Behavioural and Cognitive Psychotherapy, 2020, page 1 of 13
doi:10.1017/S1352465820000922

BABCP
www.babcp.com

MAIN

Engagement with smartphone-delivered behavioural activation interventions: a study of the MoodMission smartphone application

Abby Aizenstros[1], David Bakker[1,2,*], Stefan G. Hofmann[3], Joshua Curtiss[3] and Nikolaos Kazantzis[1,2]

[1]Cognitive Behaviour Therapy Research Unit, Monash University, Australia, [2]Cognitive Behaviour Therapy Research Unit, Institute for Social Neuroscience, Australia and [3]Department of Psychological and Brain Sciences, Boston University, USA
*Corresponding author. Email: dbakker@isn.edu.au

(Received 18 May 2020; revised 15 October 2020; accepted 2 November 2020)

Abstract

Background: Despite increased research interest in smartphone mental health applications (MHapps), few studies have examined user engagement and its determinants. MoodMission is a MHapp that targets low mood and anxiety via evidence-based techniques including behavioural activation (BA).

Aims: The present study aimed to investigate (i) whether BA interventions delivered with visual psychoeducation had greater engagement than BA interventions delivered with solely written psychoeducation, (ii) whether BA interventions targeting mastery would have greater engagement than those targeting pleasure, and (iii) the relationship between level of engagement and MHapp benefit.

Method: Participants downloaded MoodMission and completed activities and within-app evaluations over a 30-day period. Data from 238 MoodMission users were analysed via multi-level modelling and linear regression.

Results: The average number of app-based activities completed was 5.46 and the average self-reported engagement level was in the low to moderate range. As hypothesized, higher levels of engagement significantly predicted more positive activity appraisal.

Conclusions: The results suggest that BA technique beliefs are involved in MHapp engagement and future research examining user appraisals of techniques is warranted.

Keywords: app; behavioural activation; engagement; homework

Introduction

Smartphone mental health applications (MHapps) have emerged as a vehicle for the delivery of empirically supported technology-based intervention (Economides *et al.*, 2018; Firth *et al.*, 2017). 'MoodMission' (MoodMission Pty Ltd, 2019) is an MHapp directed at anxiety and depression symptoms and wellbeing (Bakker *et al.*, 2018a). Health-promoting activities called 'Missions' are provided, using techniques including mindfulness, cognitive restructuring and behavioral activation (BA) (Bakker *et al.*, 2018b).

Engagement with an MHapp is required for outcomes to be achieved (Torous *et al.*, 2018). Engagement is a broader concept than compliance or adherence and takes into consideration the subjective appraisal of task difficulty and obstacles to task completion (Holdsworth *et al.*, 2014). MHapp engagement can be influenced by privacy concerns (Sagar and Pattanayak, 2015) and whether users perceive the app as straightforward and useful (Torous *et al.*, 2018). Previously, engagement with MoodMission was found to predict greater wellbeing (Bakker and Rickard, 2019).

Social cognition theories explain engagement with mental health tasks (Kazantzis *et al.*, 2005), including Bandura's (1986) social cognitive theory, where individuals who believe that they are capable of a task are more likely to persevere with the task. Operant conditioning also helps us to consider useful triggers to engagement, where a previously reinforcing experience supports generalization (Skinner, 1938). MHapps that do not offer therapist or peer support can provide informational support to enhance self-efficacy and engagement. Psychoeducation can be provided to promote continued use, intervention satisfaction and perceived helpfulness (Hidalgo-Mazzei *et al.*, 2018). However, the type and extent of psychoeducation delivered have not been extensively studied for their impact on engagement in the MHapp context.

MoodMission incorporates BA strategies, promoting participation in positively reinforcing tasks to evoke a sense of mastery and/or pleasure (Martell *et al.*, 2010). Mastery-enhancing Missions include making a scrapbook and pleasure-enhancing Missions consist of watching films. Mastery and pleasure have long been discussed in the literature as powerful reinforcers for behaviour change. However, the questions of whether BA should focus on a specific activity type, and whether all activities yield comparable levels of engagement in BA remain unanswered (Dimidjian *et al.*, 2011).

A systematic review concluded that the assessment of the broader construct of engagement is infrequent among published studies (Kazantzis *et al.*, 2017). Additionally, inconsistent assessment methods render comparisons between studies difficult (Ng *et al.*, 2019). The Homework Rating Scale – Revised (HRS-II; Kazantzis *et al.*, 2005) is a measure of engagement and beliefs about the between-session tasks of therapy (i.e. 'homework'). The HRS-II contains three subscales: engagement, beliefs and consequences (McDonald and Morgan, 2013). The scale's internal consistency is supported by Cronbach's alphas between .71 and .91, and preliminary analysis (e.g. Sachsenweger *et al.*, 2015) has supported its theoretically derived 3-factor structure (Kazantzis *et al.*, 2005). The HRS-II has been adapted for use in MHapps (i.e. HRS-MA; Bakker and Kazantzis, 2020) and was incorporated into the present study following a test of its psychometric properties. Compared with other interventions, MHapps tend to have low retention rates (Bauer *et al.*, 2020), but as retention is only one small part of engagement, more comprehensive assessment that considers behaviour and cognitive theory determinants of engagement is needed.

