Positive Imagery Cognitive Bias Modification for Symptoms of Depression amongst University Students in Pakistan: A Pilot Study

Akhtar Bibi

Jürgen Margraf

Simon E. Blackwell

Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Bochum, Germany

Author note:

Akhtar Bibi, Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Massenbergstraße 9-13, 44787 Bochum, Germany, email: akhtar.bibi@rub.de; Simon E. Blackwell, Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Massenbergstraße 9-13, 44787 Bochum, Germany, email: simon.blackwell@rub.de; Jürgen Margraf, Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Massenbergstraße 9-13, 44787 Bochum, Germany, email: juergen.margraf@rub.de

Correspondence concerning this article should be addressed to Simon E. Blackwell, Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Massenbergstrasse 9-13, 44787 Bochum, Germany. Email address:simon.blackwell@rub.de

Abstract

Depression is highly prevalent amongst 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 amongst university students in Pakistan. Fifty-five participants scoring above a questionnaire cut-off 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 one week. Outcomes were measured at post-training and a subsequent two-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: Mental imagery, cognitive bias modification, depression, computerized intervention, interpretive bias

Positive Imagery Cognitive Bias Modification for Symptoms of Depression amongst Students in Pakistan: A Pilot Study

Access to mental health care is considered to be one of the most significant factors in the worldwide sustainable development program (Votruba, Thornicroft, & FundaMentalSDG Steering Group, 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, Thornicroft, Knapp, & Whiteford, 2007; Whiteford et al., 2013). Only specialized urban healthcare settings have facilities for providing treatments such as pharmacotherapy and psychotherapy (Jooma, Minhas, & Saxena, 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, Ayub, Gobi, & Kingdon, 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, in press). 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, Gollust, Golberstein, & Hefner, 2007). This is also the case in Pakistan, with a prevalence study of mental health problems amongst university students in Pakistan finding that 31% of participants experienced 'severe' mental health problems, and 16% experienced 'very severe' mental health problems (Saleem, Mahmood, & Naz, 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, in press).

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 due to 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, Fox, & MacLeod, 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, Blackwell, Burnett Heyes, Renner, & Raes, 2016) and negative interpretation bias (Butler & Mathews, 1983; Everaert, Podina, & Koster, 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, Lang, & Shah, 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 one

week (Blackwell & Holmes, 2010; Williams, Blackwell, Mackenzie, Holmes, & Andrews, 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, Werner-Seidler, Blackwell, & Dalgleish, 2017). The majority of studies investigating the potential of some variant of a one-week training schedule to reduce symptoms of depression have found superiority over a closely matched 'sham training' control condition (Lang, Blackwell, Harmer, Davison, & Holmes, 2012; Pictet, Jermann, & Ceschi, 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' analysis comprising those participants who completed the training schedule (Williams et al., 2015). However, the largest 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 four-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 amongst university students in Pakistan, using the

most commonly-applied one week training schedule and two week follow-up as a starting point. While imagery CBM has been successfully investigated in a variety of countries, namely the UK (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, Velten, Bieda, Zhang, & Margraf, 2017). For example, in Pakistan people with depression may often present their psychological problems in the form of bodily symptoms (Shah, Bukhsh Sajid, Sabih, & Shoaib Hanif, 2011). Previous studies have also not specifically targeted a student population (although many have included students amongst 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 one-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 (e.g., Craske, Meuret, Ritz, Treanor, & Dour, 2016; Dunn et al., in press). 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 behavioural activation, following research suggesting that imagery CBM may also help increase behavioural activation, that is, purposeful engagement in meaningful activities in everyday life (Renner, Ji, Pictet, Holmes, & Blackwell, 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 non-specific 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 non-specific effects that may be associated with a computerized cognitive training intervention (Blackwell, Woud, & MacLeod, 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', which engages spatial attention. Such a task had been used as a control condition in studies investigating other cognitive training paradigms in the context of depression (Calkins, McMorran, Siegle, & Otto, 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. A recruitment window of 12 weeks between October and December 2016 was available, and recruitment stopped after this time. Participants were recruited from Pakistani universities in the area (National University of Modern Languages (NUML), 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 (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: 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 colour-blindness, or severe visual or auditory impairment; 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/ pre-training assessment.

