JMIR Mental Health

Internet interventions, technologies and digital innovations for mental health and behaviour change
Volume 9 (2022), Issue 5 ISSN: 2368-7959 Editor in Chief: John Torous, MD

Contents

Original Papers

Smartphone Sensor Data for Identifying and Monitoring Symptoms of Mood Disorders: A Longitudinal Observational Study (e35549)
Taylor Braund, May Zin, Tjeerd Boonstra, Quincy Wong, Mark Larsen, Helen Christensen, Gabriel Tillman, Bridianne O’Dea. 3

Momentary Self-regulation: Scale Development and Preliminary Validation (e35273)
Emily Scherer, Sunny Kim, Stephen Metcalf, Mary Sweeney, Jialing Wu, Haiyi Xie, Gina Mazza, Matthew Valente, David MacKinnon, Lisa Marsch. 16

The Role of Emotion Regulation and Loss-Related Coping Self-efficacy in an Internet Intervention for Grief: Mediation Analysis (e27707)
Jeannette Brodbeck, Thomas Berger, Nicola Biesold, Franziska Rockstroh, Stefanie Schmidt, Hansjoerg Znoj. 90

Ownership, Use of, and Interest in Digital Mental Health Technologies Among Clinicians and Young People Across a Spectrum of Clinical Care Needs: Cross-sectional Survey (e30716)
Imogen Bell, Andrew Thompson, Lee Valentine, Sophie Adams, Mario Alvarez-Jimenez, Jennifer Nicholas. 103

Therapeutic Alliance in Online and Face-to-face Psychological Treatment: Comparative Study (e36775)
Josep Mercadal Rotger, Victor Gabré. 118

Engagement, Predictors, and Outcomes of a Trauma Recovery Digital Mental Health Intervention: Longitudinal Study (e35048)
Carolyn Yeager, Charles Benight. 128

Virtual Reality Behavioral Activation for Adults With Major Depressive Disorder: Feasibility Randomized Controlled Trial (e35526)
Margot Paul, Kim Bullock, Jeremy Bailenson. 146

A Group-Facilitated, Internet-Based Intervention to Promote Mental Health and Well-Being in a Vulnerable Population of University Students: Randomized Controlled Trial of the Be Well Plan Program (e37292)
Daniel Fassnacht, Kathina Ali, Joep van Agteren, Matthew Iasiello, Teri Mavrangelos, Gareth Furber, Michael Kyrios. 161

Reviews

Acceptability of Web-Based Mental Health Interventions in the Workplace: Systematic Review (e34655)
Johanna Scheutzow, Chris Attoe, Joshua Harwood. 36
Smartphone Sensor Data for Identifying and Monitoring Symptoms of Mood Disorders: A Longitudinal Observational Study

Taylor A Braund¹,², PhD; May The Zin², BMed, MD; Tjeerd W Boonstra¹,²,³, PhD; Quincy J J Wong¹,⁴, PhD; Mark E Larsen¹,², DPhil; Helen Christensen¹,², PhD; Gabriel Tillman⁵, PhD; Bridianne O’Dea¹,², PhD

¹Black Dog Institute, University of New South Wales, Sydney, Australia
²Faculty of Medicine and Health, University of New South Wales, Sydney, Australia
³Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands
⁴School of Psychology, Western Sydney University, Sydney, Australia
⁵School of Science, Psychology and Sport, Federation University, Ballarat, Australia

Abstract

Background: Mood disorders are burdensome illnesses that often go undetected and untreated. Sensor technologies within smartphones may provide an opportunity for identifying the early changes in circadian rhythm and social support/connectedness that signify the onset of a depressive or manic episode.

Objective: Using smartphone sensor data, this study investigated the relationship between circadian rhythm, which was determined by GPS data, and symptoms of mental health among a clinical sample of adults diagnosed with major depressive disorder or bipolar disorder.

Methods: A total of 121 participants were recruited from a clinical setting to take part in a 10-week observational study. Self-report questionnaires for mental health outcomes, social support, social connectedness, and quality of life were assessed at 6 time points throughout the study period. Participants consented to passively sharing their smartphone GPS data for the duration of the study. Circadian rhythm (ie, regularity of location changes in a 24-hour rhythm) was extracted from GPS mobility patterns at baseline.