Our primary aim was to examine if mode of psychoeducation delivery resulted in different levels of engagement with BA delivered in an MHapp context. It was hypothesized that the extent of engagement, measured by the number of Missions completed and the HRS-MA engagement subscale (HRS-MA-engagement), would be higher for Missions that provided greater support, classified as Missions delivering both written and diagrammatic psychoeducation as opposed to solely written (Hypothesis 1). Second, we sought to compare engagement in mastery- *versus* pleasure-enhancing BA activities. Past research has not conducted this comparison, although in accordance with self-efficacy theory (Bandura, 1986; and see review by Kazantzis and L'Abate, 2005), we hypothesized higher engagement for Missions utilizing mastery- than pleasure-enhancing BA (Hypothesis 2). Our final aim was to investigate the impact of engagement on the user-reported benefit of the app. App appraisal was collected as a post-intervention assessment and therefore was considered as an outcome of engagement. It was hypothesized that engaged participants would report more positive Mission appraisals, measured by the Mission Helpfulness Rating scale (Hypothesis 3a) and the beliefs and consequences subscales of the HRS-MA (HRS-MA-beliefs and HRS-MA-consequences, respectively; Hypothesis 3b). Additionally, participants reporting higher engagement were expected to experience larger mental health gains, assessed by participant distress ratings (Hypothesis 4a) and measures of depression, anxiety and wellbeing (Hypothesis 4b).

Table 1. Frequencies of participant demographic variables

Demographic variable	*n*	Percentage
Age		
<20 years	61	25.63
20–30 years	104	43.70
31–40 years	40	16.81
>40 years	32	13.45
Gender		
Male	47	19.75
Female	184	77.31
Other	7	2.94
Highest level of education		
Primary	10	4.20
Secondary	75	31.51
Tertiary	85	35.71
Postgraduate	54	22.69
Education status		
Full-time	97	40.76
Part-time	26	10.92
Not currently studying	115	48.32
Employment status		
Full-time	87	36.55
Part-time	73	30.67
Not working, but actively seeking work	26	10.92
Not working, and not seeking work	43	18.07
Not working due to sick leave	9	3.78

$n = 238$. Demographic percentages do not total 100% for age and highest level of education due to non-response.

Method

Participants

The sample consisted of 238 participants, aged 12 to 63 years (mean = 27.13, $SD = 10.84$). The demographic frequencies of the sample are presented in Table 1. The majority of participants were within the 20- to 30-year-old age range and female. Most participants had received tertiary education, were not currently studying, and were employed full-time. Participants were members of the general population and were not specifically a clinical sample.

Participants were voluntarily recruited upon downloading MoodMission to their personal smartphone. App download was free from the iOS and Android App Stores. MoodMission was promoted as a mental health tool on social media sites, with no monetary incentive being provided for app download. Data were collected from app users who used MoodMission since the HRS-MA's introduction to the app on 29 March 2019. Additionally, participation required app users to provide data until the 30-day evaluation time point. Of the 4260 individuals who used MoodMission within this project's data collection period, 238 app users satisfied this criterion and were included in the sample. Analyses were performed to compare the demographics of the sample against the demographics of the 4022 MoodMission users who concluded their app use prior to the 30-day time point and are reported as preliminary. There were no other inclusion or exclusion criteria for research participation, apart from requiring access to a smartphone.

Measures

Homework Rating Scale – Mobile Application Version

The Homework Rating Scale – Mobile Application Version (HRS-MA; Bakker and Kazantzis, 2020) assessed engagement with and appraisal of MoodMission. Based on an integration of cognitive and behavioural theories (Kazantzis and L'Abate, 2005), the measure consists of 12 self-report items

rated on a 5-point scale, ranging from 0 (*not at all*) to 4 (*completely/extremely/extensive*). Items 1 to 4 belong to the HRS-MA-engagement, items 5 to 9 belong to the HRS-MA-beliefs, and items 10 to 12 belong to the HRS-MA-consequences. The HRS-MA items can be seen in the Supplementary material. The psychometric properties of the HRS-MA had not been investigated prior to the present study. Therefore, a series of analyses on the HRS-MA are reported as preliminary. However, the HRS-II has been evaluated previous to its smartphone adaptation and was found to have strong psychometric properties. The scale has shown high internal consistency, with an alpha of .92 (Hara *et al.*, 2017), and a 3-factor structure (Richardson *et al.*, 2020).

Mission Helpfulness Rating Scale

MoodMission's Mission Helpfulness Rating scale (MHR) measured how helpful participants found each Mission after completion. The single question 'how helpful was that?' is answered on a continuum ranging from 0 (*not helpful at all*) to 10 (*extremely*). Individuals slide a pointer along the continuum to identify the point on the scale that quantified their perceived helpfulness. A continuous measure is consistent with prior research to assess perceived helpfulness (Duncan *et al.*, 2010).

Subjective Units of Distress Scale

The Subjective Units of Distress Scale (SUDS; Wolpe, 1969) was used to assess distress, providing a rating between 0 and 10 (Parrish *et al.*, 2016). In MoodMission's version, a rating of 0 represents 'not at all' distressed and 10 represents 'extremely' distressed. Similar to the MHR, the rating is given upon sliding a pointer along a continuum, ranging from 0 to 10. The concurrent validity of the scale has been empirically supported (Tanner, 2012).

Patient Health Questionnaire

The 9-item Patient Health Questionnaire (PHQ-9; Kroenke *et al.*, 2001) was used to measure depression. Based on the last 2 weeks, items are rated on a 4-point scale ranging from 0 (*not at all*) to 3 (*nearly every day*). One item includes 'little interest or pleasure in doing things'. A total score between 0 and 4 indicates 'minimal' depression, 5 to 9 indicates 'mild', 10 to 14 indicates 'moderate', 15 to 19 indicates 'moderately severe' and 20 to 27 indicates 'severe'. Past evaluation of MoodMission has applied the PHQ-9 to a smartphone context and reported an alpha of .90 (Bakker *et al.*, 2018a). The present study supported the high reliability of the PHQ-9 in a smartphone context, finding an alpha of .88 and McDonald's omega total of .91.