Assessment sessions took place at NUML, and were conducted by a doctoral student from Ruhr-Universität Bochum (AB). 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 Peripheral Vision Task (PVT; control condition). Outcome measures were scheduled for assessment at baseline (pre-training), after one week of training (post-training), and at a two-week follow-up.

Interventions

Imagery Cognitive Bias Modification (Imagery CBM). The imagery CBM intervention was adapted from that developed in previous experimental (e.g., Holmes et al., 2009) and

clinical (e.g., 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 in order 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. 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., 2009; Holmes, Mathews, Dalgleish, & Mackintosh, 2006), 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 (AB) with the help of two other bilingual (English and Urdu) experts. 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 towards 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 towards the end of the sentence. The contents of the previously-used training scenarios was 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 to 30 seconds. 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 seconds), and the desire to keep the length of individual sessions below 30 mins.

The imagery CBM program was implemented as a desktop application written in Java by the last author (SEB). While 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 pre-training assessment.

Peripheral Vision Task. The Peripheral Vison Task (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 centre. 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 to 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 coloured. After a further 1.5 seconds, the display would disappear and participants were asked to report the color of circle via the keyboard. Feedback ("Correct!" or "Wrong!") was provided immediately afterwards. 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 (SEB), 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. Administering questionnaires (and other study materials) in English (as also carried out in other related studies, e.g. Bibi, Blackwell, & Margraf, in press) was not anticipated to be a problem, as 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. Where figures for internal consistency are provided below, these are all Cronbach's alpha unless otherwise stated. As the study was not intended for efficacy testing, measures were not pre-specified as primary or secondary outcomes.

Outcome measures.

The Quick Inventory of Depressive Symptomatology (Rush et al., 2003). The Quick Inventory of Depressive Symptomatology (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 (DSM-IV; American Psychiatric Association, 1994). Participants are asked to rate each item according to how they have felt over the past seven days. The QIDS generally has a high internal consistency, $\alpha = .86$ (Trivedi et al., 2004). 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 (Kanter, Mulick, Busch, Berlin, & Martell, 2007).The Behavioural Activation for Depression scale (BADS) is used to measure engagement in functional and dysfunctional depression-relevant behaviours (e.g. behavioural engagement versus avoidance) over the previous week, covering the following areas: activation, avoidance/rumination, work/school impairment, and social impairment. In the current study, the short 9-item version of BADS was used (Kanter, Mulick, Busch, Berlin, & Martell, 2007). Each item is rated on seven point scale from 0 (not at all) to 6 (completely). Internal consistency, construct 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 – Positive Scale (Watson & Clark, 1994).The Positive and Negative Affect Schedule (PANAS) is a well-validated measure of positive and negative emotions. 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 have ranged from .83 to .90 (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 Items (Lovibond & Lovibond, 1995). The Depression, Anxiety and Stress Scale-21 (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*). High internal consistency is generally reported for each subscale of the DASS (depression: .72; anxiety: subscale .77; stress: .70; Tran, Tran, & Fisher, 2013). 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 (Rizvi et al., 2015). The DARS measures anhedonia across the following 4 areas: Hobbies/pastimes, food/drink, social activities, and sensory experience. 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, milk-shakes, 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 (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-D-II (Rohrbacher & Reinecke, 2014).The Ambiguous Scenarios Test (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). 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[1]. 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 (Wenzlaff, 1993). The Scrambled Sentences Task (SST) was used as an indirect measure of depressive interpretation bias (Phillips, Hine, & Thorsteinsson, 2010). Participants were asked to unscrambled 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

[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.

