Published in final edited form as:

J Behav Cogn Ther. 2020 December ; 30(4): 253–266. doi:10.1016/j.jbct.2020.10.001.

Assessing health worker competence to deliver a brief psychological treatment for depression: development and validation of a scalable measure

Juliana L. Restivo^{1,*}, Lauren Mitchell^{1,*}, Udit Joshi^{2,*}, Aditya Anand², P. Cristian Gugiu³, Daisy R. Singla⁴, Steven D. Hollon⁵, Vikram Patel^{1,6}, John A. Naslund^{1,†}, Zafra Cooper^{7,#}

¹Department of Global Health and Social Medicine, Harvard Medical School, 641 Huntington Ave. Boston, USA 02446

²Sangath, 120, Deepak Society, Chuna Bhatti, Bhopal Madhya Pradesh, India - 462016

³Clinical Outcomes Solutions, 53 West Jackson Boulevard, Suite 1150, Chicago, IL, USA 60604

⁴Department of Psychiatry, University of Toronto and Sinai Health, Mount Sinai Hospital, 600 University Ave, 914A, Toronto, ON, M5G 1X5

⁵Department of Psychology, Vanderbilt University, PMB 407817, 2301 Vanderbilt Place Nashville, TN 37240-7817 USA

⁶Department of Global Health and Population, Harvard T.H. Chan School of Public Health, 655 Huntington Ave, Boston, MA 02115, USA

⁷Department of Psychiatry, Yale School of Medicine, 300 George St #901, New Haven, CT 06511, USA

Abstract

Increased interest in disseminating and implementing psychological treatments has focused on the need for evidence-based training programs for providers, especially those without specialized training. To evaluate provider-training programs, validated outcome measures are necessary; however, the scalable measurement of training outcomes has been largely overlooked. Current methods of assessing providers' ability to deliver psychological treatments are generally time-consuming and costly, representing a major bottleneck in scaling up mental health care for commonly occurring disorders such as depression. The present study describes the development and initial validation of a scalable measure for assessing provider competence in delivering a brief behavioral activation treatment for depression, called the Healthy Activity Program, adapted for primary care settings. The measure focuses on testing knowledge about the treatment and applied knowledge regarding how to skillfully deliver the treatment, both essential features of competence. The measure was tested on a sample of 531 respondents with a variety of educational levels and

Corresponding Author: Juliana Restivo, Department of Global Health and Social Medicine, Harvard Medical School, 641 Huntington Ave., Boston, Massachusetts, USA 02115, juliana_restivo@hms.harvard.edu.

*Joint first authors

†Co-mentor author

#Mentor author

Declarations of interest: none

professional backgrounds and found to meet the requirements of the Rasch model. Three versions of the measure each of equal difficulty were derived to allow repeat testing of training outcomes over time. A scalable measure of provider competence is an essential first step towards supporting the wider dissemination and implementation of brief psychological interventions for depression, especially in low-resource settings.

Keywords

depression; behavioral activation; therapist competence; psychometrics; Rasch Analysis; training

Introduction:

There has been increasing interest in disseminating and implementing psychological treatments for depression, with particular attention to training non-specialist providers without previous mental health training to deliver these treatments in routine care settings (Singla et al., 2017). Referred to as task sharing, this approach holds potential for bridging gaps in available mental health care in both high-income and lower-income countries (Raviola et al., 2019). With health systems seeking to train non-specialist providers to deliver evidence-based psychological treatments, validated measures to assess the outcome of provider training will be necessary; however, the measurement of training outcomes has been largely overlooked (Ottman et al., 2020; Rakovshik & McManus, 2010; Rosen et al., 2017).

It has been suggested in recent discussions that the appropriate direct outcome of provider training programs is provider competence, defined as: “*the ability of the provider to deliver the treatment to the standard required to achieve its expected effect*” (Fairburn & Cooper, 2011). More specifically a competent provider needs to have the required knowledge to deliver the treatment skillfully. In this context it is necessary to assess providers’ *theoretical knowledge* about the treatment (e.g., knowing how many sessions the treatment involves, knowing what topics are covered during each session, such as activity scheduling, etc.) and their *applied knowledge of how and when to use the treatment and its strategies and procedures skillfully* (e.g., knowing how to establish a collaborative relationship, when and how to assess activation targets, etc.) (Cooper et al., 2015). Provider competence is important for a variety of reasons, not least because it is associated with improved patient outcomes (Branson et al., 2015; Ginzburg et al., 2012; Haug et al., 2016; Strunk et al., 2010).

Although there has been relatively little systematic research on the assessment of provider competence, several assessment methods have been previously employed. Many training studies have used knowledge tests to assess outcomes (Frank et al., 2020; Richmond et al., 2017; Singh & Reyes-Portillo, 2020). The majority of these knowledge tests have been devised for particular studies and as such are study-specific and lack established psychometric properties (Lewis et al., 2015; Martinez et al., 2014). Competence has also been assessed with performance based assessments such as standardized role plays (Muse & McManus, 2013), where providers’ abilities are observed in controlled conditions and rated

by experts. A recent review of the use of standardized role plays (Ottman et al., 2020) found that the majority of studies that utilized this method did not adequately validate the role play assessments. Furthermore, the extensive financial and logistical resources required to use these methods limits their large-scale implementation, especially in low-resource settings (Cooper et al., 2017; Kaslow et al., 2009). Such limitations in assessing the competence of health care workers to deliver brief psychological treatments contribute to the barriers to making these essential treatments more widely available.