Results: Although we found no association between circadian rhythm and mental health functioning at baseline, there was a positive association between circadian rhythm and the size of participants’ social support networks at baseline ($r=0.22; P=.03; R^2=0.049$). In participants with bipolar disorder, circadian rhythm was associated with a change in anxiety from baseline; a higher circadian rhythm was associated with an increase in anxiety and a lower circadian rhythm was associated with a decrease in anxiety at time point 5.

Conclusions: Circadian rhythm, which was extracted from smartphone GPS data, was associated with social support and predicted changes in anxiety in a clinical sample of adults with mood disorders. Larger studies are required for further validations. However, smartphone sensing may have the potential to monitor early symptoms of mood disorders.

(JMIR Ment Health 2022;9(5):e35549) doi:10.2196/35549

KEYWORDS
depression; bipolar disorder; sensors; mobile app; circadian rhythm; mobile phone
Introduction

Background

Circadian rhythm is an endogenous mammalian process that helps to regulate an individual’s activity over 24-hour cycles [1]. The temporal organization of the circadian rhythm is important for balancing physiological functions of body temperature and hormones [2] as well as behavioral schedules such as daily activities, sleep-wake patterns [3], and how individuals interact with each other in their social networks [4]. As such, its disruption can have substantial impacts on the health of individuals.

Circadian rhythm irregularities have been found to be closely related to the onset and clinical manifestations of mood disorders [5] such as major depressive disorder (MDD) and bipolar disorder (BD). Many studies investigating the relationship between circadian rhythm and mental health symptoms have used behavior questionnaires and self-reports of sleep-wake activity [6-11]. With different rhythm measures, studies have found correlations between circadian rhythm irregularity and greater symptoms of depression [12,13] and anxiety [13], poorer quality of life [12], and lower life satisfaction [13] and social support [14]. Moreover, disruptions in circadian rhythms differently affect episodes of mania and depression [15,16]. For example, melatonin—the hormone responsible for sustaining the sleep-wake cycle—has been found to be increased during mania and decreased during depression [16]. However, past studies are limited because of the methodological issues associated with low adherence rates for reporting and biases because of memory recall and subjectivity. Digital devices may provide a more nuanced identification and detection of circadian rhythm and present a clinical advantage for monitoring and managing mood disorders.

With advances in technology, mobile and wearable devices can now track specific behaviors and interactions [17], allowing researchers to address some of the drawbacks of traditional methods. More recently, modern smartphones equipped with powerful contextual sensors have made it feasible to capture participants’ symptoms with less burden [18]. Passive sensing through smartphones has enabled the gathering of information by operating in the background without the need for any input from the user [19] and without interrupting the user’s habitual routine. One of the commonly used sensors to study mood-related symptoms is a GPS, which maps a participant’s location [20-23]. For example, lesser mobility was found to be correlated with higher levels of mania [20] and depressive symptoms [20,21], with the relationships varying in strength depending on the timing throughout the week [22]. In addition, people with higher depressive scores showed decreased circadian rhythm, clustered around locations for longer periods, and had more disruptions to movement and variations in locations visited [23]. Other smartphone sensors including Bluetooth scans to estimate an individual’s social network based on proximity to other individuals’ mobile devices [18], device activity to denote social activity and predict communication frequencies [20], and accelerometers to infer physical and social activity [20,24], all of which may be used as proxies for behavioral markers of mood disorders. Given the ubiquity of smartphones today, this nonintrusive and objective method allows large-scale and population-level applications [25], which could help identify mood disorders early in the course of illness and provide a therapeutic tool through behavioral modification [26].

These past studies show promising findings, but they are hampered by significant limitations, including the lack of clinical participants [23] and small sample sizes [20,21,24,27]. Although some studies assessed participants through well-validated self-reported questionnaires during enrollment, most studies did not use a clinical interview conducted by a qualified clinician at baseline to determine diagnosis [18,21,22]. Even when the diagnoses were well established, the small sample sizes of most studies [20,21,24,27] limited statistical evidence and generalization. Furthermore, many studies used researcher-administered mobile phones rather than participants’ own smartphones [20,22-24,27], which may account for low adherence of completed self-assessments and missing data because participants reported forgetting to use the phone or carry it with them. Although initial studies have provided preliminary evidence for the link between smartphone measurements and clinical symptoms of mood disorders, additional research is required to provide stronger evidence for the relationship between sensor data and depressive symptoms.

