

# Positive imagery cognitive bias modification for symptoms of depression among university students in Pakistan: A pilot study

Journal of Experimental Psychopathology  
April-June 2020: 1–22  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2043808720918030  
journals.sagepub.com/home/jepp



**Akhtar Bibi**

Ruhr-Universität Bochum, Germany

**Jürgen Margraf** 

Ruhr-Universität Bochum, Germany

**Simon E. Blackwell** 

Ruhr-Universität Bochum, Germany

## Abstract

Depression is highly prevalent among university students in Pakistan, but treatment provision is inadequate. Computerized interventions may provide one means of overcoming treatment barriers. The present study piloted a computerized cognitive training paradigm involving repeated generation of positive mental imagery, imagery cognitive bias modification (imagery CBM), as a potential brief intervention for symptoms of depression among university students in Pakistan. Fifty-five participants scoring above a questionnaire cutoff indicating at least mild levels of depression were randomly assigned to either imagery CBM or a sham training control condition (peripheral vision task [PVT]). Participants were instructed to complete one training session from home daily over the course of 1 week. Outcomes were measured at post-training and a subsequent 2-week follow-up and included measures of depression, anhedonia, and positive affect. Participants provided positive feedback about the imagery CBM intervention but encountered practical problems with the study schedule, resulting in high rates of attrition, particularly at follow-up. Further, internal consistency of outcome measures was often low, and the PVT did not appear to be an adequate control condition in this study. However, overall the results suggest that with appropriate adaptations to the study methods formal investigation of efficacy is warranted.

## Keywords

Cognitive bias modification, computerized intervention, depression, interpretation bias, mental imagery

Date received: 15 November 2019; accepted: 6 March 2020

## Corresponding author:

Simon E. Blackwell, Department of Clinical Psychology and Psychotherapy, Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Bochum, Massenbergsstraße 9-13, 44787 Bochum, Germany.

Email: [simon.blackwell@rub.de](mailto:simon.blackwell@rub.de)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Access to mental health care is considered to be one of the most significant factors in the worldwide sustainable development program (United Nations, 2015; Votruba et al., 2016). In developing countries such as Pakistan, the “treatment gap” (the percentage of people who need but do not receive treatment) is about 90% for common mental health disorders (Kakuma et al., 2011; Saxena et al., 2007; Whiteford et al., 2013). Only specialized urban health care settings have facilities for providing treatments such as pharmacotherapy and psychotherapy (Jooma et al., 2009), and there is a lack of both mental health professionals (Ahmad, 2007) and evidence-based culturally applicable psychological interventions (Chisholm et al., 2007; Naeem et al., 2006). Moreover, seeking psychological support from mental health professionals is highly stigmatized (Naeem et al., 2009; Sultan, 2011).

Improved access to mental health treatments is needed across all sections of Pakistani society, and university students are one group of people in Pakistan for whom the need for treatment is particularly apparent (Bibi, Blackwell, & Margraf, in press; Bibi, Lin, & Margraf, 2020). In general, university is a time when students face many intellectual and emotional challenges, including competition for good grades, the need for career-planning, and home-sickness, which can cause physical, emotional, and social problems. As a result, students are vulnerable to developing a range of psychological problems (Eisenberg et al., 2007). This is also the case in Pakistan, with a prevalence study of mental health problems among university students in Pakistan finding that 31% of participants experienced “severe” mental health problems and 16% experienced “very severe” mental health problems (Saleem et al., 2013). Another study similarly estimated that 28.7% of university students in Pakistan are severely depressed (Bukhari & Khanam, 2015). However, very few universities offer counselling services or similar resources. Students are therefore one group for whom there is a pressing need to improve mental health treatment provision and access (Bibi, Lin, & Margraf, 2020).

Computer-based treatments that do not require a trained therapist provide a possible method for increasing the number of people who can receive treatment, both because of the reduced need for highly trained health professionals and because they may help overcome barriers such as stigma. One potential route for developing new computerized treatments has emerged from experimental psychopathology, specifically research investigating the use of computerized

cognitive training procedures to modify the dysfunctional cognitive biases that characterize many disorders. These “Cognitive Bias Modification” (CBM; Koster et al., 2009) paradigms were originally developed to investigate the potential causal role of negative biases in processes such as interpretation and attention in contributing to emotional vulnerability, but more recently they have been adapted to train positive biases with the aim of potential clinical applications (Woud & Becker, 2014). The promise of CBM procedures as interventions is that, if effective, they could help contribute to the drive to develop the easily accessible, low-cost approaches that are needed to address the high demand for mental health treatments.

One CBM paradigm developed in the context of depression aims to target the deficit in positive mental imagery (Holmes et al., 2016) and negative interpretation bias (Butler & Mathews, 1983; Everaert et al., 2017) that characterize the disorder. This “imagery CBM” was developed from the interpretation training paradigm originally developed by Mathews and Mackintosh (2000), adapted in further research to enhance the focus on generating mental imagery (Holmes et al., 2009; Holmes & Mathews, 2005). As applied in a clinical context for depression, the most frequently investigated training schedule comprises daily sessions over the course of 1 week (Blackwell & Holmes, 2010; Williams et al., 2013). In each session, participants listen to a series of training scenarios, which are descriptions of mostly everyday situations structured such that they start ambiguous but always end positively. An example training scenario might be “You wake up in the morning, and as you think about the day ahead you feel full of *energy and enthusiasm*” (positive resolution in italics). Participants are instructed to imagine themselves in the training scenarios as if actively involved in the situations described. Through repeated practice in imagining positive resolutions to ambiguous scenarios in the training sessions, the paradigm aims to train an adaptive bias to automatically imagine positive resolutions for ambiguous situations in daily life. Overall, imagery CBM appears promising in early stages of clinical investigation (Hitchcock et al., 2017). The majority of studies investigating the potential of some variant of a 1-week imagery CBM training schedule to reduce symptoms of depression have found superiority over a closely matched “sham training” control condition (Lang et al., 2012; Pictet et al., 2016; Torkan et al., 2014) or waitlist (Pictet et al., 2016; Williams et al., 2013). One further study did not find a

between-group difference versus sham training in intention-to-treat analyses but did find superiority of the imagery CBM in a “complete case (CC)” analysis comprising those participants who completed the training schedule (Williams et al., 2015). However, the largest randomized controlled trial (RCT) of imagery CBM for depression conducted to date, including 150 adults with current major depression, did not find superiority of the imagery CBM over a sham training control when using a 4-week training schedule (Blackwell et al., 2015). Overall, in these initial stages of research, imagery CBM appears promising, albeit with questions remaining about potential longer term effects or applications.

The current study therefore aimed to investigate the potential of imagery CBM for reducing depression and related symptoms among university students in Pakistan, using the most commonly applied 1-week training schedule and 2-week follow-up as a starting point. While imagery CBM has been successfully investigated in a variety of countries, namely, the United Kingdom (Blackwell et al., 2015; Blackwell & Holmes, 2010; Lang et al., 2012), Australia (Williams et al., 2013, 2015), Iran (Torkan et al., 2014), and Switzerland (Pictet et al., 2016), it cannot be assumed that it will necessarily be acceptable or suitable for any given population. Research in the field of mental health is particularly sensitive to translational problems due to differences in languages and cultures around the understanding and expression of symptoms of mental health problems (Scholten et al., 2017). For example, in Pakistan, people with depression may often present their psychological problems in the form of bodily symptoms (Shah et al., 2011). Previous studies have also not specifically targeted a student population (although many have included students among the participants). There may be differences between different populations in terms of the practicalities of delivering or assessing the outcome of an intervention, and thus, it cannot be assumed that a study protocol designed in one context can be straightforwardly applied in another. We therefore wanted to ascertain the feasibility of investigating imagery CBM as a potential low-intensity intervention for university students with depression in Pakistan, including gathering information about acceptability, study design parameters, and initial estimates of efficacy.

In addition to using a 1-week training schedule similar to that used in the majority of the previous studies, in the current study, we planned to measure

a similar set of outcomes, specifically depression and anxiety as symptom outcomes, and interpretation bias and ability to generate vivid positive mental imagery as mechanisms measures. However, following results from the most recent studies, we also planned to include a range of additional measures. These more recent studies have suggested that imagery CBM may be useful for targeting deficits in positive affect and anhedonic symptoms (Blackwell et al., 2015; Pictet et al., 2016; Williams et al., 2015), which could be particularly beneficial clinically as these aspects of depression are thought not to respond well to current treatments (Craske et al., 2016; Dunn, 2012). We therefore included specific measures of positive affect and anhedonia, which have not been typically assessed in previous clinical studies (Pictet et al., 2016). We also included a measure of behavioral activation, following research suggesting that imagery CBM may also help increase behavioral activation, that is, purposeful engagement in meaningful activities in everyday life (Renner et al., 2017).

In terms of a control condition against which to compare the imagery CBM, we aimed to use one that, as far as possible, resembled a “placebo” version of the intervention, that is, controlling for nonspecific aspects such as expectancy, researcher contact, and engagement in a regular training schedule but not including the “active ingredients” of the intervention. Previous clinical studies have generally used a closely matched “sham training” control condition that differed from the active training by just one component. For example, in several studies, the “sham training” control condition was identical to the imagery CBM intervention, except that half of the ambiguous stimuli were resolved negatively, rather than positively (Lang et al., 2012; Pictet et al., 2016; Williams et al., 2015). Such a control condition was intended to isolate the specific contribution of the training contingency to always expect a positive resolution of the ambiguity. While such control conditions are extremely useful for isolating one or more components of an intervention, in the current study, we were not interested in repeating findings about the importance of one or more specific aspect of the imagery CBM paradigm but rather to investigate whether it may provide benefits beyond the nonspecific effects that may be associated with a computerized cognitive training intervention (Blackwell et al., 2017). We therefore chose as a control condition an identical schedule of computer-based cognitive training tasks but without including the specific session-content of the imagery

CBM intervention. Instead, participants in the control condition completed a “Peripheral Vision Task” (PVT), which engages spatial attention and which we thought could not plausibly influence the specific cognitive mechanisms targeted by imagery CBM. Theoretically, any task requiring concentration and with potential face validity as cognitive training could serve as such as control condition. We chose the PVT specifically as it had been used as a control condition in studies investigating other cognitive training paradigms in the context of depression (Calkins et al., 2015) and thus appeared to have some evidence of acceptability and credibility for participants with depression.

In summary, the current pilot study aimed to investigate whether positive imagery CBM could be successfully applied in the new population and setting of university students in Pakistan, to provide information about feasibility of such a line of research, including gathering data about acceptability, initial estimates of likely effect sizes across a range of outcome measures, and information to inform design of any future studies.

## Method

### *Participants, recruitment, and settings*

As a pilot rather than hypothesis-testing study, the sample size was determined by pragmatic considerations rather than to provide a certain level of statistical power (Leon et al., 2011). A recruitment window of 12 weeks between October and December 2016 was available, and recruitment stopped after this time, at which point a total number of 55 participants had been randomized. Participants were recruited from Pakistani universities in the area (National University of Modern Languages [NUMLs], Islamabad, and International Islamic University, Islamabad) through advertisements on social media and poster advertisements in the region. Participants were offered an incentive of entry to a prize draw to win a thermos flask.

