2-years' experience of supervising LI-CBT work). The novice group was

n=79 trainee LI-CBT practitioners (90% female, with a mean age of 28). Each group received a

1-hour training session that introduced the manual and the measure.

Design

The research design demanded the generation of intra-class correlation coefficients (ICCs)

to assess levels of consistency amongst observational ratings of competency provided by multiple

raters, and was therefore a fully-crossed design (Hallgren, 2012). Fully-crossed designs allow for

systematic bias between raters to be assessed and controlled for in an ICC estimate (Putka, Le,

McCloy & Diaz, 2008).

Materials

Each of the three groups watched two sessions; (a) *video A* - a 45-minute LI-CBT

assessment session of a depressed patient (i.e. rated using the LIAC) and (b) *video* B - a 35-

minute LI-CBT treatment session for mild obsessive compulsive disorder treated with exposure

(i.e. rated using the LITC).

LI-CBT COMPETENCIES

Procedure

The three groups rated *video A* with the LIAC and *video B* with the LITC. Both sessions had been previously rated by five different experienced trainers independently; all had rated each session as competent LI-CBT.

Data Analysis

Exploratory factor analyses using principal axis factoring from a direct oblimin (oblique) rotation assessed the construct validity of the measures and an ANOVA of the ratings of the three groups tested discriminant validity. Internal reliability across items was determined via Cronbach alpha scores (scores above .70 are considered acceptable; Cronbach, 1951), item-total correlations (scores above .30 are considered acceptable; Cristobal, Flavian, & Guinaliu, 2007; Streiner & Norman, 2003) and Guttmann split-half reliabilities. Interrater reliability was assessed via two-way mixed effects intra-class correlation coefficients (ICC; Shrout & Fleiss, 1979). The degree of interrater reliability was interpreted using the Koo and Li (2016) ranges: <.5 (poor), .5-.74 (moderate), .75-.90 (good) and >.90 (excellent).

Study One: Results

Construct Validity

The Kaiser-Meyer Olkin measure of sampling adequacy suggested that both the LI-CBT assessment (KMO=.769) and treatment (KMO=.827) session ratings were adequate for factor analysis. Barlett's test indicated no sphericity concerns for either measure (LIAC: $\chi^2(15)=155.26$, p<0.001 and LITC: $\chi^2(15)=310.546$, p<0.001). A unidimensional factor solution explained 47.45% of the variance in LI-CBT assessment competencies and 54.77% of the variance in LI-CBT treatment competencies. Table 1 illustrates the six items for each measure loading onto single latent variables of LI-CBT assessment and treatment competency.

Discriminant Validity

Table 2 reports the item and total scores for novice, qualified and expert group ratings. Overall, the three groups rated the LI-CBT assessment session significantly differently (F(2,3)=

LI-CBT COMPETENCIES

6.06, *p*=.003) showing that the LIAC discriminated between levels of practitioner experience.

Novices ratings of LI-CBT assessment competencies were significantly higher than the qualified

and expert groups, representing moderate and large effect sizes respectively. Novice, qualified

and expert ratings of the LI-CBT treatment session were not significantly different (F(2,3)= 1.09,

p=.339) showing that the LITC was not able to discriminate practitioners based on their

experience level.

Internal and Interrater Reliability

Table 3 reports the internal consistency across the items for the LIAC and LITC. The

LIAC (α=.77) and the LITC (α=.83) both had good internal reliability for the full scale and

showed good correlations between items (>.3 using the Cristobal et al., 2007 cut off). Guttmann

split-half coefficients showed good internal consistency for the LIAC (*rSHG*= .74) and the LITC

(*rSHG*= .81). Good interrater agreement was found for the full-scale LIAC (ICC(2, 109)= .73,

95% CI .61 to .81) and the full-scale LITC (ICC(2, 147)= .83, 95% CI .78 to .87). The expert

group had excellent agreement on the LIAC (ICC(2,24)= .93, 95% CI .80 to .99) and moderate

agreement on the LITC (ICC (2,24)= .68, 95% CI .21 to .93). The qualified group had excellent

agreement when rating using the LIAC (ICC(2, 59)= .96, 95% CI .91 to .99) and the LITC

(ICC(2, 59)= .76, 95% CI .36 to .96). The novice group had good agreement on the LIAC (ICC

(2,30)= .80, 95% CI .46 to .97) and moderate agreement on the LITC (ICC (2,79)= .64, 95% CI

.06 to .94).

Study Two: Method

Aims and Hypotheses

The aim of this study was to test the construct, criterion, predictive and concurrent

validity of the measures and their internal and test-retest reliabilities in a longitudinal design that

used both expert raters and patient feedback. Specifically, it was hypothesized that expert raters

and patients would agree on LI-CBT competence/incompetence and that the measures would

index changes in competence according to training intervention.

LI-CBT COMPETENCIES

Participants

Participants were trainee LI-CBT practitioners who were evaluated throughout their training course using the LIAC and LITC. The dataset consisted of N=240 LI-CBT assessment sessions and N=217 LI-CBT treatment sessions. Data were collected across four University sites (trainee psychological practitioners at Site 1: n=119; Site 2: n=20; Site 3: n=21; Site 4: n=28).

Design

A longitudinal design enabling the LIAC and LITC to be evaluated over time and in relation to actor feedback, as a proxy for patient feedback.