The present study is part of a larger program of research aimed at scaling up the delivery of a brief, behavioral activation treatment for depression in adults called the Healthy Activity Program (HAP) (Patel et al., 2017). This intervention has previously been shown to be effective for treating depression when delivered by lay counselors in two different low-resource settings (Chowdhary et al., 2016; Jordans et al., 2019; Patel et al., 2017; Weobong et al., 2017). Delivery of psychological treatments such as HAP by non-specialist providers, a form of “task sharing,” has furthermore been shown to be effective in increasing access to evidence-based treatments for a range of disorders across diverse settings in both low-income and higher-income countries (Hoeft et al., 2018; Singla et al., 2017). If such psychological treatments are to be disseminated and implemented across a wide range of settings, scalable methods of training and assessing provider competence are required (Kemp et al., 2019). This is especially relevant in low resource settings. In such settings significant cost barriers and the limited availability of personnel with expertise in mental health care make it difficult to assess provider competence using existing costly and labor-intensive methods.

The present study describes the development of a brief, scalable measure of competence for assessing non-specialist providers’ ability to deliver the HAP treatment for depression together with validity data to support its use. The primary aims were to construct a measure with adequate content validity and response process integrity and to report internal validity data derived from Rasch modelling. Previously-used methods that demonstrated that it was possible to develop a brief, valid, scalable measure to assess clinician competence to deliver enhanced cognitive behavior therapy for eating disorders (CBT-E) (Cooper et al., 2015, 2017) served as a guide for the current work.

Methods:

Design

This study was conducted in two stages following the methods described by Cooper and colleagues (Cooper et al., 2015). These stages covering the development of the items and their evaluation respectively, are described below and illustrated in Figure 1. The Harvard Medical School Office of Human Research Administration determined that this research is classified using Harvard’s Data Security Policy as Level 0 data and therefore is not research as defined by the DHHS regulations (IRB19-0857).

Stage 1: Development of Items

Definition of Content of Measure: The first step to ensure adequate content validity was the development of a blueprint to specify the theoretical and applied knowledge required to deliver HAP. This provided the basis for the generation of an initial bank of assessment items. HAP is typically delivered over 6-8 sessions and covers both treatment specific and general counseling skills (Singla et al., 2014), which are important but often overlooked when training non-specialist providers in low-resource settings (Kohrt et al., 2015). Treatment specific elements include psychoeducation, behavioral assessment, activity monitoring, activity structuring and scheduling, and problem-solving (Chowdhary et al., 2016). General counselling skills include using a collaborative style, asking open-ended questions, and demonstrating empathy, warmth, and genuineness. The source material for the development of the blueprint was provided by the Program for Effective Mental Health Interventions in Under-resourced Health Systems (PREMIUM) HAP manual and the Counseling Relationship Manual (Chowdhary et al., 2016; Patel et al., 2014, 2017). These manuals are available open source [<http://www.sangath.in/evidence-based-intervention-manuals/>]. Table 1 summarizes the stages of HAP.

The initial blueprint was developed under the guidance of senior members of the research team. HAP content was organized into a list of 29 competencies. To ensure that content was adequately and comprehensively covered, the initial blueprint was reviewed by 9 experts, both those who developed behavioral activation (Dimidjian et al., 2006) and HAP (Patel et al., 2017) and those who delivered HAP in previous trials (Patel et al., 2017). These experts also advised on the relative importance of each area, by indicating which competencies were more important than others, based on their clinical experience. After expert review, the final blueprint (see Table 2) covered 24 areas of competence to be assessed by the measure.

Item Generation: Items were constructed with the goal of adequately covering the competencies in the blueprint. In addition, to ensure validity, attention was paid to the integrity of the response process by constructing items in accordance with guidelines for writing test questions in the medical sciences (Case & Donahue, 2008; Case & Swanson, 2002). All items were designed as one-best-answer multiple-choice items, which consisted of a stem and a lead-in question followed by four answer options. Items focused predominantly on assessing applied knowledge. Competencies deemed by the experts to be the most important were represented by 2-3 items while others by 1-2 items. To assess applied knowledge, item stems included short clinical vignettes presenting real-world scenarios. In all cases only items with one clearly agreed best answer were retained. To create plausible but incorrect answer options to function as distractors, items were derived from a list of common errors that providers often make in delivering HAP. These common errors were collected from a group of experienced HAP counselors.

Sample items to provide an illustration of the structure and content of the questions are available in the Supplementary Material A. The full set of items may be requested by contacting the corresponding author.

An initial sample of 43 items was informally tested with 15 individuals with no training or experience in mental health care and three experienced HAP counselors to make a

preliminary assessment of how the items functioned and their level of difficulty. Items that were correctly answered by those with no training in mental health care were deemed to be vulnerable to guessing or sophisticated test taking strategies. In addition, HAP counselors provided further comments on the clarity of the items, the HAP competencies addressed, and accuracy of answer choices. Experience gained in refining these items was used to inform the generation of further items.

Senior team members oversaw the item generation process with all items being reviewed and edited through multiple iterations. To avoid “teaching to the test” (Popham, 2001), a form of bias whereby instruction of the course materials is organized around the actual items on the final test, none of the individuals who were involved in the development of the HAP training materials had access to or contributed to the development of the test items.

In total, 109 items were developed. Further informal testing and review resulted in the exclusion of 11 items that either had one or more of the following problems: lack of clarity; no one agreed upon the correct answer; or contravention of item writing guidelines (Case & Swanson, 2002). This resulted in 98 items remaining for further testing. These 98 items were split into three broadly equivalent test versions with each version covering the full range of HAP content identified in the blueprint. Each test contained both unique (29, 30 and 30 items respectively) and 9 common anchor items. These common anchor items were included so that the tests could be subsequently linked and calibrated on a common scale. As there was no a priori reason for some items to have greater weight than others it was decided to use a binary scoring system with each item contributing equally to the total score. The total score was calculated as the sum of all correct responses. The three individual versions were formatted to be administered via an online survey link using Qualtrics (Qualtrics XM Platform, 2019). The Qualtrics platform allows scores to be calculated automatically.