Potential participants expressing an interest in the study were emailed an information sheet and invited to attend a first assessment session. Participants were instructed to bring a laptop with them and emailed a link to download the latest version of Java beforehand, to ensure that the training programs would be able to run on their computer. The following inclusion criteria were applied at assessment: university student aged 18 or above; score of 6 or more on the Quick Inventory of Depressive Symptomatology—Self Report (QIDS; Rush et al., 2003), indicating at least

mild depression; sufficient language skills (Urdu and English) to complete the study procedures; have a computer capable of running the training programs (tested via running a test Java program on their computer at assessment); and able and willing to complete all the study procedures. Exclusion criteria were as follows: currently receiving treatment for a psychiatric condition; potential suicidality (as indicated by a score of 2 or more on item 12 of the QIDS); existence of condition or circumstances that would interfere with study procedures, such as neurological impairment, red–green color-blindness, or severe visual or auditory impairment; and drug abuse, recent changes in medicine, previous history of psychiatric disorder, or bipolar disorder (according to participant self-report). Participants judged eligible at the initial eligibility assessment were invited back for the baseline/pretraining assessment.

Assessment sessions took place at NUML and were conducted by a doctoral student from Ruhr-Universität Bochum (A.B.). The study received ethical approval from the ethics committee for the Faculty of Psychology, Ruhr-Universität-Bochum, Germany (ref: 315).

### *Design*

The study used an experimental design with two parallel groups. Participants were randomly assigned in a 1:1 ratio to one of two groups, one of whom completed sessions of imagery CBM (imagery CBM condition) and the other of whom completed sessions of the PVT (control condition). Outcome measures were scheduled for assessment at baseline (pretraining), after 1 week of training (post-training), and at a 2-week follow-up.

All study procedures and tasks were administered in English unless otherwise indicated. Administering questionnaires (and other study materials) in English (as also carried out in other related studies, e.g., Bibi, Blackwell, & Margraf, in press; Bibi, Lin, & Margraf, 2020) was not anticipated to be a problem, as English is the official language at universities in Pakistan. Most university students in Pakistan will have studied English from their first year in school, and many will have had their final years of school before university taught in English.

### *Interventions*

*Imagery CBM.* The imagery CBM intervention was adapted from that developed in previous experimental (Holmes et al., 2009) and clinical (Blackwell &

Holmes, 2010) work. Training stimuli consisted of descriptions of mostly everyday scenarios, structured such that they started ambiguous but always resolved positively, and participants were instructed to imagine themselves in each scenario as it unfolded. The training schedule comprised an introductory session completed at the end of the baseline assessment, followed by six training sessions scheduled to be completed one per day over the following week from home. Each session comprised 48 training stimuli, organized into 6 blocks of 8, with the exception of the introductory session, which comprised a longer introduction followed by 3 blocks of 8 stimuli.

The introductory session started with a general introduction to mental imagery, followed by a series of examples and exercises to explain to participants how they should imagine the scenarios. In particular, participants were instructed to try to imagine the scenarios as vividly as they could, using field perspective (i.e., through their own eyes) and multiple senses, feeling actively involved in the situation described, and trying to stay absorbed in the imagery, avoiding analyzing or thinking verbally about the situations. Participants were also instructed to focus on the outcome of the scenarios. The instructions were adapted from those given face-to-face in previous studies (Blackwell et al., 2015; Holmes et al., 2009) with computer administration rather than face-to-face instruction chosen to reduce the need for administration by a researcher/clinician. Stimuli presentation was similar to that in previous studies, in that participants saw a screen saying “Close your eyes. Imagine.” for 1 s, followed by a blank screen during which time the scenario was played. After completion of the scenario, there was a pause of 1 s, before a beep then sounded and participants opened their eyes and rated the vividness of their image on a scale from 1 (*not at all vividly*) to 5 (*extremely vividly*). After participants had made their rating, the program automatically moved on to the next stimulus. Breaks were self-paced and comprised two screens. The first reminded participants of the task instructions, and the second provided visual feedback on their vividness ratings (as in Blackwell et al., 2018). The visual feedback comprised a graph showing, for each block completed so far that session, their mean, minimum, and maximum vividness ratings. Participants were encouraged to reflect on the pattern of change (if any) in vividness ratings and how they could improve for the next block. The inclusion of feedback followed previous suggestions that this may help enhance engagement

while participants were completing multiple sessions from home (Blackwell et al., 2015; Pictet et al., 2016). The six sessions completed from home had only a few brief screens of reminder instructions and one “practice” scenario before starting the training blocks.

The training stimuli used were all unique, such that there were 312 scenarios in total, order of which was randomized for each participant across all training sessions. The 312 scenarios comprised 278 adapted from previous studies (Blackwell et al., 2015; Blackwell & Holmes, 2010; Holmes et al., 2006, 2009) and 34 specifically designed for this study to be applicable to the study sample (university students in Pakistan). Scenarios were translated into Urdu by the first author (A.B.) with the help of two other bilingual (English and Urdu) experts. Although most of the study materials (including instruction screens during the imagery CBM) were presented in English, we felt that providing the training scenarios in the participants’ first language would facilitate them absorbing themselves in imagining the scenarios. The linguistic structure of the translated scenarios was similar to the original scenarios, in that they started ambiguous and the positive outcome was only revealed toward the end of the scenario. For example, one of the newly developed scenarios was

You need to take a bus to the university, but have heard that many buses aren’t running due to strikes. When you arrive at the bus stop you find that your bus *is in fact running and you relax as you see it approach.* (positive resolution in italics)

However, due to the structure of the Urdu language, word order was not identical to the English versions of the scenarios, in that the verb would be positioned toward the end of the sentence. The contents of the previously used training scenarios were also adapted according to Pakistani Muslim culture, for example, replacing mentions of Christian festivals with Muslim festivals, removing any references to alcoholic drinks, and adapting references to the weather. All the scenarios were recorded in female voice, and the duration of each was 20–30 s. The smaller number of scenarios per session compared to previous English-language studies (i.e., 48 compared to 64 in most clinical studies) was due to the longer duration of the scenario recordings, in part due to the requirements of the Urdu language (those in the English-language studies are typically 8–15 s), and the desire to keep the length of individual sessions below 30 min.

The imagery CBM program was implemented as a desktop application written in Java by the last author (S.E.B.). Although the training scenarios were recorded in Urdu, the instructions on screen were presented in English. Participants were instructed to install the latest version of Java before attending the study assessments, and the program was then copied onto their laptop during the pretraining assessment.

**PVT.** The PVT was modeled on that described in previous research (Calkins et al., 2015). In each trial, participants were presented with a display of grey circles arranged in a ring, with a white fixation cross in the center. At the start of the trial, one (randomly selected) circle was designated the starting circle, via a white circle around its perimeter. Participants were instructed to keep their eyes on the fixation cross but to focus their attention on the starting circle. They then heard a series of low tones (range: 1–9, randomly determined on a trial-by-trial basis) and were instructed with each tone to move their attention one circle around the ring in a clockwise direction. At the end of the sequence of low-pitched tones, a high-pitched tone would sound, at which point the circles would become colored. After a further 1.5 s, the display would disappear and participants were asked to report the color of circle via the keyboard. Feedback (“Correct!” or “Wrong!”) was provided immediately afterward. The PVT was adaptive, such that after four consecutive correct responses, one extra circle would be added to the display, with the size of each circle decreasing. After four consecutive incorrect responses, one circle would be removed from the display, with the size of each circle increasing. Each session started with an array of 15 circles.

As with the imagery CBM, the PVT training schedule comprised an introductory session, followed by six sessions scheduled for each day over the following week. The introductory session consisted of an extended introduction to the task, in which participants were encouraged to track the circles with a finger to help learn the task, before moving on to tracking the circles with only their attention. Four blocks of 20 trials then followed. The subsequent training session comprised 6 blocks of 20 trials, with only a brief introduction. Participants could take a self-paced break in-between blocks. The number of blocks was chosen to make the PVT sessions approximately equal in length to the imagery CBM sessions. The PVT was implemented as a Java desktop application written by the last author (S.E.B.), installed

onto the participant’s laptop, with all instructions in English.

## Measures

Measures were all administered on paper in their original English form, unless otherwise indicated. This was due to the lack of validated Urdu versions for several of the questionnaires and the desire to keep language uniform across all measures. Where figures for internal consistency are provided subsequently, these are all Cronbach’s  $\alpha$  unless otherwise stated. As the study was not intended for efficacy testing, measures were not prespecified as primary or secondary outcomes.

## Outcome measures

**The Quick Inventory of Depressive Symptomatology (QIDS).** The QIDS is 16 item self-report questionnaire, designed to measure all depressive symptoms as mentioned in the diagnostic criteria for a Major Depressive Episode in the Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition (American Psychiatric Association, 1994) (Rush et al., 2003). Participants are asked to rate each item according to how they have felt over the past 7 days. The QIDS generally has a high internal consistency,  $\alpha = .86$  (Trivedi et al., 2004), and shows convergent validity in relation to other depression measures as well as good sensitivity to change (e.g., Reilly et al., 2015; Rush et al., 2003). In our sample, the internal consistency at baseline was .46 (“Unacceptable”), at post-training was .38 (“Unacceptable”), and at follow-up was .56 (“Poor”).

**Behavioral Activation for Depression Scale (BADs).** The BADs is used to measure engagement in functional and dysfunctional depression-relevant behaviors (e.g., behavioral engagement vs. avoidance) over the previous week, covering the following areas: activation, avoidance/rumination, work/school impairment, and social impairment (Kanter et al., 2007). In the current study, the short 9-item version of BADs was used (Kanter et al., 2007). Each item is rated on 7-point scale from 0 (*not at all*) to 6 (*completely*). Internal consistency, construct validity, and predictive validity of the short form of BADs have been reported to be acceptable (Kanter et al., 2007). In our sample, the internal consistency of the BADs at baseline was .70 (“Acceptable”), at post-training was .80 (“Good”), and at follow-up was .87 (“Good”).

**Positive and Negative Affect Schedule (PANAS)—Positive scale.** The PANAS is a well-validated measure of positive and negative emotions (Watson & Clark, 1994). In the current study, a positive scale was used comprising the joviality (8 items), self-assurance (6 items), and serenity (3 items) subscales of the extended PANAS. These subscales were chosen as they were judged most relevant to the positive emotions targeted in the imagery CBM. The “past week” administration instructions were used, such that participants were asked to indicate the extent to which each item applied to how they had been feeling “over the past week,” using scales ranging from 1 (*not at all*) to 5 (*extremely*). In validation samples,  $\alpha$  reliability for positive affect scales has ranged from .83 to .90, and the scales have shown good factor and construct validity (Watson & Clark, 1994). In our sample, the internal consistency for the PANAS-P at baseline was .69 (“Questionable”), at post-training was .86 (“Good”), and at follow-up was .78 (“Acceptable”).