Objectives

The aim of this study was to determine the relationship between circadian rhythm in GPS data and symptoms of mental health measures across 6 time points among a clinical sample of adults diagnosed with MDD or BD. It was hypothesized that higher baseline circadian rhythm would be negatively associated with baseline symptoms of depression, mania, and anxiety [23] and positively associated with social support and quality of life [12-14]. Furthermore, symptoms of depression, anxiety, and mania would fluctuate over time in both disorders [28-30]. Finally, baseline circadian rhythm would predict change in mental health symptoms over time [22], with patients with BD showing a greater sensitivity to change [15,16].

Methods

Design

A 10-week prospective longitudinal design was used. Mental health measures were administered at baseline, on a fortnightly basis, and end point (10 weeks after baseline).

Participants

Participants were recruited from the Black Dog Institute Depression and Bipolar Clinic, an in-person psychiatric diagnostic and treatment service for adults located in Sydney, New South Wales, Australia (for the CONSORT [Consolidated Standards of Reporting Trials] chart, see Multimedia Appendix 1). Adults are referred to this service by their general practitioner. Clinic administrative staff approached potential participants after they had completed their in-person diagnostic assessment with the trained psychiatrist. Participants were invited to participate in the study if their psychiatric assessment revealed a diagnosis of MDD or BD according to the criteria in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

https://mental.jmir.org/2022/5/e35549
Edition [31], were aged between 18 and 65 years, fluent in
English, and owned a compatible smartphone (Android
version≥4.4 or iOS version≥8). Participants were excluded if
they had severe mood disturbance or suicidality where
intervention was required as assessed by clinic staff, had current
or past psychosis, or lacked access to a suitable smartphone for
the duration of the study. Clinic administrative staff registered
eligible participants through the study website and emailed the
consent form to them. Participation was voluntary, and no
reimbursement, financial incentive, or reward was provided.

Ethics Approval
Ethics approval was obtained from the University of New South
Wales Human Research Ethics Committee (HC17252).

Sample Size
A sample size of 76 was required to detect a weak association
(\(r=-0.3\)) between circadian rhythm and mental health
functioning, with statistical power level 0.8 and \(\alpha=0.05\). However, considering participant dropout and loss of
smartphone data for some participants, the target was set at 200,
with 100 participants for each diagnostic group.

Procedure
After they provided consent, participants were instructed to
install the study smartphone app Socialise [26] to complete the
mental health measures and enable data collection (see
Multimedia Appendix 2 for schedule). Participants were then
instructed to use their phones normally for the study duration,
with the app open in the background. Socialise was configured
to automatically record GPS data and to prompt participants to
complete the study questionnaires. Socialise conducted GPS
data acquisition scans every 3, 4, 5, and 8 minutes.

Measures
Circadian Rhythm
Circadian rhythm (also known as circadian movement [22] and
quotidian movement [32]) was defined as the extent to which
an individual’s sequence of locations followed a 24-hour rhythm
[23]. This was determined at baseline based on changes in GPS
location measured during the first 2 weeks of the study only
when sufficient data were available. Least squares spectral
analysis was performed to estimate the amount of energy that
fell into 24-hour frequency bins [33]. Circadian rhythm was then calculated as the logarithm of the sum of energy for
longitude and latitude (Multimedia Appendix 3 [22,23,33]). Hence, participants who regularly change their location at the
same time each day will show a stronger 24-hour rhythm in
their GPS data and have a higher circadian rhythm. Conversely,
participants with irregular movement patterns (changing their
location at different times each day) will have a lower circadian
rhythm.

Depressive Symptoms
The Patient Health Questionnaire-9 (PHQ-9) [34] is a 9-item
self-report questionnaire that measures the frequency of
depressive symptoms during the past 2 weeks. Items are scored
on a 4-point Likert scale of 0 (not at all) to 3 (nearly every day)
and summed for a total score (range 0 to 27), with higher scores
indicating greater depression. The total sum also corresponds
to depression severity: none to minimal (0 to 4), mild (5 to 9),
moderate (10 to 14), moderately severe (15 to 19), and severe
(20 to 27). The PHQ-9 has good sensitivity (88%) and specificity
(88%) for detecting likelihood of major depression using a cutoff
of ≥10. Test-retest reliability in adults has been found to be
acceptable (\(r=0.83\)), and internal consistency was strong
(Cronbach \(\alpha=0.89\)) [34]. The PHQ-9 was administered at
baseline, then on a fortnightly basis, with 6 surveys administered
in the study period.