Stage 2: Evaluation Stage

Recruitment of Participants: Survey participants were recruited through network referrals, email newsletters, social media, and Qualtrics panel surveys (Online Panels: Get Responses for Surveys & Research | Qualtrics, 2019). The target population included a diverse sample of individuals over the age of 18 with varying degrees of education and training in mental health care. Included amongst those with mental health training were those for whom training in a brief evidence based intervention for depression would be relevant and people who already had some training in interventions like BA or CBT, that share common elements with HAP. In addition, there were individuals who indicated that, much like the non-specialist providers to be trained in task sharing, they had no previous mental health experience. This group was selected to reflect the wide range of experiences of those who might be trained in HAP.

After consenting to complete the online survey, participants answered basic demographic questions, although no personally identifiable information was collected. Respondents were then randomized to receive one of the three versions of the measure. Respondents were stratified on the basis of their exposure to balance experience levels. To ensure the integrity of response data and avoid using data from respondents that did not actually read the questions, only survey responses that took more than 18 minutes to complete were

considered eligible for scoring. This minimum duration was determined by measuring the total time required for members of the research team to read through and answer the full test. The survey could only be submitted by participants if responses were provided for all items; therefore, only fully completed surveys could be retained for analysis.

Data Analysis using Rasch methodology: Item responses were analyzed to obtain a pool of calibrated items that fit the Rasch model (Boone, 2016). These items were then used to create three versions of the competence measure of equal difficulty. The Rasch model is a psychometric method for estimating latent ability by studying item responses (Bond & Fox, 2015; Boone et al., 2014; Rasch, 1960; Wright & Stone, 1979). The model examines several aspects of internal construct validity of a set of items (Hagquist et al., 2009; Horton & Perry, 2016; Perry & Horton, 2020; Tennant & Conaghan, 2007). These include but are not limited to: unidimensionality (items all contributing to measuring the same construct); response dependence (responses to an item having a direct impact on responses to other items); and scale targeting (the relative distribution of item locations and person locations on the same underlying continuum) (Perry & Horton, 2020). The probability of a person succeeding on an item is modelled as a logistic function of person ability (the number of correct responses by a person) and item difficulty (the number of correct responses to an item). The model enables estimates of item difficulty and person ability to be measured on a common-equal-interval logit scale. A primary advantage of the Rasch model is that person and item parameters are estimated independently from each other (Andrich, 2004; Tinsley & Dawis, 1977; B. Wright, 1968, 2005). The dichotomous Rasch model was fitted within Winsteps version 4.4.6 software (Linacre, 2019) to determine overall fit of the data to a unidimensional model and to test for local item independence. Infit and Outfit MNSQ statistics were used to examine item fit and remove misfitting items with fit statistics outside the range of 0.7 to 1.3, inclusive being identified as misfits (Smith et al., 2008).

In addition, to keep the measure relatively short for scalability purposes, further items were removed in a series of steps (see Figure 2). Items were removed if they were identified as being psychometrically redundant (providing the same amount of information about a candidate's latent ability regardless of content) on the Wright map (Wright & Stone, 1979). Items were excluded in a stepwise and cumulative fashion until there were no misfitting items. The overall fit of the retained items to a unidimensional model was assessed using Principal Component Analysis (PCA) of standardized residuals. During each step attention was paid to check that removing any psychometrically redundant items did not adversely impact covering the competencies specified by the blueprint. These items were also checked for local independence. In addition, item and person separation statistics were calculated.

Results:

Test Completion

In total, 531 individuals completed one of the three versions of the assessment measure, with the numbers completing these tests being 169 (31.8%), 193 (36.3%) and 169 (31.8%) respectively.

Participant Characteristics

The mean age of the respondents was 45.4 (SD=16.27). 258 (49%) identified as male, 271 (51%) as female and two (<1%) indicated they preferred not to answer. The highest level of education for respondents was as follows: less than high school/secondary, 11 (2%); high school graduate, diploma or equivalent, 82 (15%); some college or certificate program, 133 (25%); bachelor's degree or equivalent undergraduate degree 150 (28%); master's degree 101 (19%); PhD or doctoral equivalent 38 (7%); medical degree (MD) or equivalent 16 (3%).

Rasch Analysis

An initial examination of the infit and outfit statistics revealed that of the 98 items evaluated, there were 9 misfitting items, 8 of which were unique items and 1 was a common anchor item. These misfitting items were removed. During the iterative process of removing redundant items described above, a further 29 items were removed. In total, 38 items were removed of which 11 were misfitting items and 27 were psychometrically sound, well performing items.

The resulting 60 items with no misfit met the criteria of unidimensionality with 1.9 EI units (2.6%) of the unexplained variance belonging to the first contrast, 1.68 EI units (2.3%) to the second contrast, and 1.64 EI units (2.2%) to the third contrast. There were no items with values greater than ± 0.30 in the correlation matrix for the items with the highest co-dependency after controlling for person ability, confirming that the assumption of local item independence required for Rasch analysis was met.

Figure 4 shows the item-person map of the retained items. With the item difficulty mean set at 0 logit, the person ability estimate mean was -0.41 logits with a standard deviation of .81 logits, indicating that the distribution of person measures (ability) is nearly identical to that of item measures (difficulty), thereby confirming that the items are well-targeted, neither too hard or too easy for the population. There were no ceiling or floor effects nor any significant gaps in the measurement of ability along the continuum.

Item reliability was 0.95 confirming the item difficulty hierarchy, that is, the internal construct validity of the measure. Person reliability was less good at 0.65 indicating that there were likely not enough items in the scale to very precisely differentiate among participants.

The 60 items retained after analysis consisted of three tests each containing 26 items, 17 unique items and 9 common items, with each test covering the full set of competencies outlined in the blueprint. A comparison of these three tests is provided by the Test Characteristic Curve (TCC) (see Figure 5) that links raw scores on each of the three 26-item tests with person ability (i.e., a common scale). It is therefore possible to estimate the ability level of a participant irrespective of the test taken. The TCC of the three tests did not show significant difference in their difficulty level.