**Depression, Anxiety and Stress Scale-21 (DASS-21) items.** The DASS-21 comprises three 7-item subscales (depression, anxiety, and stress), asking about the past week. Each item is rated from 0 (*never*) to 3 (*almost always*) (Lovibond & Lovibond, 1995). High internal consistency is generally reported for each subscale of the DASS (e.g., depression: .88; anxiety: .82; stress: .90; Henry & Crawford, 2005) as well as construct validity in relation to other self-report measures (Henry & Crawford, 2005). In our sample, internal consistency for depression at baseline was .48 (“Unacceptable”), for anxiety was .31 (“Unacceptable”), and for stress was .44 (“Unacceptable”). At posttreatment, internal consistency for depression was .77 (“Acceptable”), anxiety was .72 (“Acceptable”), and stress was .77 (“Acceptable”). At follow-up, internal consistency for depression was .36 (“Unacceptable”), anxiety was .68 (“Questionable”), and stress was .66 (“Questionable”).

**Dimensional Anhedonia Rating Scale (DARS).** The DARS measures anhedonia across the following four areas: hobbies/pastimes, food/drink, social activities, and sensory experience (Rizvi et al., 2015). For each area, participants are asked to provide their own examples of two activities that they would normally enjoy (e.g., “gardening, playing the guitar” for hobbies/pastimes). Examples given by participants in the current study included cooking, biryani, milkshakes, watching movies, getting together with friends, shopping, and listening to music. For each of these categories of

activities, participant rate a series of questions assessing motivation (to take steps to engage in the activity), effort (sustained energy expenditure), interest in/desire to engage in the activity, and consummatory pleasure (enjoyment of the activity). Internal consistency for the DARS has been reported as .92, and it has shown good convergent and divergent validity in community and clinical samples (Rizvi et al., 2015). In the current study, an extended 26-item version of the DARS provided by the scale authors was used. In our sample, the internal consistency of the DARS at baseline was .65 (“Questionable”) and at post-training was .91 (“Excellent”).

**Ambiguous scenarios test (AST)-D-II.** The AST is a measure of interpretation bias, one of the cognitive processes thought to be targeted via imagery CBM, with the items developed to reflect Beck’s cognitive triad (self, world, and future (Beck, 1976; Rohrbacher & Reinecke, 2014). Participant read ambiguous descriptions of situations, and for each one have to imagine themselves in the situation and then rate it for “pleasantness” (i.e., reflecting their subjective appraisal of the situation) on a scale from  $-5$  (*extremely unpleasant*) to  $+5$  (*extremely pleasant*). Two 15-item parallel versions (A and B) have been developed and have been found to be structurally stable, internally consistent, and valid.<sup>1</sup> The two versions have been reported to show acceptable internal consistency ( $\alpha = .77$  for version A,  $\alpha = .78$  for version B; Rohrbacher & Reinecke, 2014). In our sample at baseline, the internal consistency of AST (A) was .47 (“Unacceptable”) and AST (B) was .40 (“Unacceptable”). At post-training, the internal consistency of AST (A) was .75 (“Acceptable”) and AST (B) was .77 (“Acceptable”).

**Scrambled sentences task.** The Scrambled Sentences Test (SST) was used as an indirect measure of depressive interpretation bias (Phillips et al., 2010; Wenzlaff, 1993). Participants were asked to unscramble a list of 20 mixed sequences of words (e.g., good feel very bad I usually) under a cognitive load (remembering a six digit number) and with instructions that they had limited time. The SST measures the tendency of participants to interpret ambiguous information either positively (I usually feel very good) or negatively (I usually feel very bad). A “negativity” score was generated by calculating the proportion of sentences completed correctly with a negative emotional valence. When completed under cognitive load, the paper-administered version of the Scrambled Sentences Test

generally shows strong associations with depression symptoms (e.g., Phillips et al., 2010), has been found to predict future depression symptoms (Rude et al., 2002), and also responds differentially to CBM for interpretations versus computerized Cognitive Behaviour Therapy (CBT; Bowler et al., 2012). We calculated internal consistency as split-half reliability (odd vs. even-numbered items) using the Spearman–Brown formula. The nature of the item-by-item scoring of the SST and how this contributes to the total score (i.e., only using correctly answered trials then coding these as negative or non-negative) meant that Cronbach's  $\alpha$  would not be a suitable index of internal consistency, hence our use of the split-half method (see e.g., Parsons et al., 2019). At baseline, the two halves of the SST correlated negatively for both version A ( $r = -.21$ ) and version B ( $r = -.35$ ), and we did not proceed to calculate internal consistency. At post-training, the split-half reliability of SST A was .30 ("Unacceptable") and the two halves of the SST correlated negatively for version B ( $r = -.13$ ).

**Prospective imagery test (PIT).** The PIT was used to measure the vividness with which participants could imagine positive events in their future, another process thought to be targeted by the imagery CBM (Stöber, 2000). The PIT was split into two parallel forms (A and B), order of presentation of which was counterbalanced across participants. Each administration of the PIT included five positive and five negative possible future scenarios, which participants were asked to imagine happening to them in the near future. Participants then rated the vividness of their image on a scale ranging from 1 (*no image at all*) to 5 (*very vivid*). Good internal consistencies for the PIT have been reported in both unselected community and depressed samples, and although it has not been formally validated, it correlates well with other measures of mental imagery (Blackwell et al., 2013; Ji et al., 2017). In our sample, the internal consistency of PIT (A) (Positive) at baseline was .60 ("Questionable") and PIT (B) (Positive) was .35 ("Unacceptable"), whereas the internal consistency of PIT (A) (Negative) at baseline was .53 ("Poor") and of PIT (B) ("Negative") was .51 (Poor). At post-training, the internal consistency (Cronbach's  $\alpha$ ) of PIT (A) (Positive) was .50 ("Poor") and PIT (B) (Positive) was .35 ("Unacceptable"), whereas the internal consistency (Cronbach's  $\alpha$ ) of the PIT (A) (Negative) at post-training was .67 ("Questionable") and of PIT (B) (Negative) was .61 ("Questionable").

## Other measures

**Assessment of expectancy, feedback, and acceptability.** Expectancy was measured at baseline using the three expectancy questions from the Credibility and Expectancy Questionnaire (Deville & Borkovec, 2000). At post-training, participants completed questionnaire ratings of various aspects of engagement in their allocated training task, such as task difficulty and perceptions of task usefulness. All ratings were made on a 1–9 scale (e.g., 1 = *not at all* to 9 = *extremely*). For participants in the imagery CBM condition, this included questions about task acceptability as a potential intervention. Participants could also provide more detailed written feedback.

## Randomization

Randomization to the imagery CBM or PVT group was via a Java desktop application written by one of the researchers (S.E.B.), who had no involvement in recruitment, assessment, or other procedures involving participant contact. The randomization sequence was generated by this researcher using a true number random generator (random.org) and concealed within the compiled Java application (i.e., not in a human-readable form). Randomization was stratified by gender and baseline QIDS score (mild to moderate  $< 16$  vs. severe  $\geq 16$ ). Because it was not known how many people could be recruited in the available time window, fixed short block lengths of two were used to ensure fairly balanced participant numbers in each group (a maximum discrepancy of four participants) even with small numbers of participants and unbalanced strata. For the purposes of allocation concealment, this blocking strategy was not communicated to other researchers involved in the study. Further, randomization occurred following completion of the main clinical outcome measures at the baseline assessment (all outcome measures excluding those requiring counterbalancing, i.e., AST, SST, PIT), such that participant allocation could not be known to the assessing researcher until this point. Following randomization, the researchers conducting the baseline, post-training, and follow-up sessions were not blind to participant allocation.

## Procedure

Assessments were conducted by a researcher following a written experimental protocol (in English). At the initial eligibility assessment, participants were



again provided with the study information sheet and provided written informed consent. They then completed a demographics questionnaire and the QIDS and had a test Java program installed on their laptop to check for program compatibility. Eligible participants were invited back for the baseline assessment, and appointments for the post-training and follow-up appointment were also made at this time.

At the baseline session, participants completed the following pretraining measures in this order: QIDS, PANAS-P, DARS, DASS, and BADS. They then had a brief break, in which they were then randomized to the imagery CBM or PVT group, and then completed the AST, PIT, and SST. Counterbalancing for the AST, PIT, and SST was achieved via simply alternating between completing version A for each of these at baseline and B at post-training and vice versa for each successive participant within each condition. The relevant training program was then installed on the participant's computer, and they were provided with a brief description of what the task involved (imagery CBM: "This task involves listening to descriptions of everyday situations and imagining yourself in them," PVT: "This task involves concentrating on displays presented on the computer screen and using your attention and memory to keep track of them.") before completing the Expectancy Questionnaire (EQ). Participants then went through the introductory session of their assigned training task on their own. As in previous studies (Blackwell et al., 2015; Blackwell & Holmes, 2010), to increase participant adherence to and engagement in their assigned training task, before the end of the session, the researcher discussed with the participant the practicalities of and rationale for completing the training sessions. This included helping them plan when they would complete the sessions each day, including completion of a paper "session planner" and problem-solving potential obstacles; emphasizing the importance of concentrating on and fully engaging in the tasks if they were to have a beneficial impact, and suggesting ways in which the participant could help maintain concentration (e.g., taking breaks, choosing a suitable time and location); and using the analogy of going to the gym, that is, completing the training sessions might be not very enjoyable and might not feel like it is having an immediate impact, but the more effort put in, the more chance of obtaining benefits.

The participants were then scheduled to complete one session per day at home over the subsequent week. They were sent regular standardized text

messages to remind them about the sessions, and thanking them for sessions completed.

After completion of the training schedule, participants returned to university and completed the post-training outcome measures in the same fixed order as at baseline (QIDS, PANAS, DARS, DASS, BADS-SF, AST, PIT, SST), followed by the manipulation check/task engagement and feedback questionnaires.

Two weeks later, participants returned to the university for the follow-up assessment, in which they completed the outcome measures (QIDS, PANAS, DASS, BADS) a final time. They were then debriefed. This included reminding participants that the study involved comparing two computer programs, one of which was expected to be more effective than the other for reducing symptoms of depression, and asking them which task they thought they had been assigned to (i.e., active training or control, phrased as the "more effective one" and the "not so effective one").

### Statistical analyses

As a pilot study, the aim was not hypothesis testing (Leon et al., 2011), and thus  $p$ -value-based inferential statistics were not used. For demographic variables and measures related to training engagement, such as expectancy and feedback, descriptive statistics (means and 95% confidence intervals [CIs]) were calculated for each group. Descriptive statistics and reliabilities were calculated using Statistical Package for Social Sciences software version 23.

For the outcome measures, outcomes at each time point were calculated both as raw means and as standardized effect sizes, both with 95% CIs. These were calculated both for an "intention to treat (ITT)" sample (i.e., all participants randomized, regardless of session or outcome measure completion) and for a "CC" sample (i.e., all participants providing outcome data for the specific time point).