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. We calculated internal consistency as split-half reliability (odd vs. even-numbered items) using the Spearman-Brown formula. 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 (Stöber, 2000) The Prospective Imagery Test (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. 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 (Blackwell et al., 2013; Ji, Holmes, & Blackwell, 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 (Devilly & 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 to 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 (SEB), 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, ≥ 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 4 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 pre-training measures in this order: QIDS, PANAS-P, DARS, DASS, BADS-SF. 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 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, i.e. 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, DARS, DASS, BADS-SF) 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, Davies, & Kraemer, 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) were calculated for each group. Descriptive statistics and reliabilities were calculated using SPSS (Statistical Package for Social Sciences) software version 23.

For the outcome measures, outcomes at each timepoint were calculated both as raw means and as standardized effect sizes, both with 95% confidence intervals. These were calculated both for an 'intention to treat' sample (i.e. all participants randomized, regardless of session or outcome measure completion) and a 'complete case' sample (i.e. all participants providing outcome data for the specific timepoint).

To derive means, effect sizes, and 95% CIs for the ITT sample, a mixed-model repeated measures ANOVA was fitted over the three assessment timepoints, and estimates derived from this model. This allowed an ITT approach without having to e.g. 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 (complete case) mean change by the pooled standard deviation of the baseline scores, to provide an effect size estimate reflecting a standardized mean difference less affected by potentially biasing factors such as differential drop-out 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) in order to provide intervals consistent with the unstandardized mean differences. For both ITT and complete case 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), 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 shows the demographic characteristics and scores on additional baseline variables for participants. At baseline, 29 participants were allocated to imagery CBM and 26 to PVT. Five participants in the imagery CBM condition and 9 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 drop-out (13 participants in imagery CBM condition and 9 participants in the PVT condition) occurring towards 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) was not collected. Although the post-training assessment was planned to take place one 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%

Confidence Intervals [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 (two 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 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 towards 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 timepoints of the study for both intention to treat and complete case populations. These include mean scores at each time point, and standardized effect sizes for within-group change from baseline and between-group differences in change scores, with associated 95% confidence intervals. Although the estimates generally favour 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 drop-out (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 shown in 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, amongst a specific sample in Pakistan, university students. The intervention, imagery cognitive bias modification (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.

Following previous research, the current study aimed to investigate a schedule comprising one session of imagery CBM daily for one week, with follow-up after two weeks post-training. In practice, there were some difficulties with implementing this schedule, with the post-training schedule often occurring later than one 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. This suggests that future research in this population may benefit from a more flexible schedule of training, and also collect outcome data remotely, such as online or via telephone. The current study also included only minimal incentives for study completion (i.e. a prize draw only), and in a research context larger participant incentives may be needed for a greater rate of data completeness at follow-up.

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, and this does not appear to be an effect of differential drop-out: including only those participants who completed post-training measures does not result in much better internal consistencies for the baseline measurements. The improvement in internal consistencies could be a practice or familiarity effect. 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 et al., in press). One difference in the current sample may be the relatively elevated symptoms of depression. However, in future it would be recommendable to use Urdu measures with demonstrated internal consistency. 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 Peripheral Vison Task was chosen as a comparison task in order to provide something approximating a 'placebo' control, that is, controlling for non-specific 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. Post-training ratings are potentially difficult to interpret, as they are conflated with perceived benefit; a task providing obvious benefits will of course be rated more positively. Further, this may not reflect a problem with the PVT as 'placebo' as such, but rather good credibility of the imagery CBM as an intervention: More broadly, it is not unusual for cognitive training interventions such as CBM-based interventions (e.g. attention bias modification) to be judged as having low face validity and for many participants receiving an active CBM training to think themselves in a 'sham' condition (Carlbring et al., 2012; de Voogd et al., 2017), but it may be that this is less of a problem for the imagery CBM paradigms. Further, as limited introductions to the tasks were provided with no rationale as to why they might be beneficial, they would be judged on their perceived benefits and face validity. However, 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.