Generalized Anxiety Disorder Scale

The Generalized Anxiety Disorder scale (GAD-7; Spitzer *et al.*, 2006) is a 7-item anxiety measure. Individuals reflect on the last 2 weeks and rate the frequency of items, such as 'trouble relaxing', between 0 (*not at all*) and 3 (*nearly every day*). Mild, moderate and severe anxiety levels are inferred from the cut-off points of 5, 10 and 15, respectively. The GAD-7 previously achieved an alpha of .92 (Bakker *et al.*, 2018a; Spitzer *et al.*, 2006). The alpha and omega values calculated in the present study were .89 and .92, respectively.

Warwick-Edinburgh Mental Wellbeing Scale

Wellbeing was measured by the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS; Tennant *et al.*, 2007). Fourteen statements, including 'I've been feeling useful', are rated from 1 (*none of the time*) to 5 (*all of the time*) based on the last 2 weeks. Total scores between 0 and 32 are classified as 'very low', 33 to 40 as 'below average', 41 to 59 as 'average' and 60 to 70 as 'above average'.

The WEMWBS has high reliability, with an alpha of .92 in a smartphone context (Bakker *et al.*, 2018a). In the present study, the scale achieved an alpha of .88 and omega of .92.

Procedure

A longitudinal design was used to assess participant engagement, appraisal and symptom levels. After downloading MoodMission, participants completed the PHQ-9, GAD-7 and WEMWBS, and provided demographic information. Over a 30-day period, participants decided their frequency of app use. Each app use involved reporting whether they were feeling anxious or depressed, selecting the statement that best matched their mental state, and providing a pre-Mission SUDS rating. Based on the symptom profile reported, the app generated five Mission alternatives from which the participants would self-select one. These Mission alternatives came from a pool of 220 Missions. Participants completed the Mission selected and provided a post-Mission MHR and SUDS rating. After 30 days had elapsed, participants once again completed the PHQ-9, GAD-7 and WEMWBS, in addition to the HRS-MA. All questionnaires were built into the app interface. Data collection from the app was ceased upon completion of the 30-day questionnaires.

Data analysis

Sample size

The target sample size was 266, as estimated for an MHapp intervention group by a multi-level modelling (MLM) study anticipating moderate effects, where power was set at 80% and alpha was set at .05 (Deady *et al.*, 2018). The present study also anticipated moderate effects, as this effect size has been found for the efficacy of MHapps (Firth *et al.*, 2017) and when homework is added to treatments (Kazantzis *et al.*, 2010). Missions can be considered as similar to homework because they both allow individuals to apply cognitive behavioural therapy (CBT) skills to their everyday environment. A target of 266 participants typically exceeds the sample sizes reported in the current MHapp literature (Bakker and Rickard, 2018; Giosan *et al.*, 2017).

Preliminary analyses

As there was a large disparity between the number of MoodMission users recruited ($N = 4260$) and the number included in the sample ($n = 238$), comparisons were performed between the demographics of the sample and the demographics of the group who stopped using MoodMission prior to the 30-day time point. The Mann–Whitney *U*-test was conducted to determine whether there was an age difference between the two groups. This test was chosen instead of the independent samples *t*-test due to normality assumption violation. The ages of the sample (mean rank = 1694.45, $n = 237$) were significantly lower than of those who stopped using MoodMission (mean rank = 2155.66, $n = 4022$), $U = 373382.50$, $z = -5.61$ (corrected for ties), $p < .001$, two-tailed. However, this effect was small ($r = .10$). The Mann–Whitney *U*-test was also performed to compare the groups for their highest education level. The assumption that group distributions have the same shape was met upon histogram inspection. Highest education level was not found to significantly differ between the sample (mean rank = 1559.69, $n = 224$) and those who stopped use (mean rank = 1618.05, $n = 3003$), $U = 324171$, $z = -.95$ (corrected for ties), $p = .34$, two-tailed. Pearson's chi-square test of independence with an alpha of .05 was conducted to assess whether gender was related to prematurely ending app use. The chi-square test was statistically significant, $\chi^2(2, N = 4260) = 43.51$, $p < .001$, although the effect size was small, with a Cohen's w of 0.10. A chi-square test investigating whether education status was related to stopping use was statistically significant, $\chi^2(3, N = 4260) = 74.51$, $p < .001$, and the association was small, $w = 0.13$. Finally, a chi-square test evaluating the independence of

employment status and stopping use found a significant relationship between the two variables, $\chi^2(5,N = 4260) = 54.47$, $p < .001$, although this association was small, $w = 0.11$. All expected frequencies were larger than 5, as assumed for the chi-square test. Therefore, prematurely stopping MoodMission use was related to age, gender, education status and employment status, but not highest education level.