To derive means, effect sizes, and 95% CIs for the ITT sample, a mixed-model repeated measures analysis of variance was fitted over the three assessment time points, and estimates derived from this model. This allowed an ITT approach without having to, for example, impute missing data (Gueorguieva & Krystal, 2004). Both between-group and within-group effect sizes ( $d$ ) for change from baseline were calculated as an "unbiased  $d$ " (often called Hedge's  $g$ ), dividing the estimated (ITT) or raw (CC) mean change by the pooled standard deviation of the

**Table 1.** Sample characteristics across conditions.

	Imagery CBM ( <i>n</i> = 29), <i>M</i> [95% CIs]	PVT ( <i>n</i> = 26), <i>M</i> [95% CIs]
Age	22.34 [21.79, 22.90]	22.58 [21.84, 23.31]
Gender, <i>N</i> (%)		
Male	12 (41.4%)	8 (30.8%)
Female	17 (58.6%)	18 (69.2%)
QIDS	12.31 [11.17, 13.45]	12.15 [11.18, 13.13]
DASS		
Depression	14.21 [13.29, 15.12]	14.54 [13.61, 15.47]
Anxiety	14.41 [13.38, 15.44]	14.15 [13.12, 15.19]
Stress	14.45 [13.44, 15.46]	14.38 [13.59, 15.18]

Note. Imagery CBM = imagery cognitive bias modification; PVT = peripheral vision task; CI = confidence interval; QIDS = Quick Inventory of Depressive Symptomatology; DASS = Depression Anxiety Stress Scale (21-item version).

baseline scores, to provide an effect size estimate reflecting a standardized mean difference less affected by potentially biasing factors such as differential dropout or heterogeneity of treatment effects (Cumming, 2012). The 95% CIs were calculated using estimates from the mixed model (ITT) or the variance of the change scores rather than baseline scores (CC) to provide intervals consistent with the unstandardized mean differences. For both ITT and CC analyses, outcome calculations including associated effect sizes and 95% CIs were computed using RStudio (Version 1.2.1335; RStudio Team, 2018) running R version 3.6.1 (R Core Team (2019), 2019), with mixed model analyses conducted using nlme (Pinheiro et al., 2019). Analysis scripts and formulae for derivation of the effect sizes and 95% CIs can be found at <https://osf.io/u6pdv/>.

## Results

### Participants

Seventy participants attended an initial eligibility assessment, and 55 were judged eligible and subsequently randomized (20 men, 35 women, all Pakistani). Reasons for non-eligibility were not systematically recorded but included absence of laptop, receiving psychiatric treatment, high score on the QIDS suicidality item, and being unable to install the Java program. Table 1 presents the demographic characteristics and scores on additional baseline variables for participants, which appear balanced across the two groups. At baseline, 29 participants were allocated to imagery CBM and 26 to PVT. Five participants in the imagery CBM condition and nine in the PVT

condition dropped out between baseline and post-training, giving as their reason exams (*n* = 6), sickness (*n* = 2), technical problems with their laptop (*n* = 3), seeking treatment for depression (*n* = 2), or giving no reason (*n* = 1). Only 19 participants returned for the follow-up assessment (imagery CBM: 11, PVT: 8), with a large amount of dropout (13 participants in imagery CBM condition and 9 participants in the PVT condition) occurring toward the end of the recruitment period, because of students' final exams and the date for the follow-up assessment falling outside of the university term-time. See Figure 1, for a diagrammatic overview.

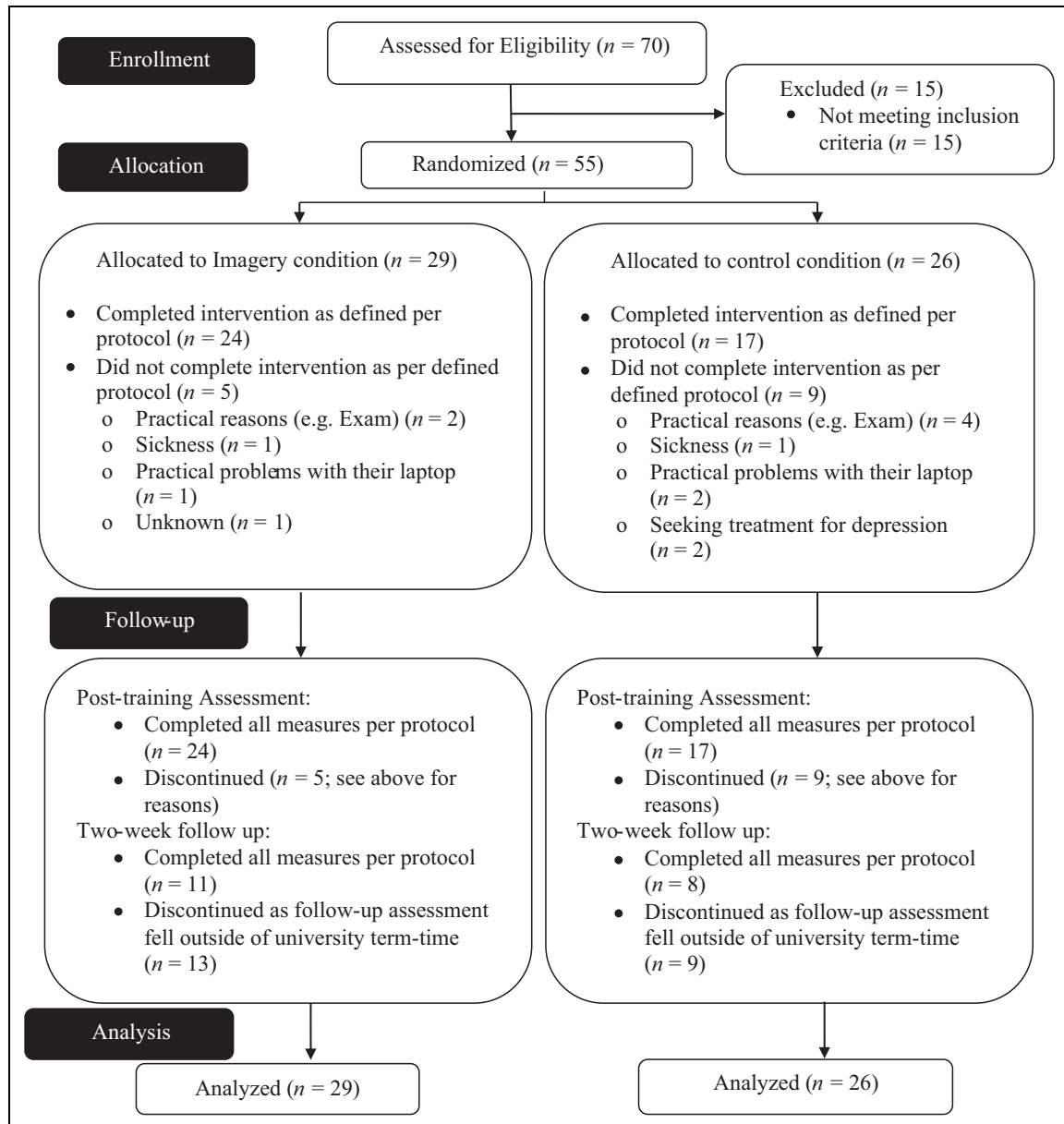
### Adherence

All participants returning for the post-training assessment had completed every session of their allocated training. For those participants dropping out prior to post-training assessment, adherence data (i.e., whether they had completed any training sessions) were not collected. Although the post-training assessment was planned to take place 1 week after baseline, in practice many participants requested for it to be moved to a later date as they were not able to complete the training sessions on a daily basis and wished to complete the schedule before returning. The mean number of days between baseline and post-training was 11.46, 95% CIs [9.49, 13.43] for the imagery CBM condition and 9.13 [7.61, 10.64] for the PVT condition. The mean number of days between post-training and follow-up was closer to the planned schedule (2 weeks, i.e., 14 days), but again with some variation: 15.69 [14.98, 16.41] for the imagery CBM condition and 15.88 [14.83, 16.92] for the PVT condition.

### Expectancy and feedback

Expectancy and feedback ratings are provided in Table 2. These indicate that in both groups expectancy prior to training was in the medium range for each question and equivalent across groups. After the end of training, feedback for the imagery CBM was relatively positive, where feedback appears more negative for the PVT. During debriefing, all participants correctly guessed their group allocation.

Feedback from participants in the imagery CBM condition included comments that they found it effective in providing relief from their depression and low mood and that they had learnt how to think positively in different situations in life, changing negative thoughts into positive ones. Several reported that



**Figure 1.** Flow of participants through the study.

participating in the research provided an entirely new experience for them as it helped them very much in understanding and dealing with their problems with a positive attitude. They reported that the study had helped them by teaching them how they could see things differently, from a different, positive, perspective, and that it is not necessary to have a negative approach toward current and upcoming life event. Some reported that they would try develop the ideas learned during the training themselves, via coming up with their own scenarios similar to those in the current study, starting ambiguously and resolving positively, to help them manage their depressed mood. They felt

that the imagery CBM program would be acceptable for Pakistani students and recommendable to other students with depressed mood. They further gave some suggestions for the improvement of program, for instance adding pictures and reducing the length of the sessions.

### *Clinical outcomes*

Table 3 presents descriptive statistics for the outcome measures across the three time points of the study for both ITT and CC populations. These include mean scores at each time point, and standardized effect

**Table 2.** Mean and 95% CIs of expectancy and feedback questionnaires.

	Imagery CBM, M [95% CIs]	PVT, M [95% CIs]
Expectancy	<i>n</i> = 29	<i>n</i> = 26
EQ Qu 1 (0–100%)	54.48 [48.95, 60.01]	58.08 [53.09, 63.06]
EQ Qu 2 (1–9)	6.28 [5.87, 6.68]	6.42 [5.99, 6.85]
EQ Qu 3 (0–100%)	59.66 [54.52, 64.79]	63.08 [58.00, 68.16]
Feedback questions (1–9 scale)	<i>n</i> = 24	<i>n</i> = 17
Difficulty/ease	6.04 [5.37, 6.72]	4.06 [3.20, 4.92]
Enjoyable	7.42 [6.89, 7.94]	5.00 [4.21, 5.79]
Useful	7.63 [7.23, 8.02]	4.82 [4.14, 5.51]
Positive effects	7.42 [6.86, 7.97]	3.59 [3.11, 4.07]
Distressing/burdensome	6.38 [5.72, 7.03]	4.41 [3.66, 5.16]
Continued usefulness	7.63 [7.20, 8.05]	—
Acceptability	7.54 [7.15, 7.94]	—
Confidence in recommending	7.58 [7.21, 7.96]	—
Willingness to try again	7.17 [6.64, 7.69]	—