Anxiety Symptoms
The Generalized Anxiety Disorder Scale (GAD-7) [35] is a
7-item self-report questionnaire used to assess anxiety symptoms
during the past 2 weeks. Each item is scored on a 4-point Likert
scale of 0 (not at all) to 3 (nearly every day). Items are added
together for a total score (range 0 to 21), with higher scores
indicating greater anxiety. The total sum also corresponds to
anxiety severity: minimal (0 to 4), mild (5 to 9), moderate (10
to 14), and severe (15 to 21). The GAD-7 has good sensitivity
(89%) and specificity (82%) for detecting likelihood of anxiety
disorder using a cutoff of ≥10 [36]. It has been found to
demonstrate good internal consistency (Cronbach \(\alpha=0.92\)) and
test-retest reliability (\(r=0.83\)) [35]. The GAD-7 was administered
at baseline, then on a fortnightly basis, and at the end point,
with 6 surveys administered in the study period.

Mania Symptoms
The 5-item Altman Self-Rating Mania Scale (ASRM) is a
self-report questionnaire used to evaluate the presence and
severity of manic symptoms in the past week [37]. There are 5
groups of statements, each corresponding to scores of 0 to 4,
with 0 being unchanged behavior and 4 being frequent manic
thoughts or behavior. The item scores are summed to give a
total score (range 0 to 20), with a score of ≥6 indicating mania
and higher scores indicating greater severity of symptoms. This
scale has good sensitivity (87.3%) and specificity (85.5%) in
adults [37]. The ASRM was administered at baseline, then on
a fortnightly basis, and at end point, with 6 surveys administered
in the study period.

Social Connectedness
The degree to which participants felt connected to others was
measured using the 20-item Social Connectedness Scale-Revised
(SCS-R) [38]. The scale consists of 20 items, with 10 positively
worded questions and 10 negatively worded questions. It
assesses participants’ experience with social inclusion, safety
in their communities, and relationships with friends and families.
The items are rated on a scale of 1 (strongly disagree) to 6
(strongly agree). Negatively worded items are reverse scored
and summed with the scores of the remaining items to obtain a
total score (range 20 to 120). Higher scores indicate a greater
social connectedness to others. The SCS-R is a highly reliable
measure with an internal consistency of Cronbach \(\alpha=0.92\) [38].
The SCS-R was administered at baseline and end point.

Social Support
The 12-item Social Support Questionnaire (SSQ) [39] was used
to measure 2 constructs of social support (size of support
network and support satisfaction). The first 6 items evaluate the
number of people that the participant feels could provide social support in the situations presented. Participants are asked to list initials and the type of relationship for each individual. The size of an individual’s support network or the SSQ number score (range 0 to 9) is then calculated by adding the number of individuals listed across all the items and dividing by 6. Higher SSQ number scores indicate a larger and more diverse social support network. The second part of the questionnaire assesses participants’ satisfaction with the support they receive in each situation. Items are scored on a 6-point scale ranging from 1 (very dissatisfied) to 6 (very satisfied). Similarly, scores are summed and divided by 6 to obtain an SSQ satisfaction score (range 1 to 6). Higher SSQ satisfaction scores indicate greater satisfaction with the support received. The SSQ has demonstrated good test-retest reliability ($r=0.83$) and high internal consistency (Cronbach $\alpha=.97$) [39]. The SSQ was administered at baseline and end point.

Quality of Life

The Satisfaction With Life Scale (SWLS) [40] was used to measure participants’ life satisfaction and subjective quality of life. The self-report measure consists of 5 items rated on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The total sum of the items ranges from 5 to 35, with a score of 20 indicating a neutral score and lower scores indicating lower satisfaction with life. The total sum also corresponds to satisfaction level: extremely dissatisfied (5 to 9), dissatisfied (10 to 14), slightly below average in life satisfaction (15 to 19), neutral (20 to 24), highly satisfied (25 to 29), and extremely satisfied (30 to 35). The SWLS has good psychometric properties, including internal consistency with the coefficient Cronbach $\alpha$ ranging from .79 to .89 and test-retest reliability of $r=0.84$ over a 1-month period [41]. The SWLS was administered at baseline and end point.

Data Collection

The Socialise app was configured to automatically record GPS data. The app conducted GPS data acquisition scans every 3, 4, 5, and 8 minutes. The app was also configured to automatically schedule the study questionnaires at fortnightly intervals and prompt participants to complete them. Using internet connectivity, the app transferred all participant data to the Black Dog Institute research platform, hosted on University of New South Wales servers. Data were then downloaded and analyzed by the research team. Because of an issue with app connectivity, an error in the final survey scheduling was detected. This resulted in the final survey being delivered at 9 weeks after baseline instead of 10 weeks.