Procedure

The training methods and national curriculum of LI-CBT practitioners in the UK define the use of observed structured clinical exams (OSCE) to assess performance (UCL, 2015). Trainee LI-CBT practitioners are assessed via OSCEs in their ability to conduct a competent LI-CBT assessment session (45-minute session) and an LI-CBT treatment session (35-minute session). Professional actors play patients during OSCEs and are provided with detailed scripts, process advice and training. For assessment OSCEs, trainees receive a brief referral letter and for treatment OSCEs, trainees received a pack describing the content of previous sessions, sessional outcomes on the Generalised Anxiety Disorder Assessment (GAD-7; Spitzer et al. 2006) and Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer & Williams, 2001) and previous homework assignments, and are then required to conduct the next treatment session. OSCEs are precise, objective and reproducible and allowed for the uniform assessment of clinical competency (Zayyan, 2011).

In University site one, additional data were provided for formative (i.e., practice sessions under exam conditions) and summative (i.e., real exams) OSCE ratings for the LIAC and the LITC, and this was used to index stability in competency measures over time (i.e., test-retest reliability) and in response to a supervisory intervention. Two-weeks following formative OSCEs, trainees then complete a summative assessment and treatment OSCE (i.e. termed

LI-CBT COMPETENCIES

summative 1 in the results). The presenting problem was changed on each occasion for both

assessment and treatment, so that assessment and treatment competencies were being tested over

differing anxiety and depression diagnoses. Practitioners who failed summative assessment or

treatment exams were provided with a one-hour competency enhancement supervision session

prior to the retake. After four-weeks, the retake assessment or treatment OSCE was completed

(i.e. termed summative 2 in the results). Ten percent of sessions at each stage (formative,

summative 1, summative 2) were blind double-rated by n=6 second expert raters. To reduce risk

of bias, expert raters did not rate trainees that they had a relationship with though LI-CBT clinical

supervision/practice groups. Expert raters are trained in the accurate assessment of LI-CBT

practitioners through joint rating of sessions and associated feedback.

Materials

A battery of psychometric measures were used to index the manner in which the

actor/patient experienced the LI-CBT practitioner during summative OSCEs. The *Working*

Alliance Inventory (WAI; Horvath & Greenberg, 1989) is a validated measure of the patient's

experience of the therapeutic alliance, in terms of bond, task and goal subscales. The *Helpful*

Aspects of Therapy Questionnaire (HAT; Llewellyn, 1988) captures and rates aspects of session

that were either *helpful* or *hindering* for the patient. The *Friends and Family Test* (FFT; NHS

England 2014) poses two questions (1) *whether the patient would return* (yes or no) and (2) *the*

likelihood of recommending the service to friends and family (on a scale of 1 = *extremely unlikely*

to 5 = *extremely likely*).

Data Analysis

Sequences of confirmatory factor analysis (CFA) tested the construct validity over time of

the LIAC and LITC (i.e. competency invariance over time; formative = time point 1 and

summative = time point 2) using the full information maximum likelihood method (FIML;

Muthén & Muthén, 2017). Established high intensity CBT competency measures recommend a

minimum cut-off score based upon a score of at least 3 on each item to ensure satisfactory and

LI-CBT COMPETENCIES

consistent levels of competency (CTS; Young & Beck, 1988 and CTS-R; Blackburn et al., 2001).

Using this criterion, a recommended cut-off score of 18 was initially used for the LIAC and LITC

from study two and the samples divided into pass (total competency score ≥ 18) and fail (total

competency score < 18). In order to test the criterion validity of the two measures, Singh's (2007)

method of calculating +/- 2 standard deviations (SDs) from the mean was then used on the LIAC

and LITC to ascertain if a similar score would be found. A sample n=157 competent assessment

sessions and n=135 competent treatment sessions (i.e. positive cases; Singh, 2007) were used to

calculate – 2SDs from that mean. A sample of n=17 incompetent assessment sessions and N=10

incompetent treatment sessions were used to calculate + 2SDs from that mean (i.e. negative

cases; Singh, 2007). The interval obtained by subtracting 2 x SD from mean and by adding 2 x

SD to mean (that is, $\mu \pm 2\sigma$) shows that the chance of an LIAC or LITC rating coming outside

this interval would be less than 5%. The lower limit of this interval (i.e. mean - 2SD) is then

considered as cut-off point for LI-CBT assessment and treatment competency.

To evaluate concurrent validity of the measures, Pearson's correlation coefficients

assessed the relationship between LI-CBT assessment and treatment competencies and the

experience of the patient/actor on the WAI (*bond, goal* and *task* scores), HAT (*helpful* and

hindering scores with the hindering scores inverted) and FFT (*likelihood of recommendation*

scores). Chi-squared tests assessed the relationship between the likelihood of returning (FFT

yes/no) and pass/fail rates to evaluate predictive validity. Area under the curve (AUC) of receiver

operating characteristic (ROC) curves determined sensitivity and specificity of the competency

cut-off scores (pass/fail) to discriminate whether actor/patients would return for another session

(FFT *yes/no*). Level of discrimination was assessed with Hosmer, Lemeshow and Sturdivant's

(2013) AUC classifications; <.5 (no better than chance), 0.5-0.7 (poor), 0.7-0.8 (acceptable), 0.8-

0.9 (excellent) and >0.9 (outstanding).