Note. CI = confidence interval; Imagery CBM = imagery cognitive bias modification; PVT = peripheral vision task; EQ = Expectancy Questionnaire; EQ Qu 1 = “By the end of the study period, how much improvement in your low mood/depression symptoms do you think will occur?”; EQ Qu 2 = “At this point, how much do you really feel that the computer program will have an effect on improving your low mood/depression symptoms?”; EQ Qu 3 = “By end of the study period, how much improvement in your low mood/depression symptoms do you really feel will occur?”; (italics for EQ questions all original) Feedback questions were (in the order presented in the table): “How difficult or easy did you find the computer task?” (1 = extremely difficult to 9 = extremely easy); “How enjoyable did you find the task?” (1 = not at all enjoyable to 9 = extremely enjoyable); “How useful did you think the task was?” (1 = not at all useful to 9 = extremely useful); “Do you think that completing the computer tasks had any positive impact on you mood/thinking/behavior?” (1 = not at all to 9 = very much so); “How distressing or burdensome did you find the computer task?” (1 = not at all to 9 = extremely distressing/burdensome); Imagery CBM group only: “Do you think that completing more sessions of these tasks over a few weeks would help to improve your mood?” (1 = not at all to 9 = definitely); “If this program was available for students in Pakistan who were feeling depressed or down to do from home, how acceptable do you think they would find it to do?” (1 = not at all to 9 = extremely acceptable); “How confident would you be about recommending this program to a friend with depression?” (1 = extremely unconfident to 9 = extremely confident); “If you were feeling depressed or down in the future, would you be willing to try this program again?” (1 = extremely unlikely to 9 = extremely likely).

sizes for within-group change from baseline and between-group differences in change scores, with associated 95% CIs. Although the estimates generally favor the imagery CBM condition with large effect sizes, given the low internal consistencies for many of these measures in this sample, the high rate of dropout (particularly at follow-up), and the indication from the feedback that the PVT may have not provided an adequate “placebo” in this study, these data should be interpreted with caution. The corresponding data for the questionnaire subscales are presented in Online Supplementary Table 1.

## Discussion

In Pakistan, the prevalence of mental health problems such as depression and anxiety has been increasing in recent years. The lack of treatment facilities and culturally adapted psychological interventions means that it is a huge challenge to provide mental health services to much of the population of Pakistan. This

pilot study investigated the feasibility of using a computerized cognitive training procedure as a potential low cost, easily disseminable intervention, among a specific sample in Pakistan, university students. The intervention, imagery CBM had been developed from experimental psychopathology research and involved repeated generation of positive mental imagery. Overall, responses to the program were positive and it appeared to be acceptable for this new population. However, the study encountered a number of problems with attrition, adherence to the study schedule, reliability of the measures used, and credibility of the chosen comparison condition, indicating changes that should be made for future research in this area.

In line with previous research, the current study aimed to investigate a schedule comprising one session of imagery CBM daily for 1 week, with follow-up after 2 weeks post-training. In practice, there were some difficulties with implementing this schedule, with the post-training schedule often occurring later than 1-week post-baseline, and with high

rates of attrition particularly at follow-up. There were various reasons that the students struggled with this schedule, such as fitting it around other commitments, and the length of training sessions. It is possible that the rating for how “distressing/burdensome” participants found the imagery CBM (see Table 2) reflects these difficulties with the training schedule, given the otherwise generally positive feedback. The problems with the 1-week training schedule in this study contrast with the generally high rates of compliance with a similar schedule in previous studies (e.g., Blackwell & Holmes, 2010; Lang et al., 2012; Torkan et al., 2014) and could potentially reflect a high rate of daily demands or unpredictable stressors in the Pakistani student population. Future research in this population may benefit from a more flexible schedule of training with shorter individual training sessions. In relation to compliance with outcome assessments, collecting outcome data remotely, such as online or via telephone, rather than requiring participants to attend face-to-face assessments, may help increase completeness of outcome data. We note that the use of a 1-week training schedule was pragmatic step in the current study, using the schedule most often used in previous research, but there is no a priori reason to assume that it is the best for lasting benefits. One possible alternative would be to provide participants with continued access to the program after completion of an initial schedule, so that they could train until they felt it was no longer needed, or implement “booster sessions” if they felt their mood start to deteriorate (Blackwell & Holmes, 2017).

The simplest explanation for the low internal consistencies of the outcome measures at baseline may be language. It was decided to administer all the questionnaires in their original English form as many did not have validated Urdu versions, and university students in Pakistan would be expected to have a sufficient level of English to complete them. Interestingly, the internal consistencies of some of the measures at post-training were very much improved, which could reflect a practice or familiarity effect. This improvement does not appear to be an effect of differential dropout, as including only those participants who completed post-training measures does not result in much better internal consistencies for the baseline measurements. It is also possible that the poor internal consistencies at baseline reflect cultural differences in the expression of mental health; however, studies using Urdu translations of measures have previously found acceptable internal consistencies. Further, other

studies using English-language questionnaires in similar samples in Pakistan have found acceptable internal consistencies, including for the DASS used in the current study (Bibi, Blackwell, & Margraf, in press; Bibi, Lin, & Margraf, 2020). One difference between our sample and those in the other studies is that our sample had relatively elevated symptoms of depression. We therefore investigated whether this may have contributed to lower internal consistencies, using the publically available data set from Bibi, Blackwell, and Margraf (in press). Restricting this unselected student sample to include only those scoring  $>8$  on the DASS Depression scale (thus matching the depression severity in the current sample) resulted in the internal consistency for this scale dropping from  $\alpha = .85$  to  $\alpha = .46$ . Thus, the elevated symptom scores in the current pilot study (resulting from using a questionnaire cutoff to identify participants) could provide a reason for the discrepancy between the poor internal consistency found in this study and the acceptable levels found using the same measures in other studies. It is possible that a higher internal consistency in an unselected student sample could arise from the larger number of low-scoring participants, as these will tend to score consistently low across all items. Conversely, in a sample in which everyone has a relatively high score there may be a multitude of problematic symptom constellations, and any subtle differences in interpretation of the items might be amplified (due to, e.g., misunderstandings of the pragmatics of a second language; Asif et al., 2019). In this context, it is interesting that the BADS showed relatively acceptable internal consistencies consistently across assessments, potentially indicating fewer language or cultural-related problems with behavioral descriptions as opposed to internal states. However, these considerations remain speculative. In the future, it would be recommendable to use Urdu measures with demonstrated internal consistency and validity not only in unselected samples but also in clinical samples. Overall, the low internal consistencies in the current study call into question whether the total scale scores can be interpreted as reflecting a valid underlying construct.

The PVT was chosen as a comparison task to approximate a “placebo” control, that is, controlling for nonspecific aspects of the imagery CBM intervention but containing no “active ingredient.” However, it is questionable whether it provided an adequate placebo in the current study, given the relatively negative post-training ratings of the task and the fact that

all participants correctly guessed their group allocation when asked at the end of the study. Post-training evaluations are potentially difficult to interpret, as they are conflated with perceived benefit. That is, a task providing obvious benefits will of course be rated more positively and would be more likely to be judged as the “more effective” treatment. Further, as limited introductions to the tasks were provided with no rationale as to why they might be beneficial, they were most likely judged on their perceived benefits and face validity. However, if participants started to have doubts about the credibility and potential utility of the PVT early on in the training schedule, this would be likely to have an impact on their engagement with the task and any expectancy-mediated placebo effects. In fact, the only measure for which the results provided evidence of a potential benefit of the PVT from baseline to both post-training and follow-up (as indicated by the 95% CIs for the within-group effect sizes not including 0) was the BADS. This could potentially reflect taking part in the study and training schedule as being activating in themselves. Overall, it would be surprising to find between-group effect sizes as large as those displayed in Table 3 if the PVT had in fact provided an adequate placebo in this study. It may be more useful for future studies to address questions about clinical efficacy or utility of imagery CBM as an intervention, and thus in terms of a comparison condition, it would be worth comparing an imagery CBM intervention to alternative low-intensity interventions that could also be considered in this setting, for example, internet-based CBT or other forms of guided self-help. That is, such research could ask whether imagery CBM is better than alternative possibilities (Blackwell et al., 2017). A “treatment as usual” comparison condition would also be useful to ascertain the usefulness of introducing any new intervention, although in practice this may be no treatment in most cases.

Notwithstanding the caveats that have to be applied to interpreting the outcome data in the current study, both due to the problems outlined earlier, and also given the limitations of the short time-scale of follow-up and lack of blinding of researchers at the outcome assessments, the study indicates that it would be useful to continue to investigate the potential for imagery CBM as a low-intensity intervention among students in Pakistan with elevated symptoms of depression. In future research, practical steps such as a more flexible training and outcome measurement schedule, and using validated Urdu measures, would

help overcome some of the problems in the current study. It would also be useful to validate the Urdu-language training scenarios in relation to their structure (ambiguous start and positive resolution) given the changes in word order required when translating from English. A further limitation of the current study is that because the intervention was installed on participants’ laptops (as opposed to an internet-based implementation), we were unable to monitor compliance remotely, and our compliance data are based on participant self-report. Given that remotely delivered computerized interventions are known to be associated with potentially high rates of attrition (e.g., Linardon & Fuller-Tyszkiewicz, 2020), future studies should aim to incorporate more robust real-time compliance monitoring and measurement. Inclusion of session-by-session measures of engagement, rather than feedback only at the end of the study, would also help better inform potential reasons for dropout (e.g., finding the task boring, as may have been the case for our control condition). Resources permitting, it would be useful to supplement the current measurement schedule with other measures such as diagnostic interviews and clinician-ratings made by blind assessors, and also to assess broader outcomes such as academic success. Further, although we were not aware of any adverse events experienced by participants during their participation in the study, this was not systematically assessed and would need to be in a future trial. Finally, the current study was conceived as pilot work whose main purpose was to inform a subsequent pre-registered randomized controlled trial, and at the time of planning, we did not consider it itself to be a study requiring registration. However, any work involving data collection or analysis can benefit from preregistration and in hindsight it would have been preferable to preregister this pilot study.

Returning to the original motivation for this line of research, mental health problems are highly prevalent in Pakistan; for example, the prevalence rate of depression and anxiety is higher in comparison with other developing countries (Asad et al., 2010; Husain et al., 2011). However, there are few trained mental health professionals such as psychiatrists in Pakistan in relation to the large number of patients, it is considered humiliating and stigmatizing to visit mental health professionals for psychological help, and even medical professionals hold many misconceptions and biases toward patients with psychological disorders (Ahmad, 2007). In such a situation, where social stigma, myths, and hostile belief structures around

**Table 3.** Means and effect sizes with 95% CIs for outcome measures over the course of the study for both ITT and CC samples.