Discussion

This study describes the development and initial validation of a brief scalable measure to assess the competence of providers to deliver HAP. The measure focuses on testing knowledge of the treatment as well as applied knowledge of how to deliver the treatment skillfully, an essential aspect of provider competence. Three versions of the measure with equivalent difficulty levels were developed to allow repeat testing of training outcomes over time (e.g., before training, after training, and at periods thereafter). As previously noted, expanding the delivery of evidence-based mental health care requires scalable methods of assessing provider competence as well as scalable methods of training (Cooper et al., 2015; Kohrt et al., 2020; Raviola et al., 2019; Singla et al., 2018).

In developing this measure of provider competence, a rigorous stepwise approach to test construction and validation was followed to achieve three psychometrically sound tests of 26 items of equivalent difficulty. The difficulty level of the test items was well matched to the ability level of participants with the items neither too hard nor too easy for the population. There were no ceiling or floor effects, nor any significant gaps in the measurement of ability along the continuum. As expected with a mixed population, the items have scope to identify greater ability on the measure than that shown by the present participants and thus to identify increased knowledge as a result of training. There were 27 additional items with sound psychometric properties that were removed to keep the measure brief. These additional items could be added to future iterations of the measure to improve the person reliability estimates. However, this would be achieved at the expense of making it longer, requiring additional time to administer and complete. A longer measure might potentially limit scalability.

This study extends the work done previously on the development of a scalable assessment of therapist competence for clinicians delivering CBT-E for eating disorders, a highly specialized psychological treatment (Cooper et al., 2015). Importantly, the current study shows that the well-established methods of test construction and Rasch analysis used for measuring specialist clinician competence described by Cooper et al can be further generalized to develop a scalable measure suitable for testing the competence of non-specialist providers delivering a brief psychological treatment for depression in primary care settings.

This study has several strengths. To ensure the validity of the measure, careful attention was paid to following the US standards for educational and psychological testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, 1999). This included the development of a blueprint, training in, and subsequent adherence to, item writing guidelines described by Case and Swanson (2002), and iterative independent review from clinicians and researchers with expertise in HAP and its ongoing delivery in primary care settings. Items were tested on a relatively large sample of participants with varying levels of education similar to those who might eventually be trained and assessed. The use of the Rasch model was a further strength. Rasch analysis focuses on the item level and lends itself to establishing a calibrated item pool from which many potential tests could be derived

in the future. These tests can be designed to closely match the ability of the group being tested in an adaptive fashion. In contrast, many knowledge tests used in the training literature to assess outcome have not been validated and their psychometric properties are unknown.

Some limitations and areas for future work should also be noted. While validity from a variety of sources has been reported, further work to establish discriminant validity by comparing a sufficiently large sample of individuals specifically trained in HAP with those not yet trained, would greatly strengthen support for the measure. In addition, further support would be provided by establishing that those who achieve high scores on the measure also score highly when their performance is observed and rated by HAP experts. This work is currently underway. It is perhaps worth noting in this regard that the brief scalable measure of therapist competence developed by Cooper and colleagues was able to efficiently predict therapist performance on a performance or skills based role play with a positive predictive value of 77% and a specificity of 78% (Cooper et al., 2017). Finally, establishing that HAP providers who score highly on the measure also achieve higher rates of improved patient outcomes would also provide further support.

Assessing competence using the measure developed in this study is likely to be more scalable than conventional methods of assessing provider competence. However, while the current measure assesses competence by assessing theoretical knowledge of HAP and applied knowledge of how to apply HAP skillfully, it does not assess trainees' actual performance at implementing the intervention, nor the ongoing quality of the treatment they provide. It may be argued that data on performance skill and the ongoing quality of treatment obtained through observer-rated measures of competence are necessary to supplement the measure. Obtaining further validity data, as described above, to support the use of the measure would indicate however that the measure could be used as a training outcome without the need for it to be supplemented with observer rated performance data on each occasion. The ongoing monitoring of the quality of treatment provided is a separate issue that also requires methods that can be scaled.

It is worth noting that the present measure was developed and tested primarily as a summative measure to assess the outcome of training. Future work could investigate its use as a formative measure to help guide training and assess progress during ongoing supervision.

Global efforts to improve the mental health care gap (Pathare et al., 2018) depend in large part on the ability to scale up training programs for non-specialist providers so that they can deliver effective brief psychological treatments (Kemp et al., 2019; Ottman et al., 2020; Singla et al., 2018). To do this, scalable measures of competence are required as are effective methods of monitoring and ensuring the quality of the ongoing provision of treatment. The development of a measure of provider competence contributes to the dissemination and implementation of brief psychological treatments and thereby to more people benefitting from such treatments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Thank you to HAP experts who provided input on the blueprint and initial test items: Sona Dimidjian, Abhijit Nadkarni, Azaz Khan, Medha Upadhye, Pranali Kundaikar, Miriam Sequeira, Urvita Bhatia, Subhash Pednekar, Shravani Rangapuri, Aarti Girap.

This study is part of a larger program of research called ESSENCE (Enabling Translation of Science to Service to Enhance Depression Care) aimed at scaling up the delivery of depression care. ESSENCE is part of a network of Scale Up Hubs for research in global mental health supported by a grant from the National Institute of Mental Health [1U19MH113211]. The funder played no role in the study design; collection, analysis, or interpretation of data; writing of the manuscript; or decision to submit the manuscript for publication.