Preliminary analyses were further conducted to test the psychometric properties of the HRS-MA. An alpha of .88 and omega of .93 supported internal consistency. HRS-MA validity was assessed by exploratory factor analysis with varimax rotation. The factor analysis evaluated the number of factors to which the HRS-MA items could be reduced, with comparisons being made to the 3-factor structure previously found for the HRS-II (Richardson *et al.*, 2020; Sachsenweger *et al.*, 2015). These three factors refer to the three subscales: engagement, beliefs and consequences. Inter-item correlations were examined to assess the factorability of the data and were used to conduct a heat map analysis. Items 3 and 4 were excluded due to their low correlations with the other items, all under the recommended minimum threshold of .30 (Jacobs *et al.*, 2017). Upon inspection of eigenvalues and a scree plot, two factors were retained in the factor analysis. Factor analysis revealed that most of the retained items loaded onto both factors. As items 3 and 4 are conceptually important to the construct of engagement, HRS-MA-engagement might not have been represented in the factor analysis. Due to the theoretical support for these engagement items (Kazantzis *et al.*, 2005), all 12 HRS-MA items were retained in a subsequent factor analysis and the main statistical analyses. When all 12 items were included, the eigenvalues and scree plot supported the retainment of three factors in the factor analysis, which aligns with there being three subscales (see also Supplementary material). Analyses were performed using IBM SPSS Statistics (version 25) and the R Project for Statistical Computing (version 3.5.3).

Data cleaning

Missing values were addressed using the maximum likelihood estimation procedure for the variables assessed via MLM (Dempster *et al.*, 1977). As linear regression is less robust to missing data, a complete sample was used for the variables assessed via linear regression models, with list-wise deletion being performed on cases with missing values (Pepinsky, 2018). Screening for outliers was based on inspection of scaled residuals and distance measures, and any outliers were removed (Bakker and Wicherts, 2014).

Assumption checking

Independence and normality assumptions of MLM were assessed (Snijders and Berkhof, 2008). Although MLM relaxes the assumption of the independence of observations, as each participant may provide multiple scores, the model's errors at the highest level of units is assumed to be independent. The normality assumption required the residuals to follow a normal distribution, which was assessed upon Q-Q plot inspection. MLM is perceived as robust to normality violation (Fung and Xu, 2010); however, in the case of significant deviations from normality for data that exhibited behaviour consistent with count data, Poisson MLM was used (Ozonur *et al.*, 2017). Poisson MLM does not assume normality. Linear regression required the independence of observations and normality assumptions to be met, and Q-Q plot inspection supported normality.

Statistical analyses

A two-tailed, paired samples *t*-test with an alpha of .05 was used to compare the pre- and post-Mission SUDS scores and the baseline and 30-day PHQ-9, GAD-7 and WEMWBS scores. Assumptions of normality and normality of difference scores were met upon Q-Q plot

inspection. MLM and linear regression were then used to investigate the predictors and impacts of engagement. MLM is appropriate for hierarchical data, in which participants are nested within a group such as an organization, and repeated measurements, where multiple observations are nested within each participant (Finch, 2017). MLM allows variables to vary between and/or within participants. Variables that are held constant are termed 'fixed effects' and variables that vary are called 'random effects'. In the present study, hypotheses comprising variables measured at different time points were assessed via MLM. Two-level MLM was used in which variables under study (level 1) were nested within participants (level 2). Linear regression was considered sufficient for variables collected at a singular time point.

To test Hypotheses 1 and 2, random slope multi-level models A and B were created, using level of support and activity type as predictor variables (level 1). MLM was used to account for participant variability associated with participants completing multiple Missions, as Missions may have provided differing extents of support and some participants may have completed both activity types at different proportions (level 2). Each predictor was dichotomously scored, with the level of support being assigned a value of 1 for Missions accompanied by both a written and diagrammatic description. A value of 0 was given for Missions accompanied solely by a written description (see Supplementary material for examples of Missions differentiated by their level of support). When scoring activity type, mastery-enhancing activities were assigned a value of 1 and pleasure-enhancing activities were assigned a value of 0. Model A used Mission completion as its outcome variable, defined as the total number of Missions completed, indicated by the number of Missions for which a post-Mission rating was completed. Due to Mission completion being a count variable, Poisson MLM was used for Model A. Model B used the HRS-MA-engagement score as its outcome variable.

Random intercept multi-level model C tested Hypothesis 3a, using Mission completion and the HRS-MA-engagement score as predictor variables and the MHR score as the outcome (level 1). Participants may have reported different appraisals as a function of time in addition to engagement, with MLM taking participant variability into consideration (level 2).

Linear regression models were used to test Hypothesis 3b, each using Mission completion and the HRS-MA-engagement score as predictor variables. Model D used the HRS-MA-beliefs score as its outcome variable, whereas Model E used the HRS-MA-consequences score.

Hypothesis 4a was tested using Model F, a random intercept multi-level model which similarly used Mission completion and the HRS-MA-engagement score as predictors. In contrast, the outcome variable used was the change in SUDS score, considered at both the between- and within-participant level (levels 1 and 2, respectively). The change in SUDS score was calculated by subtracting the post-Mission SUDS score from the pre-Mission score, with a positive score indicating an improvement in distress level.

Finally, Hypothesis 4b was tested by linear regression Models G to I. These models used Mission completion and the HRS-MA-engagement score as predictors, with the outcome variable being the change in PHQ-9 score for Model G, GAD-7 score for Model H, and WEMWBS score for Model I. The change in PHQ-9 and GAD-7 scores were calculated by subtracting the 30-day scores from the baseline scores, with a positive score indicating an improvement in symptoms. The change in WEMWBS was calculated by subtracting the baseline scores from the 30-day scores, with a positive score indicating an increase in wellbeing. The analyses were conducted using the R Project for Statistical Computing (version 3.5.3).