	Baseline (T1), M [95% CIs]	Post-training (T2), M [95% CIs]	Follow-up (T3), M [95% CIs]	Within-group d, M [95% CIs]		Between-group d, M [95% CIs]	
				T1 to T2	T1 to T3	T1 to T2	T1 to T3
<b>QIDS</b>							
ITT							
CBM	12.31 [11.28, 13.34]	9.34 [8.59, 10.09]	9.65 [8.32, 10.98]	1.05 [0.84, 1.63]	0.90 [0.41, 1.71]	1.16 [0.82, 1.98]	0.28 [-0.34, 0.69]
Control	12.15 [11.07, 13.24]	12.40 [11.54, 13.25]	10.30 [8.74, 11.85]	-0.09 [-0.49, 0.28]	0.60 [0.02, 1.38]		
CC							
CBM	12.31 [11.17, 13.45]	9.54 [8.86, 10.22]	9.82 [8.55, 11.09]	1.14 [0.82, 1.84]	1.47 [0.59, 2.86]	1.09 [0.80, 2.21]	0.57 [-0.43, 1.42]
Control	12.15 [11.18, 13.13]	12.76 [11.64, 13.89]	10.50 [8.09, 12.91]	0.06 [-0.33, 0.48]	0.89 [0.15, 1.88]		
<b>PANAS</b>							
ITT							
CBM	37.52 [35.41, 39.63]	48.61 [46.77, 50.45]	43.15 [40.50, 45.80]	1.91 [1.63, 2.78]	0.92 [0.48, 1.61]	1.93 [1.24, 2.48]	1.22 [0.31, 1.39]
Control	36.85 [34.62, 39.08]	36.86 [34.71, 39.00]	35.31 [32.22, 38.41]	0.00 [-0.45, 0.45]	-0.24 [-0.85, 0.31]		
CC							
CBM	37.52 [35.52, 39.51]	48.33 [46.32, 50.35]	41.91 [38.24, 45.58]	1.98 [1.52, 3.05]	0.97 [0.19, 1.98]	2.03 [1.49, 3.09]	1.47 [0.36, 2.41]
Control	36.85 [34.42, 39.27]	36.94 [34.84, 39.04]	35.12 [30.76, 39.49]	-0.02 [-0.51, 0.46]	-0.42 [-1.09, 0.19]		
<b>DARS</b>							
ITT							
CBM	58.00 [55.55, 60.45]	75.31 [71.55, 79.07]		2.55 [1.86, 3.19]		2.20 [1.59, 3.22]	
Control	56.19 [53.61, 58.78]	58.77 [54.58, 62.96]		0.37 [-0.04, 0.61]			
CC							
CBM	58.00 [55.76, 60.24]	74.54 [71.30, 77.78]		2.54 [1.73, 3.29]		2.16 [1.57, 3.20]	
Control	56.19 [53.28, 59.11]	60.18 [54.84, 65.52]		0.40 [-0.08, 0.69]			
<b>BADS</b>							
ITT							
CBM	13.38 [11.51, 15.25]	24.75 [22.87, 26.63]	26.18 [23.52, 28.85]	2.21 [1.98, 3.32]	2.38 [1.86, 3.41]	1.56 [0.80, 1.94]	1.53 [0.37, 1.46]
Control	16.12 [14.14, 18.09]	19.60 [17.42, 21.78]	20.95 [17.82, 24.08]	0.67 [0.22, 1.12]	0.86 [0.23, 1.43]		
CC							
CBM	13.38 [12.09, 14.67]	24.75 [22.63, 26.87]	25.55 [21.54, 29.55]	2.21 [1.71, 3.58]	2.55 [1.27, 4.35]	1.29 [0.68, 2.07]	1.22 [-0.07, 1.85]
Control	16.12 [13.58, 18.66]	18.53 [16.63, 20.43]	19.00 [14.62, 23.38]	0.92 [0.52, 1.34]	1.34 [0.32, 2.23]		
<b>DASS-D</b>							
ITT							
CBM	14.21 [13.32, 15.09]	9.02 [8.10, 9.93]	9.70 [8.34, 11.07]	2.13 [1.66, 2.74]	1.76 [1.11, 2.56]	2.22 [1.49, 2.79]	0.72 [-0.14, 0.90]
Control	14.54 [13.60, 15.47]	14.70 [13.65, 15.75]	11.81 [10.21, 13.41]	-0.06 [-0.50, 0.36]	1.03 [0.39, 2.08]		
CC							
CBM	14.21 [13.29, 15.12]	9.04 [8.07, 10.01]	9.82 [8.15, 11.49]	2.13 [1.48, 2.93]	2.03 [0.84, 3.35]	2.19 [1.65, 3.31]	0.49 [-0.57, 1.27]
Control	14.54 [13.61, 15.47]	14.82 [13.67, 15.98]	11.88 [10.12, 13.63]	-0.02 [-0.48, 0.43]	1.51 [0.49, 3.08]		

(continued)

Table 3. (continued)

	Baseline (T1), M [95% CIs]	Post-training (T2), M [95% CIs]	Follow-up (T3), M [95% CIs]	Within-group d, M [95% CIs]		Between-group d, M [95% CIs]	
				T1 to T2	T1 to T3	T1 to T2	T1 to T3
<b>DASS-A</b>							
ITT							
CBM	14.41 [13.42, 15.40]	9.38 [8.38, 10.38]	10.00 [8.05, 11.94]	1.85 [1.44, 2.48]	1.55 [0.76, 2.19]	1.95 [1.22, 2.45]	0.87 [-0.11, 0.93]
Control	14.15 [13.11, 15.20]	14.36 [13.20, 15.52]	12.14 [9.86, 14.41]	-0.07 [-0.53, 0.36]	0.68 [-0.10, 1.38]		
CC							
CBM	14.41 [13.38, 15.44]	9.25 [8.23, 10.27]	10.45 [8.28, 12.63]	1.80 [1.23, 2.59]	1.53 [0.46, 2.42]	2.05 [1.51, 3.13]	0.82 [-0.32, 1.55]
Control	14.15 [13.12, 15.19]	13.94 [12.70, 15.18]	12.00 [8.97, 15.03]	-0.21 [-0.62, 0.16]	0.72 [-0.14, 1.46]		
<b>DASS-S</b>							
ITT							
CBM	14.45 [13.57, 15.33]	9.20 [8.33, 10.07]	9.88 [8.61, 11.15]	2.16 [1.47, 2.39]	1.79 [1.14, 2.39]	2.48 [1.81, 3.20]	1.62 [0.43, 1.52]
Control	14.38 [13.45, 15.32]	15.08 [14.08, 16.08]	13.80 [12.31, 15.28]	-0.28 [-1.01, 0.15]	0.22 [-0.49, 1.07]		
CC							
CBM	14.45 [13.44, 15.46]	9.25 [8.08, 10.42]	10.27 [8.60, 11.95]	2.18 [1.30, 2.59]	1.75 [0.74, 2.68]	2.52 [2.01, 3.80]	1.53 [0.20, 2.19]
Control	14.38 [13.59, 15.18]	15.00 [14.40, 15.60]	14.00 [12.27, 15.73]	-0.31 [-1.07, 0.14]	0.28 [-0.87, 1.59]		
<b>AST</b>							
ITT							
CBM	-17.34 [-21.80, -12.89]	10.93 [6.70, 15.15]		2.29 [1.78, 3.10]		2.43 [1.86, 3.57]	
Control	-17.58 [-22.28, -12.87]	-18.89 [-23.79, -13.98]		-0.10 [-0.57, 0.34]			
CC							
CBM	-17.34 [-22.02, -12.67]	10.46 [5.90, 15.02]		2.34 [1.63, 3.34]		2.41 [1.67, 3.34]	
Control	-17.58 [-22.22, -12.93]	-19.59 [-25.04, -14.14]		-0.03 [-0.39, 0.32]			
<b>SST</b>							
ITT							
CBM	0.64 [0.58, 0.69]	0.32 [0.28, 0.36]		2.21 [2.11, 3.60]		1.97 [1.78, 3.47]	
Control	0.58 [0.53, 0.64]	0.55 [0.50, 0.59]		0.26 [-0.09, 0.65]		2.00 [1.65, 3.31]	
CC							
CBM	0.64 [0.58, 0.69]	0.32 [0.29, 0.35]		2.21 [1.95, 3.77]			
Control	0.58 [0.53, 0.64]	0.54 [0.48, 0.61]		0.24 [-0.19, 0.70]			
<b>PIT positive</b>							
ITT							
CBM	16.52 [15.57, 17.47]	19.25 [18.52, 19.99]		1.04 [0.78, 1.88]		1.23 [0.68, 2.04]	
Control	17.00 [16.00, 18.00]	16.55 [15.69, 17.41]		-0.17 [-0.63, 0.26]			
CC							
CBM	16.52 [15.59, 17.45]	19.12 [18.47, 19.78]		1.17 [0.94, 2.07]		1.46 [0.86, 2.28]	
Control	17.00 [15.93, 18.07]	16.65 [15.57, 17.72]		-0.26 [-0.90, 0.32]			

(continued)



**Table 3.** (continued)

	Baseline (T1), M [95% CIs]	Post-training (T2), M [95% CIs]	Follow-up (T3), M [95% CIs]	Within-group d, M [95% CIs]		Between-group d, M [95% CIs]	
				T1 to T2	T1 to T3	T1 to T2	T1 to T3
PIT Negative							
ITT							
CBM	15.21 [14.19, 16.23]	18.63 [17.78, 19.47]		1.21 [0.90, 1.80]		0.78 [0.36, 1.67]	
Control	14.65 [13.58, 15.73]	15.91 [14.94, 16.88]		0.44 [0.10, 0.98]			
CC							
CBM	15.21 [14.21, 16.20]	18.67 [17.64, 19.69]		1.20 [0.83, 1.83]		0.75 [0.27, 1.58]	
Control	14.65 [13.51, 15.80]	15.88 [15.05, 16.71]		0.45 [-0.01, 1.11]			

Note. CI = confidence interval; ITT = intention to treat; CC = complete case (using all available data, i.e., the mean presented at pretraining is all participants, the mean presented at post-training is from all participants who provided post-training data, and the mean presented at follow-up is from all participants who provided data at follow up. This means effect size estimates may appear not to match the means presented, as the effect size estimates only include participants who provided data at both relevant time points); QIDS = Quick Inventory of Depressive Symptomatology; CBM = cognitive bias modification; PANAS = Positive and Negative Affect Schedule; DARS = Dimensional Anhedonia Rating Scale; BADS = Behavioral Activation for Depression Scale; DASS-D/A/S = Depression/Anxiety/Stress scale of the Depression Anxiety and Stress Scale; AST = ambiguous scenarios task; SST = scrambled sentences test (negativity score); PIT-positive/negative = Vividness ratings for positive/negative items on the prospective imagery test. An effect size with a positive sign indicates an improvement (for within-group effect sizes) or greater improvement in the imagery CBM than the PVT group (for between-group effect sizes), although note that for PIT-negative it is not clear that a change in either direction is positive or negative.

mental illness make it difficult for a person to seek help from mental health professionals, low-intensity computerized interventions that could be completed from home may be particularly useful. In the longer term, it would of course be preferable to tackle the conditions that lead to such a high prevalence of depression, increase numbers of trained mental health professionals, and reduce stigma around mental health. However, at the current time, reducing the suffering of individuals via making simple interventions available could provide welcome relief.