References:

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & P. T. (1999). Standards for Educational and Psychological Testing. American Education Research Association 10.4135/9781506326139.n662
- Andrich D (2004). Supplement: Applications of Rasch Analysis in Health Care. *Care*, 42(1), 17–116. 10.1097/01.mlr.0000103528.48582.7c
- Bond TG, & Fox CM (2015). Applying the Rasch model : fundamental measurement in the human sciences. Routledge.
- Boone WJ (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, 15(4). 10.1187/cbe.16-04-0148
- Boone WJ, Yale MS, & Staver JR (2014). Rasch analysis in the human sciences In *Rasch Analysis in the Human Sciences*. Springer Netherlands 10.1007/978-94-007-6857-4
- Branson A, Shafran R, & Myles P (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy*, 68, 19–26. 10.1016/j.brat.2015.03.002 [PubMed: 25791437]
- Case SM, & Donahue BE (2008). Developing high-quality multiple-choice questions tor assessment in legal education. *Journal of Legal Education*, 58(3), 372–387.
- Case SM, & Swanson DB (2002). Constructing Written Test Questions For the Basic and Clinical Sciences. Director, 27(21), 1–181. http://www.nbme.org/PDF/ItemWriting_2003/2003IWGwhole.pdf
- Chowdhary N, Anand A, Dimidjian S, Shinde S, Weobong B, Balaji M, Hollon SD, Rahman A, Wilson GT, Verdeli H, Araya R, King M, Jordans MJD, Fairburn C, Kirkwood B, & Patel V (2016). The healthy activity program lay counsellor delivered treatment for severe depression in India: Systematic development and randomised evaluation. *British Journal of Psychiatry*, 208(4), 381–388. 10.1192/bjp.bp.114.161075
- Cooper Z, Doll H, Bailey-Straebl S, Bohn K, de Vries D, Murphy R, O'Connor ME, & Fairburn CG (2017). Assessing Therapist Competence: Development of a Performance-Based Measure and Its Comparison With a Web-Based Measure. *JMIR Mental Health*, 4(4), e51 10.2196/mental.7704 [PubMed: 29089289]
- Cooper Z, Doll H, Bailey-Straebl S, Kluczniok D, Murphy R, O'Connor ME, & Fairburn CG (2015). The development of an online measure of therapist competence. *Behaviour Research and Therapy*, 64, 43–48. 10.1016/j.brat.2014.11.007 [PubMed: 25528502]
- Dimidjian S, Hollon SD, Dobson KS, Schmalings KB, Kohlenberg RJ, Addis ME, Gallop R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunner DL, & Jacobson NS (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, 74(4), 658–670. 10.1037/0022-006X.74.4.658 [PubMed: 16881773]

- Fairburn CG, Cooper Z Therapist competence, therapy quality, and therapist training. *Behav Res Ther.* 2011;49(6-7):373–378. doi:10.1016/j.brat.2011.03.005 [PubMed: 21492829]
- Frank HE, Becker-Haimes EM, & Kendall PC (2020). Therapist training in evidence-based interventions for mental health: A systematic review of training approaches and outcomes. *Clinical Psychology: Science and Practice.* 10.1111/cpsp.12330
- Ginzburg DM, Bohn C, Höfling V, Weck F, Clark DM, & Stangier U (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy*, 50(12), 747–752. 10.1016/j.brat.2012.09.001 [PubMed: 23072975]
- Hagquist C, Bruce M, & Gustavsson JP (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, 46(3), 380–393. 10.1016/j.ijnurstu.2008.10.007 [PubMed: 19059593]
- Haug T, Nordgreen T, Öst LG, Tangen T, Kvale G, Hovland OJ, Heiervang ER, & Havik OE (2016). Working alliance and competence as predictors of outcome in cognitive behavioral therapy for social anxiety and panic disorder in adults. *Behaviour Research and Therapy*, 77, 40–51. 10.1016/j.brat.2015.12.004 [PubMed: 26708332]
- Hoefl TJ, Fortney JC, Patel V, & Unützer J (2018). Task-Sharing Approaches to Improve Mental Health Care in Rural and Other Low-Resource Settings: A Systematic Review. *Journal of Rural Health*, 34(1), 48–62. 10.1111/jrh.12229
- Horton M, & Perry AE (2016). Screening for depression in primary care: a Rasch analysis of the PHQ-9. *BJPsych Bulletin*, 40(5), 237–243. 10.1192/pb.bp.114.050294 [PubMed: 27752340]
- Jordans MJD, Luitel NP, Garman E, Kohrt BA, Rathod SD, Shrestha P, Komproe IH, Lund C, & Background VP (2019). Effectiveness of psychological treatments for depression and alcohol use disorder delivered by community-based counsellors: two pragmatic randomised controlled trials within primary healthcare in Nepal. 10.1192/bjp.2018.300
- Kaslow NJ, Grus CL, Campbell LF, Fouad NA, Hatcher RL, & Rodolfa ER (2009). Competency Assessment Toolkit for professional psychology. *Training and Education in Professional Psychology*, 3(4S), S27.
- Kemp CG, Petersen I, Bhana A, & Rao D (2019). Supervision of task-shared mental health care in low-resource settings: A commentary on programmatic experience. In *Global Health Science and Practice* (Vol. 7, Issue 2, pp. 150–159). Johns Hopkins University Press. 10.9745/GHSP-D-18-00337
- Kohrt BA, Jordans MJD, Rai S, Shrestha P, Luitel NP, Ramaiya MK, Singla DR, & 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. 10.1016/j.brat.2015.03.009 [PubMed: 25847276]
- Kohrt BA, Schafer Alison, Willhoite Ann, van't Hof E, Pedersen GA, Watts S, Ottman K, Carswell K, & van Ommeren M (2020). Ensuring Quality in Psychological Support (WHO EQUIP): developing a competent global workforce. In *World Psychiatry* (Vol. 19, Issue 1, pp. 115–116). Blackwell Publishing Ltd. 10.1002/wps.20704 [PubMed: 31922690]
- Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, & Martinez RG (2015). Outcomes for implementation science: An enhanced systematic review of instruments using evidence-based rating criteria. In *Implementation Science* (Vol. 10, Issue 1, p. 155). BioMed Central Ltd. 10.1186/s13012-015-0342-x [PubMed: 26537706]
- Linacre JM (2019). Winsteps (Version 4.4.6). www.winsteps.com
- Martinez RG, Lewis CC, & Weiner BJ (2014). Instrumentation issues in implementation science. *Implementation Science*, 9(1), 118 10.1186/s13012-014-0118-8 [PubMed: 25185799]
- Muse K, & McManus F (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33(3), 484–499. [PubMed: 23454222]
- Online Panels: Get Responses for Surveys & Research | Qualtrics. (2019). <https://www.qualtrics.com/research-services/online-sample/>
- Ottman KE, Kohrt BA, Pedersen GA, & Schafer A (2020). Use of role plays to assess therapist competency and its association with client outcomes in psychological interventions: A scoping review and competency research agenda. *Behaviour Research and Therapy*. 10.1016/j.brat.2019.103531