Notwithstanding the caveats that have to be applied to interpreting the outcome data in the current study, both due to the problems outlined above, and also given the limitations of

the sort time-scale of follow-up and lack of blinding of researchers, the study indicates that it would be useful to continue to investigate the potential for imagery CBM as a low-intensity intervention amongst 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 may be particularly 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. The use of a one-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). 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. 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 pre-

registration and in hindsight it would have been preferable to pre-register this pilot study. 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.

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 to other developing countries (Asad et al., 2010; Husain et al., 2011; Luni, Ansari, Jawad, Dawson, & Baig, 2009). 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 towards patients with psychological disorders (Ahmad, 2007). In such a situation, where social stigma, myths, and hostile belief structures around 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.

In conclusion, the current pilot study suggests that imagery CBM warrants further investigation as an easily accessible computerized intervention for symptoms of depression amongst 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 in order to detect potential problems that may otherwise compromise subsequent hypothesis-testing research.

Availability of data and material

The datasets generated and analysed 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.

Declaration of conflicting interests

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

Funding

AB was supported by HEC/DAAD. SEB and JM were employed by 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.

Acknowledgements

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

References

Ahmad, K. (2007). Shortage of psychiatrists a problem in Pakistan. *Lancet, 370*(9590), 817–818. https://doi.org/10.1016/S0140-6736(07)61397-X

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington D.C.: American Psychiatric Association.

Asad, N., Karmaliani, R., Sullaiman, N., Bann, C. M., McClure, E. M., Pasha, O., … Goldenberg, R. L. (2010). Prevalence of suicidal thoughts and attempts among pregnant Pakistani women. *Acta Obstetricia et Gynecologica Scandinavica, 89*(12), 1545–1551. https://doi.org/10.3109/00016349.2010.526185

Beck, A. T. (1976). *Cognitive Therapy and the Emotional Disorders.* New York: 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, Lin & Margraf, (in press). Salutogenic Constructs across Pakistan and Germany/ Construcciones salutogénicas en Pakistán y Alemania. *International Journal of Clinical and Health Psychology*.

Blackwell, S. E., Browning, M., Mathews, A., Pictet, A., Welch, J., Davies, J., … 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), 338–50. 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*, 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., … Giltay, E. J. (2013). Optimism and mental imagery: A possible cognitive marker to promote wellbeing? *Psychiatry Research*, *206*(1), 56–61. https://doi.org/10.1016/j.psychres.2012.09.047

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

Bukhari, S. R., & Khanam, S. J. (2015). Prevalence of depression in university students belonging to different socioeconomic status. *Journal of Postgraduate Medical Institute*, *29*(3), 156-159.

Butler, G., & Mathews, A. (1983). Cognitive processes in anxiety. *Advances in Behaviour Research and Therapy*, *5*, 51–62.

Calkins, A. W., McMorran, 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*, 578–589. https://doi.org/10.1017/S1352465814000046

Carlbring, P., Apelstrand, M., Sehlin, H., Amir, N., Rousseau, A., Hofmann, S. G., & Andersson, G. (2012). Internet-delivered attention bias modification training in individuals with social anxiety disorder - a double blind randomized controlled trial. *BMC Psychiatry*, *12*, 66. https://doi.org/10.1186/1471-244X-12-66

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

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*. New York: Routledge.

de Voogd, E. L., de Hullu, E., Burnett Heyes, S., Blackwell, S. E., Wiers, R. W., & Salemink, E. (2017). Imagine the bright side of life: A randomized controlled trial of two types of interpretation bias modification procedure targeting adolescent anxiety and depression. *PLOS ONE*, *12*(7), e0181147. https://doi.org/10.1371/journal.pone.0181147

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, *31*, 73–86. https://doi.org/10.1016/S0005-7916(00)00012-4

Dunn, B. D., German, R. E., Khazanov, G., Xu, C., Hollon, S. D., & DeRubeis, R. J. (in press). Changes in positive and negative affect during pharmacological treatment and cognitive therapy for major depressive disorder: a secondary analysis of two randomized controlled trials. *Clinical Psychological Science.* https://doi.org/10.1177/2167702619863427

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

Gueorguieva, 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*, 310–317.