Statistical Analysis

We used 2-tailed $t$ tests to examine differences between participants with MDD and those with BD in circadian rhythm and mental health outcomes, social support, social connectedness, and quality of life at baseline. Chi-square tests were used to test whether the proportion of participants meeting cutoff scores of mental health disorder diagnoses at baseline differed between participants with MDD and those with BD. Linear regression models were used to test whether baseline circadian rhythm was associated with baseline mental health, social support, social connectedness, or quality of life. Interaction terms between circadian rhythm and diagnostic group were first included to test whether the relationships differed between diagnostic groups.

Mixed linear models were used to test whether mental health functioning changed over time. Models included mental health, social support, or social connectedness as the dependent variable; participant as the random effect; and time point as a fixed effect. Interaction terms between the fixed effects of time point and diagnostic group were first included to test whether changes in mental health and social functioning differed between people with MDD and those with BD. Post hoc tests were performed using estimated marginal means, with the false discovery rates (FDRs) for all $P$ values corrected to $q$ values using the Benjamini and Hochberg [42] procedure.

Mixed linear models were also used to test whether circadian rhythm moderated changes in mental health or social support/connectedness over time. Models included mental health, social support, or social connectedness as the dependent variable; subject as the random effect; and time point as a fixed effect. Interaction terms between the fixed effects of time point and circadian rhythm were then used to test whether baseline circadian rhythm moderated changes in mental health and social functioning. Interaction terms among the fixed effects of time point, circadian rhythm, and diagnostic group were first included to test whether the moderating effects of baseline circadian rhythm on changes in mental health and social functioning differed between people with MDD and those with BD. Post hoc tests were performed using estimated marginal means, with the FDRs for all $P$ values corrected to $q$ values using the Benjamini and Hochberg [42] procedure.

All analyses were conducted using R software (version 3.5.1; The R Foundation for Statistical Computing) [43]. Mixed linear models were tested using the lm4 package in R [44]. $P$ values for mixed linear models were calculated using the lmerTest package in R [45]. Post hoc tests were performed using the emmeans package in R [46] (see Multimedia Appendix 4 for supplementary analyses).

Results

Participants

During the recruitment period, 219 adults attended the Black Dog Institute clinic for a psychiatric assessment (Multimedia Appendix 1), of whom 162 (74%) were deemed eligible for the study by clinic administrative staff and were invited to participate. Of these 162 participants, 149 (92%) consented, downloaded the Socialise app, and completed the baseline mental health assessment; however, of these 149 participants, 3 (2%) withdrew shortly after completion of baseline assessment and 25 (16.8%) did not have sufficient GPS data, leaving a final sample that consisted of 121 (83.2%) participants, with attrition for the study questionnaires outlined in Multimedia Appendix 2.

Demographic and Clinical Characteristics

Table 1 outlines the baseline demographic and clinical characteristics of the final sample ($n=121$). The mean age of...
participants was 41.4 (SD 13.6; range 18 to 70) years; 65.3% (79/121) were women; and 66.1% (80/121) used iPhones. People with MDD had significantly higher anxiety ($t_{119}=-2.47; P=.02; \text{Cohen } d=-0.47, 95\% \text{ CI } -0.85 \text{ to } -0.09$) and likely cases of anxiety ($\chi^2_{1}=6.1; P=.01; \phi=0.24, 95\% \text{ CI } 0.06-0.42$), lower mania ($t_{119}=2.74; P=.007; \text{Cohen } d=0.52, 95\% \text{ CI } 0.14-0.90$), and a smaller social support network ($t_{119}=2.57; P=.01; \text{Cohen } d=0.49, 95\% \text{ CI } 0.11-0.87$) than participants with BD. There was no significant relationship between diagnosis and likely cases of depression, $\chi^2_{1}=0.9; P=.35$, or mania, $\chi^2_{1}=3.0; P=.08$. No other differences in mental health or social support/connectedness were found.