Results

Missions completed by each participant ranged from 1 to 65 (mean = 5.46, $SD = 8.31$, 95% CI [4.37, 6.55]), with the majority of participants completing one Mission and consequently scoring low on this engagement measure. Figure 1 displays Missions completed by the sample

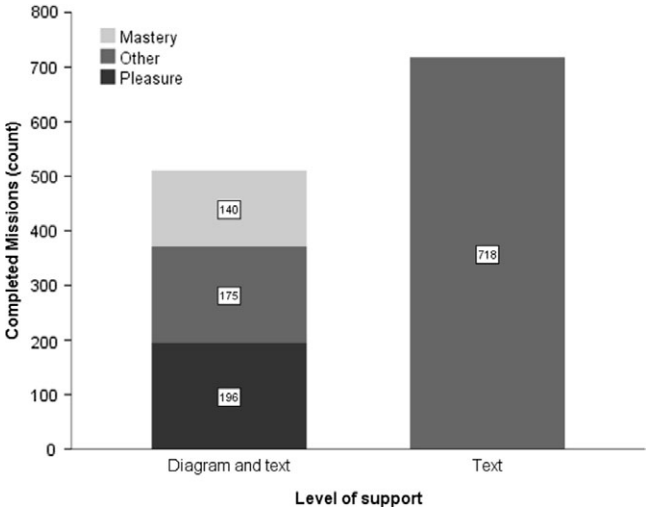

Figure 1. Completed missions by their level of support and activity type. 'Mastery' denotes missions based on mastery-enhancing behavioural activation (BA); 'Other' denotes missions based on a cognitive behaviour therapy technique other than BA; 'Pleasure' denotes missions based on pleasure-enhancing BA.

according to their level of support and activity type. Completed Missions were typically accompanied solely by text (58.40%) and based on a CBT technique other than BA (72.70%).

On the HRS-MA, the most common response was in the middle range for the majority of items, with a typical rating of 2 (range 0–4), indicating a moderate level of engagement and appraisal. Conversely, lower scores were typical for the 'difficulty' and 'obstacles' items of the HRS-MA-engagement and higher scores were generally given for the 'comprehension', 'rationale' and 'specificity' items of the HRS-MA-beliefs. The mean and standard deviation scores of the HRS-MA items are given in the Supplementary material.

The average HRS-MA-engagement score was low to moderate, compared with the maximum possible score of 16. The mean score on the HRS-MA-beliefs, HRS-MA-consequences and MHR were all moderate, with possible score ranges of 0 to 20, 0 to 12, and 0 to 10, respectively. These scores indicate that the sample tended to provide moderately positive appraisals of Missions. Average scores on the pre-Mission SUDS and baseline PHQ-9, GAD-7 and WEMWBS were also located in the middle of these questionnaires' score ranges, with maximum possible scores of 10, 27, 21 and 70, respectively.

Participants reported a positive change upon completing the post-Mission SUDS and 30-day PHQ-9 and WEMWBS. Participants SUDS significantly decreased with a medium effect size ($t_{1228} = 24.86$, $p < .001$, $d = 0.66$). Non-significant differences were observed for the PHQ-9 ($t_{234} = 1.34$, $p = .18$, $d = 0.08$), GAD-7 ($t_{233} = -0.36$, $p = .72$, $d = -0.02$) and WEMWBS ($t_{58} = 0.02$, $p = .99$, $d = 0.002$). According to the cut-off scores of these questionnaires, participants tended to remain exhibiting moderate distress, depression and anxiety severity, and below average mental wellbeing.

Hypothesis 1

The level of informational support was examined for its impact on engagement. When Missions were accompanied solely by a text description, the average engagement level was represented by a fixed effects intercept of 1.24 for the number of Missions completed ($SE = 0.13$, $p < .001$) and 6.15 for HRS-MA-engagement ($SE = 0.15$, $p < .001$). On average, Missions that provided both a

diagram and text description were not significantly associated with an increase in Mission completion ($b = 0.02$, $SE = 0.09$, $p = .83$) or HRS-MA-engagement ($b < 0.001$, $SE < 0.001$, $p = 1$). The random effects slopes revealed minimal individual difference (variance < 0.001).

Hypothesis 2

Engagement with mastery-enhancing BA activities was compared to engagement with pleasure-enhancing BA activities. Pleasure-enhancing activities were associated with a fixed effects intercept of 1.60 for Mission completion ($SE = 0.20$, $p < .001$) and 6.31 for HRS-MA-engagement ($SE = 0.29$, $p < .001$). Mastery-enhancing activities did not significantly predict an increase in Mission completion ($b = 0.004$, $SE = 0.18$, $p = .98$) or HRS-MA-engagement ($b < 0.001$, $SE < 0.001$, $p = 1$). The random effects slopes revealed minimal individual difference (variance < 0.001).

Hypothesis 3a

Engagement was assessed for its impact on the MHR score. An increase in Mission completion did not significantly predict an increase in the MHR score ($b = 0.02$, $SE = 0.02$, $p = .31$). Similarly, the HRS-MA-engagement score failed to significantly predict an increase in the score on the MHR ($b < 0.001$, $SE = 0.13$, $p = 1$).