At a broader level, this pilot study can serve as an example for attempting to translate cognitive training interventions to other countries and cultures and raises some general recommendations. First, in any new context, a pilot study should be considered an essential first step, as otherwise potential problems will not be uncovered in time to allow necessary modifications. Where possible, all aspects of a planned future trial should be piloted, including the intervention, control condition, measures, assessment schedule, and any other procedures related to trial conduct (Leon et al., 2011). Some potential problems are also likely to be detectable with even a small number of participants, for example, via a single case series, which would likely be a useful precursor to a pilot study using a randomized design (e.g., as in the original translation of imagery CBM to a clinical sample; Blackwell & Holmes, 2010). Second, the content of the intervention needs to be checked and potentially adapted for cultural acceptability (e.g., adaptation of training scenarios as in the current study). Third, detailed collection of both quantitative and qualitative feedback and data related to acceptability and attrition is recommended to detect and allow full exploration of any potential problems. Fourth, main outcome measures that have been shown to have acceptable psychometric properties and validity within a treatment context (i.e., not just general population or unselected sample) should be used where possible, and indices of reliability such as internal consistency calculated and reported for both well-established and any novel untested measures used (e.g., Parsons et al., 2019). Finally, publication of pilot work can be highly valuable in raising awareness of potential problems or considerations in novel applications of interventions, and hence publication should be one aim of such piloting (whether this is a randomized study or single case work). Steps to reduce bias and increase transparency, such as preregistration, would therefore also be recommended.

In conclusion, the current pilot study suggests that imagery CBM warrants further investigation as an easily accessible computerized intervention for symptoms of depression among students in Pakistan. The current study does not allow any confident conclusions to be drawn about the likely efficacy of imagery CBM in this population, due to problems with the measures, attrition, and the control comparison. However, it provides a platform for future research in this population using these paradigms and indicates potentially valuable routes for further investigations. Finally, the study highlights the importance of pilot studies when moving paradigms to new samples to detect potential problems that may otherwise compromise subsequent hypothesis-testing research.

### Availability of data and material

The data sets generated and analyzed during the current study and analysis scripts are available in the Open Science Framework repository, with the exception of individual-level demographic data due to concerns about potential identifiability (available on request): <https://osf.io/u6pdv/>. Materials are available from the corresponding author on request.

### Acknowledgments

The authors thank Raheela Hayat and Wajiha Kanawal for helping with data collection. They also thank Greg J. Siegle for providing example computer scripts for the Peripheral Vision Task.



### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Akhtar Bibi was supported by HEC/DAAD. Simon E. Blackwell and Jürgen Margraf were employed by Ruhr-Universität Bochum. The authors acknowledge support by the DFG Open Access Publication Funds of the Ruhr-Universität Bochum. The funders had no role in the design of the study, collection, analysis, and interpretation of data, or in the writing of the manuscript.

### ORCID iDs

Jürgen Margraf  <https://orcid.org/0000-0001-5207-7016>  
Simon E. Blackwell  <https://orcid.org/0000-0002-3313-7084>

## Supplemental material

Supplemental material for this article is available online.

## Note

1. Due to a misreading of the paper by Rohrbacher and Reinecke, in our sample, the “A” and “B” versions used did not correspond to those specified in their paper but rather the first 15 (“A”) and last 15 (“B”) scenarios listed in their Appendix.

## References

- Ahmad, K. (2007). Shortage of psychiatrists a problem in Pakistan. *Lancet (London, England)*, *370*(9590), 817–818. [https://doi.org/10.1016/S0140-6736\(07\)61397-X](https://doi.org/10.1016/S0140-6736(07)61397-X)
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Author.
- Asad, N., Karmaliani, R., Sullaiman, N., Bann, C. M., McClure, E. M., Pasha, O., Wright, L. L., & Goldenberg, R. L. (2010). Prevalence of suicidal thoughts and attempts among pregnant Pakistani women. *Acta Obstetrica et Gynecologica Scandinavica*, *89*(12), 1545–1551. <https://doi.org/10.3109/00016349.2010.526185>
- Asif, M., Deng, Z., Hussain, Z., & Rasool, S. (2019). The case study of pragmatic failure in second language of Pakistani students. *International Journal of English Linguistics*, *9*(4), 200. <https://doi.org/10.5539/ijel.v9n4p200>
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press.
- Bibi, A., Blackwell, S. E., & Margraf, J. (in press). Mental health, suicidal ideation, and experience of bullying among university students in Pakistan. *Journal of Health Psychology*. <https://doi.org/10.1177/1359105319869819>
- Bibi, A., Lin, M., & Margraf, J. (2020). Salutogenic constructs across Pakistan and Germany: A cross sectional study. *International Journal of Clinical and Health Psychology*, *20*(1), 1–9. <https://doi.org/10.1016/j.ijchp.2019.10.001>
- Blackwell, S. E., Browning, M., Mathews, A., Pictet, A., Welch, J., Davies, J., Watson, P., Geddes, J. R., & Holmes, E. A. (2015). Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: A randomized controlled trial. *Clinical Psychological Science*, *3*(1), 91–111. <https://doi.org/10.1177/2167702614560746>
- Blackwell, S. E., & Holmes, E. A. (2010). Modifying interpretation and imagination in clinical depression: A single case series using cognitive bias modification. *Applied Cognitive Psychology*, *24*(3), 338–350. <https://doi.org/10.1002/acp.1680>
- Blackwell, S. E., & Holmes, E. A. (2017). Brightening the day with flashes of positive mental imagery: A case study of an individual with depression. *Journal of Clinical Psychology*, *73*(5), 579–589. <https://doi.org/10.1002/jclp.22455>
- Blackwell, S. E., Rius-Ottenheim, N., Schulte-van Maaren, Y. W. M., Carlier, I. V. E., Middelkoop, V. D., Zitman, F. G., Spinhoven, P., Holmes, E. A., & Giltay, E. J. (2013). Optimism and mental imagery: A possible cognitive marker to promote well-being? *Psychiatry Research*, *206*(1), 56–61. <https://doi.org/10.1016/j.psychres.2012.09.047>
- Blackwell, S. E., Westermann, K., Woud, M. L., Cwik, J. C., Neher, T., Graz, C., Nyhuis, P. W., & Margraf, J. (2018). Computerized positive mental imagery training versus cognitive control training versus treatment as usual in inpatient mental health settings: Study protocol for a randomized controlled feasibility trial. *Pilot and Feasibility Studies*, *4*(1), 133. <https://doi.org/10.1186/s40814-018-0325-1>
- Blackwell, S. E., Woud, M. L., & MacLeod, C. (2017). A question of control? Examining the role of control conditions in experimental psychopathology using the example of cognitive bias modification research. *Spanish Journal of Psychology*, *20*, e54. <https://doi.org/10.1017/sjp.2017.41>
- Bowler, J. O., Mackintosh, B., Dunn, B. D., Mathews, A., Dalgleish, T., & Hoppitt, L. (2012). A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: Effects on anxiety, depression, attentional control, and interpretive bias. *Journal of Consulting and Clinical Psychology*, *80*(6), 1021–1033. <https://doi.org/10.1037/a0029932>
- Bukhari, S. R., & Khanam, S. J. (2015). Prevalence of depression in university students belonging to different socioeconomic status. *Journal of Postgraduate Medical Institute*, *29*, 156–159.
- Butler, G., & Mathews, A. (1983). Cognitive-processes in anxiety neurosis. *Bulletin of the British Psychological Society*, *36*(MAY), A49–A49.
- Calkins, A. W., McMorrin, K. E., Siegle, G. J., & Otto, M. W. (2015). The effects of computerized cognitive control training on community adults with depressed mood. *Behavioural and Cognitive Psychotherapy*, *43*(5), 578–589. <https://doi.org/10.1017/S1352465814000046>
- Chisholm, D., Flisher, A., Lund, C., Patel, V., Saxena, S., Thornicroft, G., & Tomlinson, M. (2007). Scale up services for mental disorders: A call for action. *Lancet*,

- 370(9594), 1241–1252. [https://doi.org/10.1016/S0140-6736\(07\)61242-2](https://doi.org/10.1016/S0140-6736(07)61242-2)
- Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., & Dour, H. J. (2016). Treatment for Anhedonia: A neuroscience driven approach. *Depression and Anxiety*, 33(10), 927–938. <https://doi.org/10.1002/da.22490>
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, 31(2), 73–86. [https://doi.org/10.1016/S0005-7916\(00\)00012-4](https://doi.org/10.1016/S0005-7916(00)00012-4)
- Dunn, B. D. (2012). Helping depressed clients reconnect to positive emotional experience: Current insights and future directions. *Clinical Psychology and Psychotherapy*, 19(4), 326–340. <https://doi.org/10.1002/cpp.1799>
- Eisenberg, D., Gollust, S. E., Golberstein, E., & Hefner, J. L. (2007). Prevalence and correlates of depression, anxiety, and suicidality among university students. *American Journal of Orthopsychiatry*, 77(4), 534–542. <https://doi.org/10.1037/0002-9432.77.4.534>
- Everaert, J., Podina, I. R., & Koster, E. H. W. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, 58, 33–48. <https://doi.org/10.1016/j.cpr.2017.09.005>
- Georguieva, R., & Krystal, J. H. (2004). Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. *Archives of General Psychiatry*, 61(3), 310–317.
- Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 44(2), 227–239. <https://doi.org/10.1348/014466505X29657>
- Hitchcock, C., Werner-Seidler, A., Blackwell, S. E., & Dalgleish, T. (2017). Autobiographical episodic memory-based training for the treatment of mood, anxiety and stress-related disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, 52, 92–107. <https://doi.org/10.1016/j.cpr.2016.12.003>
- Holmes, E. A., Blackwell, S. E., Burnett Heyes, S., Renner, F., & Raes, F. (2016). Mental imagery in depression: Phenomenology, potential mechanisms, and treatment implications. *Annual Review of Clinical Psychology*, 12, 249–280. <https://doi.org/10.1146/annurev-clinpsy-021815-092925>
- Holmes, E. A., Lang, T. J., & Shah, D. M. (2009). Developing interpretation bias modification as a “cognitive vaccine” for depressed mood—Imagining positive events makes you feel better than thinking about them verbally. *Journal of Abnormal Psychology*, 118(1), 76–88. <https://doi.org/10.1037/a0012590>
- Holmes, E. A., & Mathews, A. (2005). Mental imagery and emotion: A special relationship? *Emotion*, 5(4), 489–497. <https://doi.org/10.1037/1528-3542.5.4.489>
- Holmes, E. A., Mathews, A., Dalgleish, T., & Mackintosh, B. (2006). Positive interpretation training: Effects of mental imagery versus verbal training on positive mood. *Behavior Therapy*, 37(3), 237–247. <https://doi.org/10.1016/j.beth.2006.02.002>
- Husain, N., Chaudhry, N., Tomenson, B., Jackson, J., Gater, R., & Creed, F. (2011). Depressive disorder and social stress in Pakistan compared to people of Pakistani origin in the UK. *Social Psychiatry and Psychiatric Epidemiology*, 46(11), 1153–1159. <https://doi.org/10.1007/s00127-010-0279-y>
- Ji, J. L., Holmes, E. A., & Blackwell, S. E. (2017). Seeing light at the end of the tunnel: Positive prospective mental imagery and optimism in depression. *Psychiatry Research*, 247, 155–162. <https://doi.org/10.1016/j.psychres.2016.11.025>
- Jooma, R., Minhas, F. A., & Saxena, S. (2009). *WHO-AIMS report on mental health institutes in Pakistan*. World Health Organisation Office, Islamabad, Pakistan. Retrieved August 29, 2017, from [https://www.who.int/mental\\_health/pakistan\\_who\\_aims\\_report.pdf](https://www.who.int/mental_health/pakistan_who_aims_report.pdf)
- Kakuma, R., Minas, H., Van Ginneken, N., Dal Poz, M. R., Desiraju, K., Morris, J. E., Saxena, S., & Scheffler, R. M. (2011). Human resources for mental health care: Current situation and strategies for action. *The Lancet*, 378(9803), 1654–1663. [https://doi.org/10.1016/S0140-6736\(11\)61093-3](https://doi.org/10.1016/S0140-6736(11)61093-3)
- Kanter, J. W., Mulick, P. S., Busch, A. M., Berlin, K. S., & Martell, C. R. (2007). The Behavioral Activation for Depression Scale (BADs): Psychometric properties and factor structure. *Journal of Psychopathology and Behavioral Assessment*, 29, 191–202. <https://doi.org/10.1007/s10862-006-9038-5>
- Koster, E. H. W., Fox, E., & MacLeod, C. (2009). Introduction to the special section on cognitive bias modification in emotional disorders. *Journal of Abnormal Psychology*, 118(1), 1–4. <https://doi.org/10.1037/a0014379>
- Lang, T. J., Blackwell, S. E., Harmer, C. J., Davison, P., & Holmes, E. A. (2012). Cognitive bias modification using mental imagery for depression: Developing a novel computerized intervention to change negative thinking