- Patel V, Weobong B, Nadkarni A, Weiss HA, Anand A, Naik S, Bhat B, Pereira J, Araya R, Dimidjian S, Hollon SD, King M, McCambridge J, McDaid D, Murthy P, Velleman R, Fairburn CG, & Kirkwood B (2014). The effectiveness and cost-effectiveness of lay counsellor-delivered psychological treatments for harmful and dependent drinking and moderate to severe depression in primary care in India: PREMIUM study protocol for randomized controlled trials. *Trials*, 15(1), 101 10.1186/1745-6215-15-101 [PubMed: 24690184]
- Patel V, Weobong B, Weiss HA, Anand A, Bhat B, Katti B, Dimidjian S, Araya R, Hollon SD, King M, Vijayakumar L, Park A-L, McDaid D, Wilson GT, Velleman R, Kirkwood BR, & Fairburn CG (2017). The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. *The Lancet*, 389(10065), 176–185.
- Pathare S, Brazinova A, & Levav I (2018). Care gap: A comprehensive measure to quantify unmet needs in mental health. *Epidemiology and Psychiatric Sciences*, 27(5), 463–467. 10.1017/S2045796018000100 [PubMed: 29521609]
- Perry AE, & Horton M (2020). Assessing vulnerability to risk of suicide and self-harm in prisoners: a Rasch analysis of the suicide concerns for offenders in the prison environment (SCOPE-2). *BMC Psychiatry*, 20(1), 164 10.1186/s12888-020-02569-1 [PubMed: 32295553]
- Popham WJ (2001). Teaching to the test? *Educational Leadership*, 58(6), 16–20. <http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=507758868&site=ehost-live&scope=site>
- Qualtrics XM Platform. (2019). <https://www.qualtrics.com>
- Rakovshik SG, & McManus F (2010). Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. In *Clinical Psychology Review* (Vol. 30, Issue 5, pp. 496–516). 10.1016/j.cpr.2010.03.004 [PubMed: 20488599]
- Rasch G (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. MESA Press.
- Raviola G, Naslund JA, Smith SL, & Patel V (2019). Innovative Models in Mental Health Delivery Systems: Task Sharing Care with Non-specialist Providers to Close the Mental Health Treatment Gap. In *Current Psychiatry Reports* (Vol. 21, Issue 6, p. 44). Current Medicine Group LLC 1. 10.1007/s11920-019-1028-x [PubMed: 31041554]
- Richmond H, Copsey B, Hall AM, Davies D, & Lamb SE (2017). A systematic review and meta-analysis of online versus alternative methods for training licensed health care professionals to deliver clinical interventions. In *BMC Medical Education* (Vol. 17, Issue 1, p. 227). BioMed Central Ltd. 10.1186/s12909-017-1047-4 [PubMed: 29169393]
- Rosen RC, Ruzek JI, & Karlin BE (2017). Evidence-based training in the era of evidence-based practice: Challenges and opportunities for training of PTSD providers. In *Behaviour Research and Therapy* (Vol. 88, pp. 37–48). Elsevier Ltd. 10.1016/j.brat.2016.07.009 [PubMed: 28110675]
- Singh T, & Reyes-Portillo JA (2020). Using Technology to Train Clinicians in Evidence-Based Treatment: A Systematic Review. *Psychiatric Services*, 71(4), 364–377. 10.1176/appi.ps.201900186 [PubMed: 31960775]
- Singla DR, Kohrt BA, Murray LK, Anand A, Chorpita BF, & Patel V (2017). Psychological Treatments for the World: Lessons from Low- and Middle-Income Countries. *Annual Review of Clinical Psychology*, 13(1), 149–181. 10.1146/annurev-clinpsy-032816-045217
- Singla DR, Raviola G, & Patel V (2018). Scaling up psychological treatments for common mental disorders: a call to action. In *World Psychiatry* (Vol. 17, Issue 2, pp. 226–227). Blackwell Publishing Ltd. 10.1002/wps.20532 [PubMed: 29856556]
- Singla DR, Weobong B, Nadkarni A, Chowdhary N, Shinde S, Anand A, Fairburn CG, Dimidjian S, Velleman R, Weiss H, & Patel V (2014). Improving the scalability of psychological treatments in developing countries: An evaluation of peer-led therapy quality assessment in Goa, India. *Behaviour Research and Therapy*, 60(100), 53–59. 10.1016/j.brat.2014.06.006 [PubMed: 25064211]
- Smith AB, Rush R, Fallowfield LJ, Velikova G, & Sharpe M (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8(1), 33 10.1186/1471-2288-8-33 [PubMed: 18510722]