### Table 1. Baseline demographic and clinical characteristics of participants (N=121).

<table>
<thead>
<tr>
<th></th>
<th>Total sample</th>
<th>Depression (n=79)</th>
<th>Bipolar disorder (n=42)</th>
<th>$t$ test ($df$)</th>
<th>Chi-square test ($df$)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>41.41 (13.62)</td>
<td>41.63 (13.94)</td>
<td>41.00 (13.16)</td>
<td>$-0.24 (119)$</td>
<td>N/A</td>
<td>.81</td>
</tr>
<tr>
<td>Depressive symptoms (PHQ-9$^b$), mean (SD)</td>
<td>11.75 (6.67)</td>
<td>12.51 (7.09)</td>
<td>10.33 (5.61)</td>
<td>$-1.85 (119)$</td>
<td>N/A</td>
<td>.07</td>
</tr>
<tr>
<td>Anxiety symptoms (GAD-7$^c$), mean (SD)</td>
<td>8.21 (5.76)</td>
<td>9.14 (5.93)</td>
<td>6.48 (5.04)</td>
<td>$-2.60 (119)$</td>
<td>N/A</td>
<td>.01</td>
</tr>
<tr>
<td>Mania symptoms (ASRM$^e$), mean (SD)</td>
<td>4.12 (3.06)</td>
<td>3.58 (2.63)</td>
<td>5.14 (3.56)</td>
<td>2.50 (119)</td>
<td>N/A</td>
<td>.02</td>
</tr>
<tr>
<td>Quality of life (SWLS$^f$), mean (SD)</td>
<td>16.08 (7.51)</td>
<td>15.22 (7.43)</td>
<td>17.71 (7.49)</td>
<td>1.76 (119)</td>
<td>N/A</td>
<td>.08</td>
</tr>
<tr>
<td>Social connectedness (SCS-R$^g$), mean (SD)</td>
<td>73.57 (4.25)</td>
<td>73.71 (4.51)</td>
<td>73.29 (3.99)</td>
<td>$-0.21 (19)$</td>
<td>N/A</td>
<td>.83</td>
</tr>
<tr>
<td>Social support, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSQNS$^h$</td>
<td>2.63 (1.60)</td>
<td>2.37 (1.38)</td>
<td>3.13 (1.86)</td>
<td>2.57 (119)</td>
<td>N/A</td>
<td>.01</td>
</tr>
<tr>
<td>SSQSS$^i$</td>
<td>4.31 (1.29)</td>
<td>4.25 (1.24)</td>
<td>4.41 (1.39)</td>
<td>0.65 (119)</td>
<td>N/A</td>
<td>.52</td>
</tr>
<tr>
<td>Circadian rhythm, mean (SD)</td>
<td>$-5.46 (3.58)$</td>
<td>$-5.43 (3.71)$</td>
<td>$-5.50 (3.38)$</td>
<td>$-0.09 (94)$</td>
<td>N/A</td>
<td>.93</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>79 (65.3)</td>
<td>50 (63.3)</td>
<td>29 (69)</td>
<td>0.2 (1)</td>
<td>N/A</td>
<td>.67</td>
</tr>
<tr>
<td>Likely clinical case, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive symptoms (PHQ-9 score≥10)</td>
<td>69 (57)</td>
<td>48 (60.8)</td>
<td>21 (50)</td>
<td>N/A</td>
<td>0.9 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Anxiety symptoms (GAD-7 score≥10)</td>
<td>39 (32.2)</td>
<td>32 (40.5)</td>
<td>7 (16.7)</td>
<td>N/A</td>
<td>6.1 (1)</td>
<td>.01</td>
</tr>
<tr>
<td>Mania symptoms (ASRM score≥26)</td>
<td>33 (27.3)</td>
<td>17 (21.5)</td>
<td>16 (38.1)</td>
<td>N/A</td>
<td>3.0 (1)</td>
<td>.08</td>
</tr>
</tbody>
</table>

$^a$N/A: not applicable.
$^b$PHQ-9: Patient Health Questionnaire-9.
$^c$GAD-7: Generalized Anxiety Disorder Scale.
$^d$Italicization indicates values that met the significance threshold ($P<.05$).
$^e$ASRM: Altman Self-Rating Mania Scale.
$^f$SWLS: Satisfaction With Life Scale.
$^g$SCS-R: Social Connectedness Scale-Revised.
$^h$SSQNS: Social Support Questionnaire number score.
$^i$SSQSS: Social Support Questionnaire satisfaction score.