The internal consistency of the measures in Study Two were determined using the same

methods and guidelines as Study One. One-way random effects intraclass correlation coefficients

LI-CBT COMPETENCIES

(ICC; Shrout & Fleiss, 1979) were calculated for absolute agreement to determine the interrater reliability between double raters across training institutions (interpreted with the same thresholds used in Study 1). Test-retest reliability was examined using paired sample t-tests to assess the sensitivity of the measures in detecting changes in competency in response to a supervision intervention by comparing summative 1 versus summative 2 mean scores.

Results: Study 2

Construct Validity

Assessment and treatment competency ratings were assessed for influential multivariate outliers by calculating Mahalanobis distances. Values greater than 32.910 ($p < 0.001$) or less than 2.214 ($p < 0.001$) were considered outliers; no multivariate outliers were identified on either the LIAC or LITC. Initially, a baseline configural invariance model (with no equality constraints imposed on parameters) was completed to establish whether the same factors and pattern of factor loadings demonstrated adequate model fit over time. The configural invariance model indicated acceptable model fit for both the LIAC ($\chi^2(47)$: 78.275, p<0.05; CFI: 0.945; RMSEA: 0.061; SRMR: 0.068) and LITC ($\chi^2(47)$: 74.878, p<0.05; CFI: 0.948; RMSEA: 0.063; SRMR: 0.064). A second test of invariance (metric invariance) assesses whether the relationships between items of a construct were the same over time by adding constraints to the item factor loadings. The metric invariance model also demonstrated acceptable model fit for both the LIAC ($\chi^2(52)$: 81.21, p<0.05; CFI: 0.948; RMSEA: 0.056; SRMR: 0.081) and LITC ($\chi^2(52)$: 77.933, p<0.05; CFI: 0.952; RMSEA: 0.058; SRMR: 0.075). The third test of invariance (scalar invariance) assessed whether differences on the observed competency items were only due to differences on the common factors by constraining item intercepts over time. Whilst the LITC showed acceptable model fit ($\chi^2(52)$: 88.941, p<0.05; CFI: 0.941; RMSEA: 0.061; SRMR: 0.076), the LIAC showed a significant decrement in model fit (l (change in $\chi^2(5)$ = 23.673, p<0.001; change in CFI = -0.033). Strict invariance testing was not completed due to there being no theoretical justification for the residual variances of observed competency subsection scores being the same across time. The

LI-CBT COMPETENCIES

online supplementary materials report the two-factor model figures. The CFA results overall

demonstrate a two-factor solution showing that the six assessment and treatment competencies

loaded onto a single latent variable at two time points (formative; time point 1 and summative;

time point 2).

Criterion Validity

Employing the method of +/-2SDs from the mean (Singh, 2007), cut-off score estimations

for the LIAC based on passed (n=157; *M* = 23.18, *SD* = 3.07, *range* = 18-32) and failed sessions

(n=17; *M* = 14.18, *SD* = 2.78, *range* = 6-17) indicated cut-off scores of 17.03 (rounded down to

17) and 19.74 (rounded up to 20). Cut-off score estimations for the LITC based on passed (n=135;

M = 23.97, *SD* = 3.47, *range* = 18-31) and failed sessions (n=10; *M* = 14.35, *SD* = 1.93, *range* =

12-17) indicated cut-off scores of 17.03 (rounded down to 17) and 18.21 (rounded down to 18).

Considering patient safety issues and the scoring methods of extant high intensity CBT competency

measures, scores for competent LI-CBT assessment and treatment were held at 18 (i.e. requiring

that each of the six sections for each scale needed to score 3 or more on the Dreyfus rating system

(i.e. demonstrating a 'competent' level for all competency skills).

Concurrent Validity

Table 4 reports correlations between competency ratings and actor/patient feedback (i.e.

WAI, HAT and FFT). Significant positive correlations were demonstrated for the LIAC total

scale and each of the working alliance subscales and working alliance total score. Associations

between treatment competency and the working alliance were more variable. The LITC total

scale score showed significant positive correlations with the working alliance subscales and total

working alliance scores. However, the working alliance total score only correlated with three of

the treatment competency items (*engagement, information gathering,* and *change method)*. Only

the LITC showed a significant correlation with helpful aspects of therapy scores, implying higher

levels of competency during LI-CBT treatment are associated with patient ratings of helpfulness.

Both the LIAC and LITC significantly correlated with the inverted hindrance scale scores. This

LI-CBT COMPETENCIES

implies lower assessment and treatment competency ratings are associated with the actor/patient

experiencing the LI-CBT practitioner as more hindering during sessions. The Helpful Aspects of

Therapy form allows for some qualitative feedback; 28% of competent LI-CBT practitioners

received hindering aspect comments, compared to 73% of incompetent LI-CBT practitioners.

Predictive Validity

The LIAC and LITC total scores and the FFT recommendation scores showed significant

positive correlations. Thirty percent of incompetent LI-CBT practitioners following an

assessment session would be not be recommended, compared to 5% of competent practitioners.

Following an incompetent treatment session, then 21% of LI-CBT practitioners would not be

recommended, versus 4 % of competent practitioners. Actor/patients were four times as likely to

be willing to return for treatment following a competent versus incompetent assessment ($X^2 =$

4.276, *p*=.039, Odds Ratio = 4.0) and were over nine times as likely to be willing to continue with

treatment following a competent versus incompetent treatment session ($X^2 = 4.210$, *p*=.040; Odds

Ratio = 9.5). For assessment sessions, the area under the curve was 0.70 showing that LIAC

classifications of LI-CBT practitioners as competent/incompetent were able to adequately predict

whether actor/patients would return for treatment or not. For treatment sessions, the area under

the curve was 0.93 showing LITC competency classifications were able to predict (at an

outstanding level) whether patients would continue with treatment or not.