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., 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

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

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 Menal Health Institutes in Pakistan*. Islamabad, Pakistan: World Health Organization.

Kakuma, R., Minas, H., Van Ginneken, N., Dal Poz, M. R., Desiraju, K., Morris, J. E., … 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

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,* 626-629. https://doi.org/10.1016/j.jpsychires.2010.10.008

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sydney: Psychology Foundation.

Luni, F. K., Ansari, B., Jawad, A., Dawson, A., & Baig, S. M. (2009). Prevalence of

depression and anxiety in a village in Sindh. *Journal of Ayub Medical College Abbottabad*, *21*(2), 68-72.

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.

Phillips, W. J., Hine, D. W., & Thorsteinsson, E. B. (2010). Implicit cognition and depression: a meta-analysis. *Clinical Psychology Review*, *30*, 691–709. https://doi.org/10.1016/j.cpr.2010.05.002

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

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2019). _nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-140, https://CRAN.R-project.org/package=nlme

R Core Team (2019). (2019). R: A language and environment for statistical computing. *Accessed 1st April 2019*.

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,* 369–80. 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,* 109–119. https://doi.org/10.1016/j.psychres.2015.07.062

RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc.,Boston, MA. http://www.rstudio.com

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

Rush, J. A., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., … 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

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

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). Patttern of somatic symptoms in anxiety and depression. *Pakistan Armed Forces Medical Journal*, *61*, 447–452.

Stöber, J. (2000). Prospective cognitions in anxiety and depression: Replication and methodological extension. *Cognition & Emotion*, *14*(5), 725–729. https://doi.org/10.1080/026999300501176 93

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

Tran, T. D., Tran, T., & Fisher, J. (2013). Validation of the depression anxiety stress scales (DASS) 21 as a screening instrument for depression and anxiety in a rural community-based cohort of northern Vietnamese women. *BMC Psychiatry*, *13*(1), 24. https://doi.org/10.1186/1471-244X-13-24

Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., … 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*(01), 73–82. https://doi.org/doi:10.1017/S0033291703001107

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. University of Iowa, Iowa City.

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). Englewood Cliffs, NJ: Prentice Hall.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A., Ferrari, A. J., Erskine, H. E., … 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

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

Figure 1. Flow of participants through the study.