- styles. *European Journal of Personality*, 26(2), 145–157. <https://doi.org/10.1002/per.855>
- Leon, A. C., Davies, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, 45(5), 626–629. <https://doi.org/10.1016/j.jpsychires.2010.10.008>
- Linardon, J., & Fuller-Tyszkiewicz, M. (2020). Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal of Consulting and Clinical Psychology*, 88(1), 1–13. <https://doi.org/10.1037/ccp0000459>
- Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales* (2nd ed.). Psychology Foundation.
- Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology*, 109(4), 602–615. <https://doi.org/10.1037/0021-843X.109.4.602>
- Naeem, F., Ayub, M., Gobi, M., & Kingdon, D. (2009). Development of Southampton adaptation framework for CBT (SAF-CBT): A framework for adaptation of CBT in non-western culture. *Journal of Pakistan Psychiatric Society*, 6(2), 79–84.
- Naeem, F., Ayub, M., Javed, Z., Irfan, M., Haral, F., & Kingdon, D. (2006). Stigma and psychiatric illness. A survey of attitude of medical students and doctors in Lahore, Pakistan. *Journal of Ayub Medical College, Abbottabad : JAMC*, 18(3), 46–49.
- Parsons, S., Kruijt, A. W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science*, 2(4), 378–395. <https://doi.org/10.1177/2515245919879695>
- Phillips, W. J., Hine, D. W., & Thorsteinsson, E. B. (2010). Implicit cognition and depression: A meta-analysis. *Clinical Psychology Review*, 30(6), 691–709. <https://doi.org/10.1016/j.cpr.2010.05.002>
- Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D., (2019). *nlme: Linear and nonlinear mixed effects Models*. R package version 3.1-140, Retrieved June 11, 2019, from <https://CRAN.R-project.org/package=nlme>
- Pictet, A., Jermann, F., & Ceschi, G. (2016). When less could be more: Investigating the effects of a brief internet-based imagery cognitive bias modification intervention in depression. *Behaviour Research and Therapy*, 84, 45–51. <https://doi.org/10.1016/j.brat.2016.07.008>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved August 27, 2019, from <https://www.R-project.org/>
- Reilly, T. J., MacGillivray, S. A., Reid, I. C., & Cameron, I. M. (2015). Psychometric properties of the 16-item quick inventory of depressive symptomatology: A systematic review and meta-analysis. *Journal of Psychiatric Research*, 60, 132–140. <https://doi.org/10.1016/j.jpsychires.2014.09.008>
- Renner, F., Ji, J. L., Pictet, A., Holmes, E. A., & Blackwell, S. E. (2017). Effects of engaging in repeated mental imagery of future positive events on behavioural activation in individuals with major depressive disorder. *Cognitive Therapy and Research*, 41, 369–380. <https://doi.org/10.1007/s10608-016-9776-y>
- Rizvi, S. J., Quilty, L. C., Sproule, B. A., Cyriac, A., Bagby, M. R., & Kennedy, S. H. (2015). Development and validation of the Dimensional Anhedonia Rating Scale (DARS) in a community sample and individuals with major depression. *Psychiatry Research*, 229(1–2), 109–119. <https://doi.org/10.1016/j.psychres.2015.07.062>
- Rohrbacher, H., & Reinecke, A. (2014). Measuring change in depression-related interpretation bias: Development and validation of a parallel ambiguous scenarios test. *Cognitive Behaviour Therapy*, 43(3), 239–250. <https://doi.org/10.1080/16506073.2014.919605>
- Rude, S. S., Wenzlaff, R. M., Gibbs, B., Vane, J., & Whitney, T. (2002). Negative processing biases predict subsequent depressive symptoms. *Cognition and Emotion*, 16(3), 423–440.
- Rush, J. A., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, 54(5), 573–583. [https://doi.org/10.1016/S0006-3223\(02\)01866-8](https://doi.org/10.1016/S0006-3223(02)01866-8)
- RStudio Team. (2018). *RStudio: Integrated development for R*. RStudio, Inc. Retrieved June 11, 2019, from <http://www.rstudio.com/>
- Saleem, S., Mahmood, Z., & Naz, M. (2013). Mental health problems in university students: A prevalence study. *FWU Journal of Social Sciences*, 7(2), 124–130.
- Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: Scarcity, inequity, and inefficiency. *Lancet*, 370(9590), 878–889. [https://doi.org/10.1016/S0140-6736\(07\)61239-2](https://doi.org/10.1016/S0140-6736(07)61239-2)

- Scholten, S., Velten, J., Bieda, A., Zhang, X. C., & Margraf, J. (2017). Testing measurement invariance of the depression, anxiety, and stress scales (DASS-21) across four countries. *Psychological Assessment, 29*(11), 1376–1390. <https://doi.org/10.1037/pas0000440>
- Shah, M., Bukhsh Sajid, W., Sabih, F., & Shoaib Hanif, M. (2011). Pattern of somatic symptoms in anxiety and depression. *Pakistan Armed Forces Medical Journal, 61*(3), 447–452.
- Stöber, J. (2000). Prospective cognitions in anxiety and depression: Replication and methodological extension. *Cognition & Emotion, 14*(5), 725–729. <https://doi.org/10.1080/02699930050117693>
- Sultan, S. (2011). Stigmatization: Addressing self-esteem and personal growth in patients with psychological and physiological illness. *Pakistan Journal of Social Sciences, 31*(1), 29–36.
- Torkan, H., Blackwell, S. E., Holmes, E. A., Kalantari, M., Neshat-Doost, H. T., Maroufi, M., & Talebi, H. (2014). Positive imagery cognitive bias modification in treatment-seeking patients with major depression in Iran: A pilot study. *Cognitive Therapy and Research, 38*(2). <https://doi.org/10.1007/s10608-014-9598-8>
- Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., Crismon, M. L., Shores-Wilson, K., Toprac, M. G., Dennehy, E. B., Witte, B., & Kashner, T. M. (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: A psychometric evaluation. *Psychological Medicine, 34*(1), 73–82. <https://doi.org/doi:10.1017/S0033291703001107>
- United Nations (2015). Transforming our world: The 2030 Agenda for sustainable development. United Nations. Retrieved August 29, 2017, from <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>
- Votruba, N., Thornicroft, G., & FundaMentalSDG Steering Group. (2016). Sustainable development goals and mental health: Learnings from the contribution of the FundaMentalSDG global initiative. *Global Mental Health (Cambridge, England), 3*, e26. <https://doi.org/10.1017/gmh.2016.20>
- Watson, D., & Clark, L. A. (1994). *The PANAS-X—Manual for the positive and negative affect schedule—Expanded form*. Retrieved October 13, 2011, from [https://ir.uiowa.edu/psychology\\_pubs/11/](https://ir.uiowa.edu/psychology_pubs/11/)
- Wenzlaff, R. M. (1993). The mental control of depression: Psychological obstacles to emotional well-being. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 239–257). Prentice Hall.
- Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J. L., & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the global burden of disease study 2010. *The Lancet, 382*(9904), 1575–1586. [https://doi.org/10.1016/S0140-6736\(13\)61611-6](https://doi.org/10.1016/S0140-6736(13)61611-6)
- Williams, A. D., Blackwell, S. E., Mackenzie, A., Holmes, E. A., & Andrews, G. (2013). Combining imagination and reason in the treatment of depression: A randomized controlled trial of internet-based cognitive-bias modification and internet-CBT for depression. *Journal of Consulting and Clinical Psychology, 81*(5), 793–799. <https://doi.org/10.1037/a0033247>
- Williams, A. D., O'Moore, K., Blackwell, S. E., Smith, J., Holmes, E. A., & Andrews, G. (2015). Positive imagery cognitive bias modification (CBM) and internet-based cognitive behavioral therapy (iCBT): A randomized controlled trial. *Journal of Affective Disorders, 178*, 131–141. <https://doi.org/10.1016/j.jad.2015.02.026>
- Woud, M. L., & Becker, E. S. (2014). Editorial for the special issue on cognitive bias modification techniques: An introduction to a time traveller's tale. *Cognitive Therapy and Research, 38*(2), 83–88. <https://doi.org/10.1007/s10608-014-9605-0>

### Author biographies

**Akhtar Bibi** is a PhD scholar in the Department of Clinical Psychology and Psychotherapy, Mental Health Research and Treatment Center at Ruhr-Universität Bochum. Her current research focuses on positive mental health, cognitive biases, and psychological treatments.

**Jürgen Margraf** is Alexander von Humboldt-Professor of Clinical Psychology and Psychotherapy as well as co-director of the Mental Health Research and Treatment Center at Ruhr-Universität Bochum. His current research focuses on positive mental health, emotional disorders and psychological treatments.

**Simon E. Blackwell** is a postdoctoral researcher in the Department of Clinical Psychology and Psychotherapy, Mental Health Research and Treatment Center at Ruhr-Universität Bochum. His current research focuses on positive mental imagery and translational research methods.