Internal, Interrater and Test-Retest Reliability.

Table 5 reports the internal reliability results. The LIAC (α=.83) and the LATC (α=.84)

both demonstrated good internal consistency. Average inter-item correlation coefficients for all

items and total scale scores were correlated (>.3 using the Cristobal et al., 2007 cut off). Item-

total analysis indicated good correlations between the items (>.3). Guttmann split-half

coefficients showed good reliability for the LIAC (*rSHG*= .78) and the LIAC (*rSHG*= .77).

Interrater reliabilities for the full scales was good for the LIAC (ICC = .83; 95% CI .66 to .92)

and moderate for the LITC (ICC = .74; 95% CI .54 to .86). Four of the six LIAC items had good

interrater agreement (ICC range .76 to .82), apart from introducing the session (poor agreement; ICC = .38; 95% CI .12 to .67) and information giving (moderate agreement; ICC = .73; 95% CI .47 to .87). All LITC items showed moderate to good interrater agreement (ICC range .52 to .80), apart from the interpersonal skills item (poor agreement; ICC = .34; 95% CI .01 to .60). Table 6 reports the test-retest reliability results. There was a significant increase in both LI-CBT assessment (n=17; summative 1 M=17.94, SD=3.81; summative 2 M=22.47, SD=2.90; $t(16)$= -5.12 p<.001) and LI-CBT treatment competency in response to one-hour of specific LI-CBT competency supervision (n=15; summative 1 M=19.33, SD=2.77; summative 2 M=23.63, SD=4.76; $t(14)$= -4.33 p=.001).

Discussion

This project developed and then evaluated observational competency measures of the abilities to assess for and treat with LI-CBT. The sample was large compared to previous studies that have developed high intensity CBT competency measures and a wider range of reliability and validity indices were tested (Limon, 2017). Short, six-item and single-factor LI-CBT assessment and treatment competency measures were found. A score of 18 on both measures differentiated competent from incompetent LI-CBT assessment and treatment practice. The LIAC demonstrated good internal consistency, good interrater and test-retest reliability and construct, concurrent, discriminant, criterion and predictive validity. The LITC had good internal consistency and test-retest reliability, moderate-good interrater reliability and demonstrated construct, concurrent, and criterion validity. The LITC however did not demonstrate discriminant validity for distinguishing between the ratings of practitioners with different levels of experience. These results overall provide some initial support for use of the manuals and measures for the governance of the increasing number of services worldwide that are making use of LI-CBT interventions (Singla et al. 2017).

A strength of the initial design of the measures was their basis in behaviour change model (Michie et al. 2011) and so the theoretical foundations of LI-CBT work (UCL, 2015). Both the

measures had good internal consistency and these findings are consistent with high-intensity CBT competency measures in suggesting that the potential overlap in individual items may reflect the integrative nature of the CBT framework (Muse, McManus, Rakovsjik, & Thwaites, 2017). The ICC ranges (across items and total scores) reflects the variation found when assessing interrater reliability of high intensity CBT competency scales (.40-.98, median .65; Loades & Armstong, 2016). The more variable agreement observed for the LITC may also be related to the variable levels of competence displayed during treatment sessions. Von Cronsbruch et al. (2012) found that higher levels of interrater agreement tend to occur when incompetent practice is apparent. The LI-CBT change method item in the LITC manual usefully enables the labelling of when the LI-CBT practitioner incompetently drifts into trying to implement high intensity CBT methods (Waller & Turner, 2016).

It has previously been illustrated that experts in the therapy being evaluated tend to demonstrate higher interrater reliability (Barber, Crits-Christoph & Lubrosky, 1996). A similar pattern was observed in this study, with interrater reliability estimates between expert and qualified LI-CBT practitioners being generally higher than novices. Previous studies have shown that a high level of training is needed to achieve good interrater reliability for high intensity CBT competency rating scales (Barber et al., 2007; Blackburn et al., 2001; Gordon, 2006; Muse et al., 2017). The levels of interrater agreement found here without any intensive training appear therefore promising. In terms of the applied usage of the scales, then the clinical experience of the rater influences the ratings made of assessment sessions, with ratings by novice LI-CBT practitioners likely to be inflated and influenced by cognitive bias (Kruger & Dunning, 1999). Therefore, the self-rating of practice by novice LI-CBT practitioners in clinical services should be supplemented with a supervisor rating for balance.