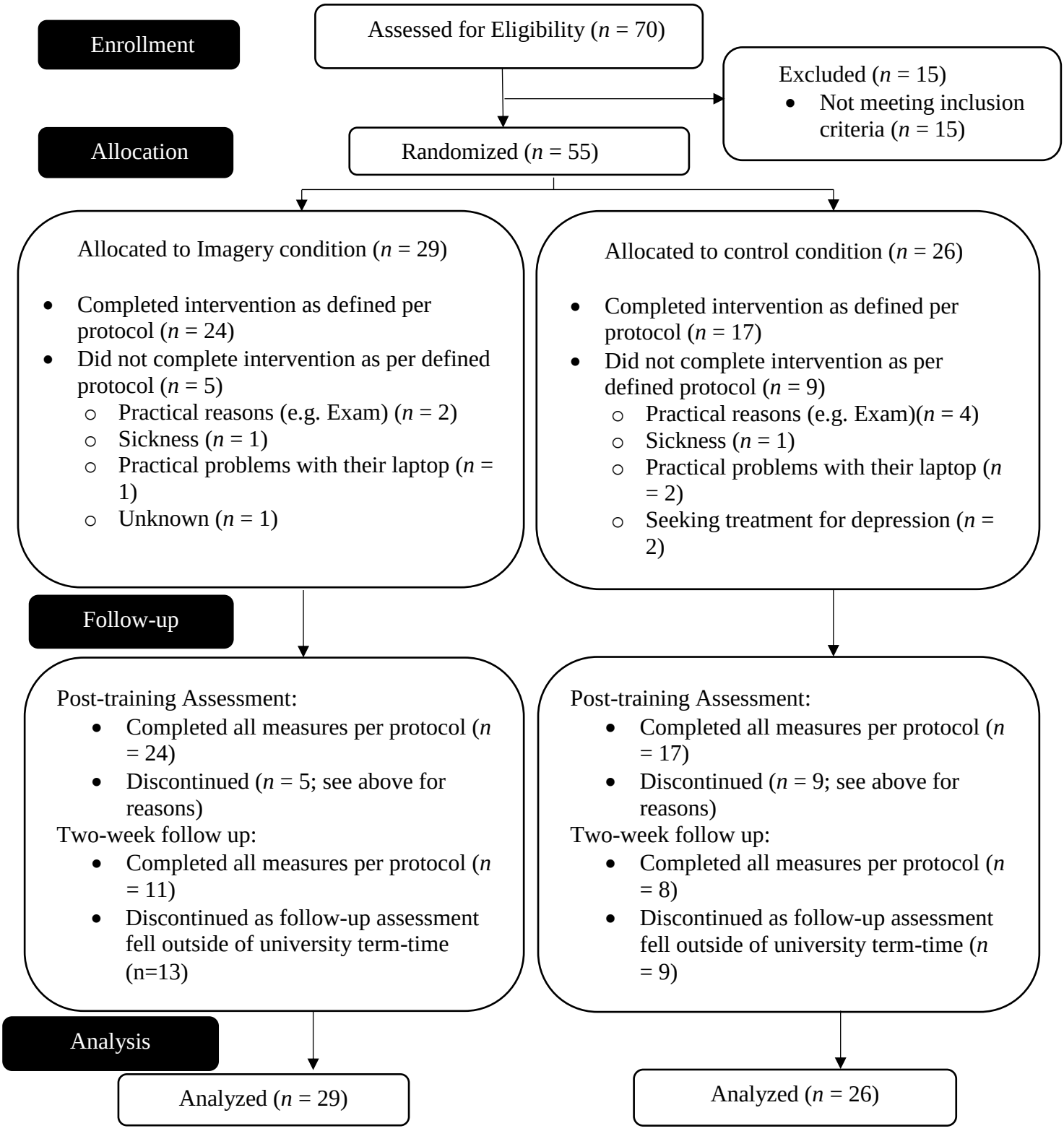

Table 1. Sample characteristics across conditions

	Imagery CBM (n = 29) *M* [95% CIs]	PVT (n = 26) *M* [95% CIs]
Age	22.34 [21.79, 22.90]	22.58 [21.84, 23.31]
Gender, *N* (%)		
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. QIDS = Quick Inventory of Depressive Symptomatology; DASS = Depression Anxiety Stress Scale (21-item version). Imagery CBM = Imagery Cognitive Bias Modification; PVT = Peripheral Vision Task.

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

	Imagery CBM *M* [95% CIs]	PVT *M* [95% CIs]
Expectancy	*n* = 29	*n* = 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)	*n* = 24	*n* = 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. EQ Qu1 = "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 Qu3 = "By end of the study period, how much improvement in your low mood/ depression symptoms do you really *feel* will occur?"; 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*).

Table 3. Means and effect sizes with 95% Confidence Intervals for outcome measures over the course of the study for both intention to treat (ITT) and complete case (CC) samples.