- Strunk DR, Brotman MA, DeRubeis RJ, & Hollon SD (2010). Therapist Competence in Cognitive Therapy for Depression: Predicting Subsequent Symptom Change. *Journal of Consulting and Clinical Psychology*, 78(3), 429–437. 10.1037/a0019631 [PubMed: 20515218]
- Tennant A, & Conaghan PG (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? In *Arthritis Care and Research* (Vol. 57, Issue 8, pp. 1358–1362). 10.1002/art.23108 [PubMed: 18050173]
- Tinsley HEA, & Dawis RV (1977). Test-Free Person Measurement with the Rasch Simple Logistic Model. *Applied Psychological Measurement*, 1(4), 483–487. 10.1177/014662167700100404
- Weobong B, Weiss HA, McDaid D, Singla DR, Hollon SD, Nadkarni A, Park A-L, Bhat B, Katti B, Anand A, Dimidjian S, Araya R, King M, Vijayakumar L, Wilson GT, Velleman R, Kirkwood BR, Fairburn CG, & Patel V (2017). Sustained effectiveness and cost-effectiveness of the Healthy Activity Programme, a brief psychological treatment for depression delivered by lay counsellors in primary care: 12-month follow-up of a randomised controlled trial. *PLOS Medicine*, 14(9), e1002385 10.1371/journal.pmed.1002385 [PubMed: 28898283]
- Wright B (1968). Sample-free test calibration and person measurement. In Paper presented at the Invitational Conference on Testing Problems <https://eric.ed.gov/?id=ED017810>
- Wright B (2005). A History of Social Science Measurement. *Educational Measurement: Issues and Practice*, 16(4), 33–45. 10.1111/j.1745-3992.1997.tb00606.x
- Wright Benjamin, & Stone M (1979). Best test design. *Measurement and Statistics*. <https://research.acer.edu.au/measurement/1>

Figure 1.
Stages of Competency Assessment Tool Development and Evaluation.
(HAP: Healthy Activity Program)

Figure 2.

A stepwise systematic process of sequential item deletion using RASCH modelling to create three equivalent tests

Figure 3.

Wright Map of Competency Scale: The left side of the map shows the distribution of participant ability measured from most able at the top to least able at the bottom. Right side of the map shows distribution of items based on the item difficulty measured with the most difficult at the top and the least difficult at the bottom. The mean item difficulty was defined as zero on the scale. The final pool of items (60 in number) which used to create Test A, Test B and Test C each having 17 unique items and 9 common items were used for plotting the Wright Map.

Figure 4.

Person Measure – Item Difficulty Histogram: A standard histogram of the person ability distribution and item difficulty distribution arranged horizontally. The graph illustrates that there are enough items to cover the person ability range of the participants.

Figure 5.

Test Characteristic Curve: Test Characteristic Curve (TCC) is an S-shaped curve, demonstrating the fact that the functional relationship between the measures and the expected scores is nonlinear. Plotting TCC for the final three tests made up of all retained items showed that there is no significant difference in difficulty level between the three tests.

Table 1:

Phases of HAP (Patel et al., 2017)

Phase of HAP	Goals of Phase
Early Phase (1-2 sessions)	<ul style="list-style-type: none"> • Engaging and establishing an effective counseling relationship • Helping patients understand the HAP model • Eliciting commitment for counseling
Middle Phase (3-6 sessions)	<ul style="list-style-type: none"> • Assessing activation targets and encouraging activation • Identifying barriers to activation and learning how to overcome them • Helping patients solve (or cope with) life problems
Ending Phase (1 session)	<ul style="list-style-type: none"> • Reviewing and strengthening gains that the patient made during treatment in order to prevent relapse

Table 2:**Blueprint**

COMMON SKILLS				
Number	Competency	Knowledge	Application	Number of Questions per Test Version
1	Active Listening	<ul style="list-style-type: none"> • Definition of reflective listening and paraphrasing 	<ul style="list-style-type: none"> • Uses reflective listening and paraphrasing • Summarizes • Asks clarifying questions • Allows silences • Uses minimal encouragers 	2
2	Open-ended questions	<ul style="list-style-type: none"> • Definition of open-ended and close-ended questions 	<ul style="list-style-type: none"> • Uses open vs close-ended questions at appropriate times • Elicits more than “yes-no” responses • Uses “tell me more” 	2
3	Collaborative style	<ul style="list-style-type: none"> • Different styles of counseling (more directive vs more listening/reflective) • Difference between friendly chat and counseling 	<ul style="list-style-type: none"> • Use the different styles of counseling appropriately • Does not simply tell the patient what to do • Ensures active participation of the patient • Elicits and incorporates patient input into treatment plans • Asks the patient to provide feedback on the sessions • Learns together collaboratively • Introduces self to patient properly • Elicits and address the patient’s chief complaint at some point during the session 	1
4	Appropriate language	<ul style="list-style-type: none"> • Knows local terms 	<ul style="list-style-type: none"> • Uses a level of language appropriate for patient • Uses local terms • Uses clear terminology • Avoids jargon 	Version A: 2 Version B: 1 Version C: 1
5	Demonstrating empathy, warmth, and genuineness		<ul style="list-style-type: none"> • Demonstrates friendly and warm behavior • Uses a genuine and sincere manner • Has a non-judgmental attitude • Demonstrates acceptance • Acknowledges and validates patient experience 	2
6	Affirmations and hope		<ul style="list-style-type: none"> • Discusses with the patient how life would be different if their stress were addressed and provides hope that through counseling, this can happen • Express appreciation of patient’s efforts and strengths in coping • Provide reassurance as a way of giving patients the courage to face a problem • Provides affirmations of the patient’s strengths and successes • Provides encouragement 	2
7	Goal setting and expectations	<ul style="list-style-type: none"> • Reasonable and realistic goals that could be suggested 	<ul style="list-style-type: none"> • Asks about goals • Discusses goals with the patient and whether they are realistic and achievable • Encourages and checks realistic expectations with an appropriate timeline 	1
8	Assessing self-harm/suicide	<ul style="list-style-type: none"> • Risk/protective factors for suicide • Misconceptions about suicide (e.g. asking about suicide will plant the idea in the patient’s head) 	<ul style="list-style-type: none"> • Asks about harm to self or others • Asks specifically about suicide • Works to develop a safety plan • Manages patients with high suicide risk appropriately 	2
9	Managing Crises	<ul style="list-style-type: none"> • Steps of crisis counseling 	<ul style="list-style-type: none"> • Assess patients in crisis appropriately • Manages personal crises • Manages bereavement • Manages domestic violence (for both victims and perpetrators) 	Version A: 1 Version B: 2 Version C: 2