The LIAC and LITC were both able to index improvements to competency over short time periods and in response to bespoke clinical supervision. Previous studies of high intensity CBT competency measures have also shown significant increases in ratings over training time

LI-CBT COMPETENCIES

(Blackburn et al., 2001; Muse et al, 2017; Liness et al. 2019). Clearly, despite the psychoeducational and guided nature of LI-CBT, a crucial factor is the ability to engage the patient in an effective alliance to enable the psychoeducation to take place. Guided self-help outperforms pure self-help (Pleva & Wade, 2007) and so the alliance appears important in terms of enabling the psychoeducation. A strong relationship was found between competency measures and the *goal* WAI subscale, and this was expected as the LI-CBT approach is particularly goal-focussed. ROC curves demonstrated that the competency of LI-CBT practitioners very accurately predicted the desire to return for treatment following assessment and the desire to stay in treatment. The LIAC was also able to distinguish less experienced practitioners from those with more experience in terms of their competency ratings. This supports Brosnan et al.'s (2008) findings that trainee practitioners are likely to produce over-optimistic ratings of clinical competence. The LITC did not distinguish novice raters from qualified or expert raters, so the discriminant validity of the treatment competency measure could not be fully confirmed. This may be due to the fact that low intensity assessments are very highly structured (and so easier to rate) and LI-CBT treatment sessions are more complex (and so harder to rate), due to the LI-CBT practitioner implementing a change method in these sessions.

The results of the study need to be considered in the light of some methodological limitations. Conclusions are naturally limited to LI-CBT work with mild-to-moderate anxiety and depression presentations. More inductive methods could have informed the item generation process (Hinkin, 1995). Although study one recruited groups with different levels of experience, all raters rated the competence of single exemplar assessment and treatment sessions, and a greater range of sessions could have been rated (Koo & Li, 2016). Study one also utilised ratings from a large number of participants, and corresponding ICCs may have been inflated, as previous studies have showed that interrater reliability decreases when the number of raters is reduced (Karterud et al., 2013). Furthermore, ICCs were higher for incompetent practitioners (von Consbruch et al. 2012) and therefore, the interrater reliability results in study two may not have

provided an accurate assessment of agreement across all levels of practitioner competence.

Similarly, the longitudinal data in study two was collected from a homogenous sample group (i.e. LI-CBT practitioner trainees) and therefore the sensitivity results may not generalise beyond a training context. The OSCE method is not without some pitfalls, such as being labour intensive, costly and logistically difficult to organise (Barman, 2005). The FFT outcome measure has not been psychometrically evaluated and so these results need to be viewed with due caution. The only moderate evidence of interrater reliability and discriminant validity suggests that further refinement and evaluation is of the measures indicated. Suggestions for this include further more detailed elaboration of the competences in the manuals to facilitate accurate ratings (Fairburn & Cooper, 2011), conducting longitudinal studies of qualified LI-CBT practitioners and using groups of novice, qualified and expert raters to assess sessions across a range of apparent competency.

To conclude, accurate assessment of competency is needed to ensure the skilful application of evidenced-based psychotherapy, regardless of the level of treatment intensity (Fairburn & Cooper, 2011). This study has produced measures of the assessment and treatment competencies required in the delivery of LI-CBT to patients with mild-to-moderate anxiety and depression and the measures appear to have reasonable reliability and validity. The measures and manuals can be utilised during supervision to support the ongoing competency development of LI-CBT practitioners. Future clinical trials of LI-CBT will also benefit from employing the measures in their methods, as this is an important aspect of the integrity of trial design. As the number of the LI-CBT practitioners expands worldwide, the current research will aid in the clinical governance of these now widely delivered CBT-based interventions for anxiety and depression.

References

Ali, S., Littlewood, E., McMillan, D., Delgadillo, J., Miranda, A., Croudace, T., & Gilbody, S.

(2014). Heterogeneity in patient-reported outcomes following low-intensity mental health

interventions: A multilevel analysis. *PloS ONE, 9,* e99658.

doi:10.1371/journal.pone.0099658

Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of adherence and competence

on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical

Psychology, 64,* 619-622. doi:10.1037/0022-006X.64.3.619

Barber, J. P., Sharpless, B, A., Klostermann, S., & McCarthy, K, S. (2007). Assessing

intervention competence and its relation to therapy outcome: A selected review derived

from the outcome literature. *Professional Psychology: Research and Practice*, 38, 493-

500. doi:10.1037/0735-7028.38.5.493

Barman A. (2005). Critiques on the Objective Structured Clinical Examination. *Annals of

Academic Medicine Singapore, 34,* 478-482.

Bennett-Levy, J., Richards, D. A., & Farrand, P. (2010). Low intensity CBT interventions: A

revolution in mental health care. In J. Bennett-Levy, D. A. Richards, P. Farrand, H.

Christensen, K. M. Griffiths, D. J. Kavanaugh, . . . C. Williams (Eds.), *Oxford guide to

low intensity CBT interventions* (pp. 3-18). New York, NY, US: Oxford University Press

Blackburn, I. M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F.

K. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties.

Behavioural and Cognitive Psychotherapy, 29, 431-446.

doi:10.1017/S1352465801004040

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness

and efficiency. *The British Journal of Psychiatry, 186* (1), 11-17.

doi:10.1192/bjp.186.1.11

Brosan, L., Reynolds, S., & Moore, R. (2008). Self-evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy, 36*, 581-587. doi:10.1017/S1352465808004438

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*, 431-443. doi:10.1017/S1352465815000491

Consbruch, V.K., Clark, D.M., & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: psychometric properties of the cognitive therapy competence scale for social phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy, 40*, 149-161. doi:10.1017/s1352465811000622

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

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

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

Cuijpers, P. (1997). Bibliotherapy in unipolar depression: a meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 28,* 139-147. doi: 10.1016/s0005-7916(97)00005-0

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*, 449-458. doi: 10.1016/j.cpr.2009.04.005

Delgadillo J. (2018). Guided self-help in a brave new world. *British Journal of Psychiatry*, 212, 65-66. doi: org/10.1192/bjp.2017.17

Delgadillo, J., Rhodes, L., Moreea, O, McMillan, D., Gilbody, S., Leach, C., …Ali, S. (2018).
Relapse and recurrence of common mental health problems after low intensity cognitive
behavioural therapy: The WYLOW longitudinal cohort study. *Psychotherapy and
Psychosomatics, 87,* 116-117. doi:10.1159/000485386

Dreyfus, H. L. (1989). The Dreyfus model of skill acquisition. *In* J. Burke (Ed.) *Competency Based
Education and Training.* London: Falmer Press.