Hypothesis 3b

The effect of engagement on scores on the HRS-MA subscales assessing Mission appraisal, HRS-MA-beliefs and HRS-MA-consequences, was considered. Mission completion and HRS-MA-engagement accounted for a significant 38% of the variability in HRS-MA-beliefs, $R^2 = .38$, adjusted $R^2 = .35$, $F_{2,45} = 13.87$, $p < .001$. Higher scores on Mission completion and HRS-MA-engagement were significantly associated with an increase in HRS-MA-beliefs, $b = 0.16$, $SE = 0.05$, $p = .003$ and $b = 0.88$, $SE = 0.21$, $p < .001$, respectively. A significant 59% of the variability in HRS-MA-consequences was attributable to the two engagement measures, $R^2 = .59$, adjusted $R^2 = .57$, $F_{2,45} = 31.73$, $p < .001$. Mission completion and HRS-MA-engagement each significantly predicted an increase in HRS-MA-consequences, $b = 0.08$, $SE = 0.03$, $p = .003$ and $b = 0.74$, $SE = 0.10$, $p < .001$, respectively.

Hypothesis 4a

The present study investigated the impact of engagement on the change in SUDS score. The average change in SUDS score decreased for each Mission completed ($b = -0.001$, $SE = 0.03$), although this relationship was not statistically significant ($p = .98$). Likewise, an increase in HRS-MA-engagement did not significantly predict a decrease in the average change in SUDS score ($b = -0.16$, $SE = 0.21$, $p = .47$).

Hypothesis 4b

Engagement was assessed for the ability to predict a change in PHQ-9, GAD-7 and WEMWBS score. Mission completion and HRS-MA-engagement contributed to a non-significant 2% of the variability in the change in PHQ-9 score ($R^2 = .02$, adjusted $R^2 = -.03$, $F_{2,45} = 0.34$, $p = .71$), 1% of the variability in the change in GAD-7 score ($R^2 = .01$, adjusted $R^2 = -.03$, $F_{2,45} = 0.23$, $p = .77$), and 15% of the variability in the change in WEMWBS score ($R^2 = .15$, adjusted $R^2 = -.04$, $F_{2,9} = 0.80$, $p = .48$). An increase in the number of Missions completed was not significantly associated with an increase in the change in PHQ-9 score ($b = 0.06$, $SE = 0.08$, $p = .45$),

GAD-7 score ($b = 0.02$, $SE = 0.07$, $p = .79$) or WEMWBS score ($b = 0.03$, $SE = 0.59$, $p = .96$). Similarly, an increase in HRS-MA-engagement did not significantly predict a change in PHQ-9 score ($b = -0.12$, $SE = 0.33$, $p = .72$), GAD-7 score ($b = -0.20$, $SE = 0.29$, $p = .50$) or WEMWBS score ($b = 0.99$, $SE = 0.80$, $p = .25$).

Discussion

The present study examined engagement with smartphone-delivered CBT interventions in the MoodMission app. Engagement was measured by the number of Missions completed and the HRS-MA-engagement. We tested the hypothesis that interventions with diagrammatic and text descriptions would have higher levels of engagement than those with text alone (Hypothesis 1). We also tested the hypothesis that the type of BA intervention (i.e. mastery enhancing *vs* pleasure enhancing) would yield different engagement levels (Hypothesis 2). Significant differences were not obtained and Hypotheses 1 and 2 were not supported. One possible reason for our results is the limited amount of psychoeducation delivered per technique in MoodMission, a natural limitation of the amount of information that can be presented on a small screen. Diagrams were also limited in their complexity and range. However, this is the first study to examine the mode of rationale delivery and the type of CBT intervention within a smartphone app and further research is required. Future studies should investigate the effects of other levels of psychoeducation support delivered via MHapp, including videos, photos, vignettes and audio recordings (Hidalgo-Mazzei *et al.*, 2018).

We did find significant relationships between measures of engagement and HRS-MA beliefs and consequences. The HRS measure has been adapted from the face-to-face delivery of CBT and examines both engagement and the theoretically meaningful determinants of engagement (see review in Kazantzis *et al.*, 2005). We tested the HRS modification for MHapps (HRS-MA: Bakker and Kazantzis, 2020) and found acceptable levels of reliability and factorial validity. Therefore, it is meaningful that participants who were more engaged with the app and used it more frequently also had a more positive belief in their app use, and experienced a greater sense of collaboration, progress and mastery, and having a clearer rationale for participating in the activities. Our findings underscore the importance of theory as the basis for designing the manner in which rationales for interventions are communicated to MHapp users (Torous *et al.*, 2018). Future research should examine whether engagement–outcome relations in smartphone-delivered CBT interventions are similar to those found in face-to-face therapy (Kazantzis *et al.*, 2018).

The present study had limitations that should be acknowledged. First, the present investigation was a naturalistic study; there was no experimental control over which Missions were chosen by participants and for what reasons. This offers advantages, such as improved ecological validity, but may have introduced a degree of variance that made it difficult to detect meaningful effects. The present sample was also modest compared with previous research with MoodMission (e.g. Bakker and Rickard, 2019) and there was a high level of attrition. Larger sample sizes are required to detect the small effect sizes of MHapp interventions, particularly when those participants are naturally recruited as typical app users without financial incentive. Retention rates for MHapps without financial incentive are commonly low when compared with other interventions (Bauer *et al.*, 2020), so MoodMission's retention rate (5.5%) was consistent with average Apple and Android app retention rates in the general population; the rate was 4% after 90 days (Statista, 2016).

The present study provides further support for the HRS items, and the modification as a measure of engagement in an MHapp context. Other MHapps could incorporate the HRS-MA and assess for user engagement. Further research is required to examine the engagement–outcome relationship, its mediators, and those stable individual characteristics that serve as

moderators, using a larger sample size. Increasing understanding of engagement, including how it can be measured and the importance of user appraisal, can be used as a basis for enhancing MHapp design.

Acknowledgements. None.