	Baseline (T1)	Post-training (T2)	Follow-up (T3)	Within-group d [95% CIs]		Between-group d [95% CIs]	
	M [95% CIs]	M [95% CIs]	M [95% CIs]	T1 to T2	T1 to T3	T1 to T2	T1 to T3
QIDS							
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]		
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]	1.16 [0.82, 1.98]	0.28 [-0.34, 0.69]
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]		
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]	1.09 [0.80, 2.21]	0.57 [-0.43, 1.42]
PANAS							
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]		
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]	1.93 [1.24, 2.48]	1.22 [0.31, 1.39]
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]		
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]	2.03 [1.49, 3.09]	1.47 [0.36, 2.41]
DARS							
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]			
BADS							
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]		
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]	1.56 [0.80, 1.94]	1.53 [0.37, 1.46]
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]		
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]	1.29 [0.68, 2.07]	1.22 [-0.07, 1.85]
DASS-D							
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]		
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]	2.22 [1.49, 2.79]	0.72 [-0.14, 0.90]
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]		
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]	2.19 [1.65, 3.31]	0.49 [-0.57, 1.27]
DASS-A							
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]		
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]	2.05 [1.51, 3.13]	0.82 [-0.32, 1.55]
DASS-S							
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]		
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]	2.48 [1.81, 3.20]	1.62 [0.43, 1.52]
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]		
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]	2.52 [2.01, 3.80]	1.53 [0.20, 2.19]
AST							
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]			
SST							
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]	
CC				
CBM	0.64 [0.58,0.69]	0.32 [0.29,0.35]	2.21 [1.95,3.77]	2.00 [1.65,3.31]
Control	0.58 [0.53,0.64]	0.54 [0.48,0.61]	0.24 [-0.19,0.70]	
PIT positive				
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]	
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. ITT = Intention to Treat, CC = Complete Case (using all available data, i.e. the mean presented at pre-training 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 Symptoms, PANAS = Positive and Negative Affect Schedule, DARS = Dimensional Anhedonia Rating 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 compared to 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.

Supplementary Table 1

Means and effect sizes with 95% Confidence Intervals for subscales of the Dimensional Anhedonia Rating Scale and Positive and Negative Affect Schedule over the course of the study.

	Baseline (T1)	Post-training (T2)	Follow-up (T3)	Within-group d [95% CIs]		Between-group d [95% CIs]	
	M [95% CIs]	M [95% CIs]	M [95% CIs]	T1 to T2	T1 to T3	T1 to T2	T1 to T3
DARS Hobbies							
ITT							
CBM	21.17 [19.86, 22.49]	25.99 [24.67, 27.31]		1.32 [1.17, 2.16]		1.25 [1.11, 2.58]	
Control	19.92 [18.53, 21.32]	20.23 [18.75, 21.72]		0.08 [-0.22, 0.38]			
CC							
CBM	21.17 [19.96, 22.39]	25.75 [24.68, 26.82]		1.35 [1.15, 2.25]		1.27 [1.15, 2.65]	
Control	19.92 [18.36, 21.49]	20.12 [18.24, 22.00]		0.10 [-0.29, 0.48]			
DARS Food							
ITT							
CBM	13.24 [12.26, 14.22]	16.96 [15.92, 18.01]		1.37 [1.09, 2.06]		0.90 [0.55, 1.89]	
Control	12.46 [11.42, 13.50]	13.76 [12.58, 14.93]		0.47 [0.11, 0.76]			
CC							
CBM	13.24 [12.31, 14.17]	17.00 [16.04, 17.96]		1.36 [1.00, 2.13]		0.99 [0.63, 2.01]	
Control	12.46 [11.32, 13.61]	14.35 [12.75, 15.96]		0.38 [0.01, 0.69]			