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

Fairburn, G., & Patel, V. (2014). The global dissemination of psychological treatments: A road
map for research and practice. *American Journal of Psychiatry, 171,* 495-498. doi:
10.1176/appi.ajp.2013.13111546.

Farrand, P., & Woodford, J. (2013). Impact of support on the effectiveness of written cognitive
behavioural self-help: a systematic review and meta-analysis of randomised controlled
trials. *Clinical Psychology Review, 33*, 182-95. doi: 10.1007/s12160-015-9689-0

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. doi: 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*, 747–752. doi: 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, 343. doi:10.1037/pas0000372

Green, H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT
Psychological Wellbeing Practitioners (PWPs): A multilevel modelling and mixed

methods analysis. *Behaviour Research and Therapy, 63*. 43-54. doi:

10.1016/j.brat.2014.08.009

Hallgren K. A. (2012). Computing inter-rater reliability for observational data: An overview and

tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34.

Horvath, A. O., & Greenberg, L. S. (1986). Development and validation of the Working Alliance

Inventory. *Journal of Counselling Psychology, 36,* 223-233. doi:10.1037/0022-

0167.36.2.223

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd

ed.). Hoboken, NJ: Wiley.

Hinkin T.R. (1995). A review of scale development practices in the study of organizations. *Journal

of Management, 21,* 967–988. doi: 10.1016/0149-2063(95)90050-0

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., et al (2017). Efficacy

of self-guided internet-based cognitive behavioral therapy in the treatment of depressive

symptoms: a meta-analysis of individual participant data. *JAMA Psychiatry, 74*, 351-359.

doi: 10.1001/jamapsychiatry.2017.0044.

Karterud, S., Pedersen, G., Johansen, M. S., Johansen, P., Schlüter, C., Engen, M., … Bateman, A.

(2013). The mentalization-based treatment adherence and competence scale (MBT-ACS):

Development, structure and reliability. *Psychotherapy Research*, *23,* 674–98.

doi:10.1080/10503307.2012.708795.

Kohrt, B., Jordans, M., Rai, S., Shrestha, P., Luitel, N. P., Ramaiya, M., …Patel, V. (2015).

Therapist competence in global mental health: Development of the Enhancing Assessment

of Common Therapeutic factors (ENACT) rating scale. *Behaviour Research and Therapy,

69,* 11-21. doi:10.1016/j.brat.2015.03.009

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.

doi: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*, 606–613. doi:10.1046/j.1525-1497.2001.016009606.x

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77***,** 1121-1134. doi: 10.1037//0022-3514.77.6.1121

Limon, E. (2017). Measures of clinical competency of the delivery of cognitive behaviour therapy; Evidence of absence? MSc thesis, University of Sheffield.

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

Llewelyn, S. (1988). Psychological therapy as viewed by clients and therapists. *British Journal of Clinical Psychology*, 27, 223-237. doi:10.1111/j.2044-8260.1988.tb00779.x

Loades, M. E., & Armstrong, P. (2016). The challenge of training supervisors to use direct assessments of clinical competence in CBT consistently : A systematic review and exploratory training study. *The Cognitive Behaviour Therapist, 9,* e27. doi:101.1017/S1754470X15000288

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research*, *35*, 382-386. doi:10.1097/00006199-198611000-00017.

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science, 6:* 42. doi:10.1186/1748-5908-6-42

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*, 542-555. doi:10.1037/pas0000372

Muthén, L.K., & Muthén, B.O. (2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA:

Muthén & Muthén.

NHS England. (2014). Friends and Family Test. Retrieved from: https://www.england.nhs.uk/wp-

content/uploads/2014/07/fft-imp-guid-14.pdf

Polit, D., & Beck, C. (2006). The content validity index: Are you sure you know what's being

reported? Critique and recommendations. *Research in Nursing Health*, *29*, 489-497.

doi:10.1002/nur.20147

Pollecoff, M. (2016). *Trends in psychotherapy in the UK.* UK: United Kingdom Council for

Psychotherapy.

Pleva, J. & Wade, T.D. (2007). Guided self-help versus pure self-help for perfectionism: a

randomised controlled trial. *Behaviour Research and Therapy, 45,* 849-61. doi:

10.1016/j.brat.2006.08.009

Putka, D.J., Le, H., McCloy, R.A., & Diaz, T. (2008). Ill-structured measurement designs in

organizational research: Implications for estimating interrater reliability. *Journal of

Applied Psychology, 93*, 959–981. doi: 10.1037/0021-9010.93.5.959

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.* (2nd Ed.). Rethink, UK.