Financial support. None.

Conflicts of interest. None.

Ethics statements. The authors abided by the Ethical Principles of Psychologists and Code of Conduct as indicated by the ARC/NHMRC guidelines in accordance with the APS. The research was approved by the Monash University Human Research Ethics Committee (reference number 18294). MoodMission users consented to their data being used for research by consenting to MoodMission's privacy policy. For minors, the privacy policy contains an explanatory statement to be read by guardians.

Supplementary material. To view supplementary material for this article, please visit: https://doi.org/10.1017/S1352465820000922

References

Bakker, D., & Kazantzis, N. (2020). *Homework Rating Scale – Mobile Application Version* [Measurement instrument]. In N. Kazantzis (ed), *Using Homework Assignments in Cognitive Behavior Therapy*. Routledge.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2018a). A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behaviour Research and Therapy*, *109*, 75–83. https://doi.org/10.1016/j.brat.2018.08.003

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2018b). Development and pilot evaluation of smartphone-delivered cognitive behavior therapy strategies for mood- and anxiety-related problems: MoodMission. *Cognitive and Behaviour Practice*, *25*, 496–514. https://doi.org/10.1016/j.cbpra.2018.07.002

Bakker, D., & Rickard, N. (2018). Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. *Journal of Affective Disorders*, *227*, 432–442. https://doi.org/10.1016/j.jad.2017.11.016

Bakker, D., & Rickard, N. (2019). Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. *Australian Psychologist*, *54*, 245–260. https://doi.org/10.1111/ap.12383

Bakker, M., & Wicherts, J. M. (2014). Outlier removal and the relation with reporting errors and quality of psychological research. *PLoS One*, *9*, e103360. https://doi.org/10.1371/journal.pone.0103360

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall.

Bauer, M., Glenn, T., Geddes, J., Gitlin, M., Grof, P., Kessing, L. V., Monteith, S., Faurholt-Jepsen, M., Severus, E., & Whybrow, P. C. (2020). Smartphones in mental health: a critical review of background issues, current status and future concerns. *International Journal of Bipolar Disorders*, *8*. https://doi.org/10.1186/s40345-019-0164-x

Deady, M., Johnston, D. A., Glozier, N., Milne, D., Choi, I., Mackinnon, A., Mykletun, A., Calvo, R. A., Gayed, A., Bryant, R., Christensen, H., & Harvey, S. B. (2018). Smartphone application for preventing depression: study protocol for a workplace randomised controlled trial. *BMJ Open*, *8*, e020510. https://doi.org/10.1136/bmjopen-2017-020510

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*, 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Dimidjian, S., Barrera, M., Martell, C., Munoz, R. F., & Lewinsohn, P. M. (2011). The origins and current status of behavioral activation treatments for depression. *Annual Review of Clinician Psychology*, *7*, 1–38. https://doi.org/10.1146/annurev-clinpsy-032210-104535

Duncan, S. F., Childs, G. R., & Larson, J. H. (2010). Perceived helpfulness of four different types of marriage preparation interventions. *Family Relations*, *59*, 623–636. https://doi.org/10.1111/j.1741-3729.2010.00628.x

Economides, M., Martman, J., Bell., M. J., & Sanderson, B. (2018). Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. *Mindfulness*, *9*, 1584–1593. https://doi.org/10.1007/s12671-018-0905-4

Finch, H. (2017). Multilevel modeling in the presence of outliers: a comparison of robust estimation methods. *Psicologica*, *38*, 57–92. https://www.uv.es/psicologica/articulos1.17/3FINCH.pdf

Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *Journal of Affective Disorders*, *218*, 15–22. https://doi.org/10.1016/j.jad.2017.04.046

Fung, W. K., & Xu, X. C. (2010). Estimation and robustness of linear mixed models in credibility context. *Variance: Advancing the Science of Risk*, *2*, 66–80. http://www.variancejournal.org/issues/04-01/66.pdf

Giosan, C., Cobeanu, O., Mogoaşe, C., Szentagotai, A., Mureşan, V., & Boian, R. (2017). Reducing depressive symptomatology with a smartphone app: study protocol for a randomized, placebo-controlled trial. *Trials*, *18*, e215. https://doi.org/10.1186/s13063-017-1960-1

Hara, K. M., Aviram, A., Constantino, M. J., Westra, H. A., & Antony, M. M. (2017). Therapist empathy, homework compliance, and outcome in cognitive behavioral therapy for generalised anxiety disorder: partitioning within- and between-therapist effects. *Cognitive Behaviour Therapy*, *46*, 375–390. https://doi.org/10.1080/16506073.2016.1253605

Hidalgo-Mazzei, D., Reinares, M., Mateu, A., Nikolova, V. L., Bonnin, C. D. M., Samalin, L., Garcia-Estela, A., Perez-Sola, V., Young, A. H., Strejilevich, S., Vieta, E., & Colom, F. (2018). OpenSIMPLe: a real-world implementation feasibility study of a smartphone-based psychoeducation programme for bipolar disorder. *Journal of Affective Disorders*, *241*, 436–445. https://doi.org/10.1016/j.jad.2018.08.048

Holdsworth, E., Bowen, E., Brown, S., & Howat, D. (2014). Client engagement in psychotherapeutic treatment and associations with client characteristics, therapist characteristics, and treatment factors. *Clinical Psychology Review*, *34*, 428–450. https://doi.org/10.1016/j.cpr.2014.06.004

Jacobs, N. W., Berduszek, R. J., Dijkstra, P. U., & van der Sluis, C. K. (2017). Validity and reliability of the Upper Extremity Work Demands Scale. *Journal of Occupational Rehabilitation*, *27*, 520–529. https://doi.org/10.1007/s10926-016-9683-9

Kazantzis, N., Brownfield, N., Mosely, L., Usatoff, A., & Flighty, A. (2017). Homework in cognitive behavioral therapy: a systematic review of adherence assessment in anxiety and depression treatment (2011–2016). *Psychiatric Clinics of North America*, *40*, 625–639. https://doi.org/10.1016/j.psc.2017.08.001

Kazantzis, N., Deane, F., Ronan, K. R., & L'Abate, L. (eds) (2005). *Using Homework Assignments in Cognitive Behavior Therapy*. Routledge.