COMMON SKILLS				
Number	Competency	Knowledge	Application	Number of Questions per Test Version
10	Involving a significant others (SO)	<ul style="list-style-type: none"> • When to involve a SO • Precautions when involving a SO 	<ul style="list-style-type: none"> • Includes SOs in session activities • Conducts sessions with SOs appropriately • Handles challenging situations when a SO is involved • Encourages interactions between patient and SO • Does role-plays with patients of interactions with SOs if needed 	1
TREATMENT (HAP) SPECIFIC SKILLS				
Number	Competency	Knowledge	Application	Number of Questions per Test Version
11	Understanding depression	<ul style="list-style-type: none"> • Symptoms of depression • Causes of depression • Consequences of untreated depression (e.g. suicide) • Misconceptions about depression • Tools to screen (PHQ-9) • Mild vs moderate vs severe depression and associated treatment • Treatment of depression • Common mental and physical issues that can accompany depression 	<ul style="list-style-type: none"> • Identifies/diagnoses depression • Uses the PHQ-9 correctly • Explains depression to patients • Addresses/know when to refer for: • Thinking too much • Anxiety and feeling tense • Problems with someone close to the patient • Difficulties with sleep • Substance use (tobacco) 	2
12	Communicating HAP Model and initiating treatment	<ul style="list-style-type: none"> • Benefits of counseling • Counselor's training/supervision • Number of sessions in HAP and over what time • Where sessions will be conducted • How a counselor can be contacted outside of the sessions) 	<ul style="list-style-type: none"> • Effectively initiates HAP • Explains HAP and counseling • Confirms patient understanding and elicits commitment to proceed with treatment • Establishes agreement about the length of the first meeting 	2
13	Exploring patient's feelings		<ul style="list-style-type: none"> • Asks "what happened" (identifying factors that contribute to or maintain depression) • Offers possibilities of common factors if needed • Asks "how did you feel" in the context of what happened • Offers common feelings/experiences if needed • Gets to know the patient 	2
14	Assessing functioning and links between functioning, activity, and mood	<ul style="list-style-type: none"> • Common effects of mood on the functioning • Different methods to rate mood 	<ul style="list-style-type: none"> • Asks "what did you do" in response to the events and subsequent feelings • Asks "what is the connection between what you do/don't do and how you feel and the problems in your life" • Explains the link between behavior and mood and stressor • Ask about activities that make patients feel good • Uses patient's activity calendar for assessing activity and mood 	2
15	Targets patient activation	<ul style="list-style-type: none"> • Common activation targets and barriers 	<ul style="list-style-type: none"> • Identifies relevant activation targets • Starts with small targets • Breaks down activities into small steps • Helps patient anticipate barriers to activation and strategizes to overcome them 	2
16	Problem-solving	<ul style="list-style-type: none"> • Common useful solutions that can be suggested • Knows the problem-solving approach 	<ul style="list-style-type: none"> • Guides patient to identify problems, generate solutions, apply the most appropriate one, and review the effectiveness of the chosen solution 	2

COMMON SKILLS				
Number	Competency	Knowledge	Application	Number of Questions per Test Version
			<ul style="list-style-type: none"> Teaches patients these steps so they can use them for future problems 	
17	Assessing and ensuring progress	<ul style="list-style-type: none"> Knows what the HAP homework entails 	<ul style="list-style-type: none"> Explains, plans, and reviews homework and activity calendar If a patient does not complete homework, asks why and addresses barriers Completes homework in session with the patient if needed Reviews progress with the PHQ-9 	2
18	Addressing barriers to treatment	<ul style="list-style-type: none"> Common barriers to attending sessions, completing treatment 	<ul style="list-style-type: none"> Asks about barriers Addresses barriers Promotes attendance Prevents treatment dropout 	2
19	Ending treatment	<ul style="list-style-type: none"> Criteria for stopping treatment 	<ul style="list-style-type: none"> Reviews skills (asks the patient to tell what they have learned from the sessions; highlights specific actions they have taken) Emphasizes patient's own role in getting better Motivates patients to use the strategies they learned across other life situations Summarizes steps to help the patient stay well over time Helps patient identify possible challenging future situations 	1
LOGISTICS				
Number	Competency	Knowledge	Application	Number of Questions per Test Version
20	Logistics/management of the session	<ul style="list-style-type: none"> Recommended structure for sessions Recommended frequency of sessions Recommended HAP timeline 	<ul style="list-style-type: none"> Sets an agenda at the beginning of sessions Summarizes at the end of sessions 	1
21	Engaging with the patient outside of the health center	<ul style="list-style-type: none"> Resources available to patients if they have an emergency 	<ul style="list-style-type: none"> Conducts home visits appropriately Conducts effective counseling sessions over the phone Keeps in contact with patients outside of sessions Handles patient emergencies 	1
22	Confidentiality	<ul style="list-style-type: none"> Counselor-patient confidentiality agreement 	<ul style="list-style-type: none"> Explains confidentiality Breaks confidentiality when needed Maintains confidentiality during the session (if the session is interrupted, etc.) 	2
23	Personal well-being and safety	<ul style="list-style-type: none"> Signs of burnout Self-care strategies Strategies to ensure personal safety during sessions (e.g. keeping self between patient and door) 	<ul style="list-style-type: none"> Maintains boundaries with patients Responds appropriately when feeling unsafe during the session Manages burnout 	1
24	Medication	<ul style="list-style-type: none"> Medication side effects Knows when medications are used 	<ul style="list-style-type: none"> Works with patient to address barriers to taking medication 	1