Roth, A., & Pilling, S. (2007). Using an evidence-based methodology to identify the competences

required to deliver effective cognitive and behavioural therapy for depression and anxiety

disorders. *Behavioural and Cognitive Psychotherapy*, *36*, 129-147.

doi:10.1017/s1352465808004141

Singh, G. (2007). Determination of cutoff score for a diagnostic test. *Internet Journal of

Laboratory Medicine, 2,* 1-4.

Singla, D.R., Kohrt, B.A. Murray, L.K., Anand, A., Chorpita, B.F. Patel, V. (2017).

Psychological treatments for the world: lessons from low- and middle-income countries.

LI-CBT COMPETENCIES

Annual Review of Clinical Psychology, 13, 149-181. doi: 10.1146/annurev-clinpsy-032816-045217

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428. doi:10.1037/0033-2909.86.2.420

Southam-Gerow, M. A. & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: Introduction to Special Issue. *Clinical Psychology: Science & Practice, 20*, 1-13. doi:10.1111/cpsp.12019

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

Turpin, G., Clarke, J., Duffy, R., & Hope, R. (2009). A new workforce to deliver IAPT: A case study. *Journal of Mental Health Training, Education and Practice, 4,* 37-46. doi:10.1108/17556228200900017

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

von Consbruch, K., Clark, D. M., & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: Psychometric properties of the Cognitive Therapy Competence Scale for Social Phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy, 40*, 149-161. doi:10.1017/S1352465811000622

Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy*, 77, 129-137. doi: 10.1016/j.brat.2015.12.005

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,* 200-211. doi:10.1037/a0018912.

Young, J. E., & Beck, A. T. (1988). Cognitive Therapy Scale. Unpublished Manuscript,

University of Pennsylvania, Philadelphia, PA.

Zarafonitis-Müller, S., Kuhr, K., & Bechdolf, A. (2014). The relationship between therapist's

competence and adherence to outcome in cognitive-behavioural therapy - results of a

meta-analysis. *Fortschritte der Neurologie Psychiatrie, 82*, 502–510.

Zayyan, M. (2011). Objective structured clinical examination: the assessment of choice. *Oman

Medical Journal*, *26*, 219–222. doi:10.5001/omj.2011.55

LI-CBT COMPETENCIES

Table 1 *Exploratory factor analysis factor loadings (study one)*

Low intensity assessment competencies (LIAC)	Loading
Introduction to the assessment	0.540
Initial engagement	0.836
Interpersonal skills	0.822
Information gathering relevant to the problem	0.763
Information giving relevant to the problem	0.725
Shared planning and decision making	0.548
Low intensity treatment competencies (LITC)	Loading
Focusing the treatment session	0.728
Maintaining engagement	0.827
Interpersonal skills	0.768
Information gathering specific to change	0.740
LI change method	0.651
Homework planning	0.694

LI-CBT COMPETENCIES

Table 2 *Mean scores and ANOVA results for expert, qualified and novice raters (study one)*

	Groups			Between-group *d*			*F* (df=2)	*p*	Tukey post-hoc
	Expert raters	Qualified raters	Novice raters	E v Q	E v N	Q v N			
LIAC - assessment	(*n*=24)	(*n*=59)	(*n*=30)						
Introduction	3.46 (0.49)	3.93 (0.71)	4.17 (0.59)	-0.72	-1.30	-0.36	8.46	<.000*	E < Q, N
Engagement	2.79 (0.86)	3.21 (0.71)	3.33 (0.69)	-0.56	-0.70	-0.17	3.92	.023*	E < N
Interpersonal	3.10 (0.90)	2.80 (0.73)	3.15 (0.66)	0.38	-0.07	-0.50	2.65	.076	-
Information gathering	3.40 (0.78)	3.29 (0.58)	3.58 (0.59)	0.17	-0.27	-0.50	2.08	.130	-
Information giving	3.38 (0.70)	3.16 (0.75)	3.72 (0.65)	0.30	-0.51	-0.78	6.10	.003*	N > Q
Shared planning	2.83 (0.88)	2.92 (0.92)	3.53 (0.76)	-0.10	-0.86	-0.70	5.95	.004*	N > Q, E
Total LIAC score	*18.96 (2.28)*	*19.03 (3.91)*	*21.48 (2.77)*	*-0.02*	*-0.98*	*-0.69*	*6.06*	*.003*	*N > Q, E*
LITC - treatment	(*n*=24)	(*n*=59)	(*n*=79)						
Focusing session	3.94 (0.65)	3.92 (0.70)	3.80 (0.60)	0.03	0.23	0.19	0.74	.481	-
Engagement	3.71 (0.95)	3.85 (0.70)	3.68 (0.54)	-0.18	0.05	0.28	1.06	.350	-
Interpersonal	3.96 (0.83)	3.82 (0.88)	3.64 (0.79)	0.16	0.40	0.22	1.64	.198	-
Information gathering	3.85 (0.83)	3.93 (0.62)	3.89 (0.58)	-0.12	-0.06	0.07	0.13	.882	-
LI change method	3.73 (0.86)	3.95 (0.75)	3.81 (0.68)	-0.28	-0.11	0.20	0.91	.407	-
Homework planning	3.96 (0.62)	4.28 (0.65)	3.83 (0.73)	-0.50	0.18	0.64	6.63	.002*	Q > E
Total LITC score	*22.50 (3.90)*	*23.43 (3.67)*	*22.63 (2.82)*	*-0.25*	*-0.04*	*0.25*	*1.09*	*.339*	-

*Note. * p<.05*

N = novice rater group; Q = qualified rater group; E = expert rater group; d = between-groups Cohen's d effect size.