Kazantzis, N., & L'Abate, L. (2005). Theoretical foundations. In N. Kazantzis, F. P. Deane., K. R. Ronan., & L. L'Abate (eds), *Using Homework Assignments in Cognitive Behavior Therapy* (pp. 9–33). Routledge.

Kazantzis, N., Luong, H. K., Usatoff, A. S., Impala, T., Yew, R. Y., & Hofmann, S. G. (2018). The processes of cognitive behavioral therapy: a review of meta-analyses. *Cognitive Therapy and Research*, *42*, 349–357. https://doi.org/10.1007/s10608-018-9920-y

Kazantzis, N., Whittington, C., & Dattilio, F. (2010). Meta-analysis of homework effects in cognitive and behavioral therapy: a replication and extension. *Clinical Psychology: Science and Practice*, *17*, 144–156. https://doi.org/10.1111/j.1468-2850.2010.01204.x

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Martell, C. R., Dimidjian, S., & Lewinsohn, P. M. (2010). Behavioral activation therapy. In N. Kazantzis, M. A. Reinecke, & A. Freeman (eds), *Cognitive and Behavioral Theories in Clinical Practice* (pp. 193–217). Guilford Press.

McDonald, B. R., & Morgan, R. D. (2013). Enhancing homework compliance in correctional psychotherapy. *Criminal Justice and Behaviour*, *40*, 814–828. https://doi.org/10.1177/0093854813480781

MoodMission Pty Ltd (2019). MoodMission (version 1.4.7) [Mobile application software]. https://apps.apple.com/au/app/moodmission/id1140332763

Ng, M. M., Firth, J., Minen, M., & Torous, J. (2019). User engagement in mental health apps: a review of measurement, reporting, and validity. *Psychiatric Services*, *70*, 538–544. https://doi.org/10.1176/appi.ps.201800519

Ozonur, D., Akdur, H. T. K., & Bayrak, H. (2017). Comparisons of tests of distributional assumption in Poisson regression model. *Communications in Statistics – Simulation and Computation*, *46*, 6197–6207. https://doi.org/10.1080/03610918.2016.1202267

Parrish, D. E., Oxhandler, H. K., Duron, J. F., Swank, P., & Bordnick, P. (2016). Feasibility of virtual reality environments for adolescent social anxiety disorder. *Research on Social Work Practice*, *26*, 825–835. https://doi.org/10.1177/1049731514568897

Pepinsky, T. (2018). A note on listwise deletion versus multiple imputation. *Political Analysis*, *26*, 480–488. https://doi.org/10.1017/pan.2018.18

Richardson, D., Hosemans, D., & Kazantzis, N. (2020). *Engagement with Health-Promoting Activities: Psychometric Evaluation of the Homework Rating Scale in a Large Community Sample*. Manuscript in preparation.

Sachsenweger, M. A., Fletcher, R. B., & Clarke, D. (2015). Pessimism and homework in CBT for depression. *Journal of Clinical Psychology*, *71*, 1153–1172. https://doi.org/10.1002/jclp.22227

Sagar, R., & Pattanayak, R. D. (2015). Use of smartphone apps for mental health: can they translate to a smart and effective mental health care? *Journal of Mental Health and Human Behaviour*, *20*, 1–3. https://doi.org/10.4103/0971-8990.164791

Skinner, B. F. (1938). *The Behaviour of Organisms: An Experimental Analysis*. Appleton-Century.

Snijders, T. A., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In J. de Leeuw, E. Meijer, H. Goldstein, & J. de Deleeuw (eds), *Handbook of Multilevel Analysis* (pp. 141–175). Springer.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalised anxiety disorder. *Archives of Internal Medicine*, *166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Statista (2016, Sept 15). Worldwide mobile app retention rate during the first 90 days of ownership as of March 2016, by mobile platform. Available at: https://www.statista.com/statistics/243728/worldwide-mobile-app-user-retention-by-mobile-plaform/

Tanner, B. A. (2012). Validity of global physical and emotional SUDS. *Applied Psychophysiology and Biofeedback*, *37*, 31–34. https://doi.org/10.1007/s10484-011-9174-x

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, *5*, 63. https://doi.org/10.1186/1477-7525-5-63

Torous, J., Nicholas, J., Larsen, M. E., Firth, J., & Christensen, H. (2018). Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evidence Based Mental Health*, *21*, 116–119. https://doi.org/10.1136/eb-2018-102891

Wolpe, J. (1969). *The Practice of Behavior Therapy*. Pergamon Press.

Cite this article: Aizenstros A, Bakker D, Hofmann SG, Curtiss J, and Kazantzis N. Engagement with smartphone-delivered behavioural activation interventions: a study of the MoodMission smartphone application. *Behavioural and Cognitive Psychotherapy*. https://doi.org/10.1017/S1352465820000922