DARS Social							
ITT							
CBM	12.86 [11.95, 13.78]	17.32 [16.21, 18.43]		1.76 [1.29, 2.52]		1.26 [0.53, 1.87]	
Control	12.81 [11.84, 13.77]	14.12 [12.83, 15.41]		0.51 [0.02, 0.86]			
CC							
CBM	12.86 [11.95, 13.77]	17.17 [16.20, 18.13]		1.81 [1.27, 2.64]		1.39 [0.62, 1.99]	
Control	12.81 [11.79, 13.82]	14.35 [12.63, 16.08]		0.43 [-0.14, 0.89]			
DARS Sensory							
ITT							
CBM	10.72 [9.95, 11.50]	14.67 [13.59, 15.75]		1.84 [1.20, 2.50]		1.72 [0.66, 2.02]	
Control	11.00 [10.18, 11.82]	11.30 [10.03, 12.57]		0.14 [-0.36, 0.58]			
CC							
CBM	10.72 [9.98, 11.47]	14.62 [13.66, 15.59]		1.87 [1.19, 2.56]		1.78 [0.68, 2.07]	
Control	11.00 [10.12, 11.88]	11.35 [9.76, 12.95]		0.11 [-0.50, 0.67]			
PANAS- J							
ITT							
CBM	17.79 [16.56, 19.03]	23.28 [22.18, 24.38]	20.66 [19.30, 22.02]	1.61 [1.26, 2.29]	0.81 [0.38, 1.51]		
Control	17.38 [16.08, 18.69]	17.18 [15.90, 18.47]	15.99 [14.39, 17.58]	-0.06 [-0.54, 0.41]	-0.38 [-1.18, 0.17]	1.70 [0.99, 2.17]	1.24 [0.35, 1.43]
CC							
CBM	17.79 [16.55, 19.04]	23.25 [22.03, 24.47]	20.27 [18.42, 22.13]	1.63 [1.14, 2.43]	0.94 [0.15, 2.03]	1.79 [1.15, 2.65]	1.47 [0.29, 2.31]
Control	17.38	17.35	16.00	-0.14	-0.44		

	[16.05, 18.72]	[16.06, 18.65]	[14.27, 17.73]	[-0.71, 0.39]	[-1.44, 0.30]		
PANAS-SA							
ITT							
CBM	13.41	16.90	14.56	1.45	0.46		
	[12.54, 14.29]	[16.09, 17.71]	[13.22, 15.91]	[1.24, 2.27]	[-0.10, 1.06]		
Control	12.73	12.96	12.84	0.09	0.04	1.37	0.43
	[11.81, 13.66]	[12.02, 13.89]	[11.27, 14.40]	[-0.30, 0.49]	[-0.57, 0.66]	[0.82, 1.97]	[-0.26, 0.77]
CC							
CBM	13.41	16.58	14.09	1.57	0.43		
	[12.58, 14.25]	[15.85, 17.32]	[12.38, 15.80]	[1.28, 2.54]	[-0.29, 1.18]		
Control	12.73	12.82	12.62	0.14	-0.29	1.45	0.77
	[11.73, 13.73]	[11.74, 13.90]	[10.39, 14.86]	[-0.33, 0.63]	[-0.85, 0.29]	[1.08, 2.56]	[-0.23, 1.65]
PANAS-S							
ITT							
CBM	6.31	8.49	7.71	1.44	0.88		
	[5.76, 6.86]	[7.97, 9.02]	[6.99, 8.42]	[1.17, 2.42]	[0.37, 1.70]		
Control	6.73	6.76	6.64	0.02	-0.05	1.45	0.97
	[6.15, 7.31]	[6.13, 7.38]	[5.80, 7.48]	[-0.45, 0.48]	[-0.70, 0.58]	[0.59, 1.70]	[0.06, 1.11]
CC							
CBM	6.31	8.50	7.55	1.43	0.92		
	[5.81, 6.81]	[8.03, 8.97]	[6.63, 8.46]	[1.09, 2.46]	[0.10, 2.03]		
Control	6.73	6.76	6.50	0.00	-0.15	1.45	1.13
	[6.08,7.38]	[5.98,7.55]	[5.32,7.68]	[-0.59,0.59]	[-0.93,0.61]	[0.64, 2.01]	[-0.04, 1.89]

Note. ITT = Intention to Treat, CC = Complete Case (using all available data, i.e. the mean presented at pre-training 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), PANAS-J/SA/S = Joviality/Self-Assurance/Serenity subscale of the Positive and Negative Affect Schedule, DARS Hobbies/ Food/ Social/ Sensory= Dimensional Anhedonia Rating Scale Hobbies/ Food/ Social/ Sensory subscale.