LI-CBT COMPETENCIES

Table 3 *Reliability of the LIAC and LITC (study one)*

Competency items	Item-total (if deleted)	Cronbach alpha (if deleted)
Low intensity assessment competency (LIAC)	(*n*=113)	
Introduction	.40	.77
Engagement	.57	.73
Interpersonal	.59	.72
Information gathering	.51	.75
Information giving	.56	.73
Shared planning	.52	.75
Low intensity treatment competency (LITC)	(*n*=162)	
Focusing session	.43	.84
Engagement	.66	.80
Interpersonal	.72	.78
Information gathering	.58	.81
LI change method	.58	.81
Homework planning	.68	.79

Table 4 *Associations between competency scale scores and actor/patient completed measures (study two)*

	Working Alliance Inventory				Helpful Aspects of Therapy		Recommendation
	Task	Bond	Goal	*Total*	Helpful	Hinderance[a]	
LIAC (*n*=123)							
Introduction	.33 (.06)	.34 (.05)*	.34 (.05)*	.36 (.04)*	-	-	.32 (.00)**
Engagement	.47 (.01)**	.43 (.01)**	.62 (.00)**	.54 (.00)**	-	-	.47 (.00)**
Interpersonal	.52 (.00)**	.51 (.00)**	.51 (.00)**	.54 (.00)**	-	-	.40 (.00)**
Information gathering	.52 (.00)**	.48 (.00)**	.64 (.00)**	.58 (.00)**	-	-	.43 (.00)**
Information giving	.67 (.00)**	.60 (.00)**	.56 (.00)**	.64 (.00)**	-	-	.17 (.10)
Shared planning	.49 (.00)**	.33 (.06)	.47 (.00)**	.46 (.00)**	-	-	.13 (.22)
Total scale score	.66 (.00)**	.57 (.00)**	.69 (.00)**	.67 (.00)**	.29 (.11)	.49 (.01)**	.35 (.00)**
LITC (*n*=94)							
Focusing session	.17 (.34)	.06 (.74)	.15 (.40)	.08 (.64)	-	-	.28 (.06)
Engagement	.47 (.01)**	.46 (.00)**	.41 (.02)*	.42 (.02)*	-	-	.47 (.00)**
Interpersonal	.22 (.23)	.26 (.15)	.24 (.18)	.17 (.37)	-	-	.39 (.01)**
Info gathering	.34 (.06)	.35 (.05)*	.31 (.09)	.36 (.04)*	-	-	.35 (.02)*
LI change method	.66 (.00)**	.61 (.00)**	.64 (.00)**	.65 (.00)**	-	-	.66 (.00)**
Homework planning	.28 (.11)	.22 (.22)	.27 (.14)	.24 (.20)	-	-	.31(.04)*
Total scale score	.51 (.00)**	.47 (.01)**	.49 (.00)**	.46 (.00)**	.69 (.00)**	.48 (.01)**	.55 (.00)**

Note. *= *p*<.05, **= *p*<.01; [a] hinderance scores have been inverted so a positive correlation represents increased competency associated with a lower hindering rating

Table 5 *Reliability of the LIAC and LITC ratings (study two)*

Competency items	Item-total (if deleted)	Cronbach alpha (if deleted)
Low intensity assessment competency (LIAC)	(*n*=240)	
Introduction	.50	.82
Engagement	.70	.78
Interpersonal	.68	.78
Information gathering	.62	.80
Information giving	.61	.80
Shared planning	.53	.82
Low intensity treatment competency (LITC)	(*n*=217)	
Focusing session	.53	.83
Engagement	.68	.81
Interpersonal	.67	.81
Information gathering	.65	.81
LI change method	.61	.82
Homework planning	.60	.82

Table 6 *Item and total competency scale score means (SD) over time (study two)*

LIAC item and total scale	Formative ratings	Summative 1 ratings	Summative 2 ratings
Introduction	4.02 (.66)	4.27 (.84)	4.55 (.71)
Engagement	3.61 (.66)	3.46 (.79)	3.75 (.62)
Interpersonal	3.61 (.70)	3.84 (.89)	3.91 (.73)
Information gathering	3.44 (.68)	3.72 (.79)	3.70 (.55)
Information giving	3.53 (.65)	3.54 (.82)	3.77 (.73)
Shared planning	3.38 (.75)	3.18 (1.02)	3.52 (.67)
Total LIAC scale score	*21.26 (3.18)*	*22.31 (3.89)*	*23.13 (3.08)*
LITC items and total scale	Formative ratings	Summative 1 Ratings	Summative 2 ratings
Focusing session	4.61 (.68)	4.31 (.88)	5.03 (1.16)
Engagement	3.88 (.68)	3.83 (.83)	3.63 (.83)
Interpersonal	4.18 (.70)	4.04 (.78)	4.00 (.82)
Information gathering	4.07 (.59)	3.92 (.83)	3.69 (.86)
LI change method	4.07 (.68)	3.72 (1.07)	3.56 (.98)
Homework planning	3.78 (.75)	3.47 (.95)	3.81 (1.12)
Total LITC scale score	*24.51 (2.98)*	*23.27 (4.19)*	*23.72 (4.62)*