### Association Among Baseline Circadian Rhythm, Mental Health, and Social Support/Connectedness

There were no significant interactions between circadian rhythm and diagnostic group in predicting mental health or social support/connectedness; therefore, data were pooled across diagnostic groups. At baseline, higher circadian rhythm was associated with a larger social support network ($\beta=0.088, 95\% \text{ CI } 0.009-0.168; t_{94}=2.20; P=.03$; Figure 1). Circadian rhythm was not significantly associated with any other measure of mental health or social support/connectedness.
Figure 1. Association between circadian rhythm and the size of social support networks (Social Support Questionnaire number score [SSQNS]). Shading represents 95% CIs.

Changes in Mental Health and Social Support/Connectedness Over Time

There was no interaction between diagnosis and time point in predicting PHQ-9 scores ($F_{5,359}=0.94; P=0.46$); therefore, diagnostic groups were pooled together for analysis. There was a significant main effect of time point in predicting PHQ-9 scores ($F_{5,363}=4.92; P<0.001$). Post hoc tests found that PHQ-9 scores decreased from baseline to time point 2 ($t_{367}=2.58; q=0.03$; Cohen $d=0.35$, 95% CI 0.08-0.62), time point 3 ($t_{370}=3.05; q=0.01$; Cohen $d=0.44$, 95% CI 0.15-0.72), time point 4 ($t_{371}=4.09; q<0.001$; Cohen $d=0.62$, 95% CI 0.32-0.92), time point 5 ($t_{370}=3.77; q=0.002$; Cohen $d=0.61$, 95% CI 0.29-0.94), and time point 6 ($t_{368}=2.95; q=0.01$; Cohen $d=0.73$, 95% CI 0.24-1.23; Figure 2A). There were no other significant differences among the time points. There were no other significant interactions between diagnosis and time points or main effects for the time points for other measures of mental health or social support/connectedness.

There was a significant interaction between time point and diagnosis in predicting mania on the ASRM ($F_{5,360}=2.96; P=0.01$). Post hoc tests found that ASRM scores in people with BD decreased from baseline to time point 4 ($t_{375}=4.06; q=0.004$; Cohen $d=1.06$, 95% CI 0.55-1.59; Figure 2B). Compared with baseline ASRM scores in people with BD, ASRM scores in people with MDD were also lower at time point 2 ($t_{293}=3.60; q=0.01$; Cohen $d=0.91$, 95% CI 0.41-1.42) and time point 3 ($t_{305}=3.63; q=0.01$; Cohen $d=0.95$, 95% CI 0.43-1.47).
Circadian Rhythm Predicting Change in Mental Health or Social Support/Connectedness Over Time

A significant interaction was found among time point, baseline circadian rhythm, and diagnoses ($F_{5.275}=3.65; P=.003$) in predicting GAD-7 scores. To further explore the 3-way interaction, MDD and BD were analyzed separately. A significant interaction between time point and circadian rhythm was found for people with BD only ($F_{5.91}=3.25; P=.009$; Figure 3A [47]). Post hoc tests found that anxiety in people with circadian rhythm scores approximately 1 SD below the mean (ie, –8.73) decreased between time points 1 and 5 ($t_{94.6}=3.0; P=.003; q=.02; \text{Cohen } d=1.34, 95\% \text{ CI } 0.44\text{ to }2.25$). Conversely, anxiety in people with circadian rhythm scores 1 SD above the mean (ie, –2.20) significantly increased between time points 1 and 5; yet, this did not survive FDR correction ($t_{94.6}=–2.32; P=.02; q=.11; \text{Cohen } d=–1.05, 95\% \text{ CI } –1.98\text{ to }–0.13$). No interaction was found between time points and circadian rhythm for people with MDD (Figure 3B [47]). No significant interactions were found among time point, circadian rhythm, and diagnosis category or time point and circadian rhythm for any other measures of mental health (Multimedia Appendix 5) or social support/connectedness.

Figure 2. (A) Depressive symptoms (Patient Health Questionnaire-9 [PHQ-9]) and (B) mania symptoms (Altman Self-Rating Mania Scale [ASRM]) at each study time point. Error bars represent 95% CIs.
**Discussion**

**Principal Findings**

This study investigated the association among circadian rhythm, mental health symptoms, and social support/connectedness in adults with MDD or BD. Although we found no significant association between circadian rhythm and mental health scores, we found a positive association between circadian rhythm and social support at baseline. Both clinical groups showed a decrease in depressive symptoms over time; yet, only those with BD showed a decrease in mania symptoms. Finally, we found that circadian rhythm at baseline moderated the change in anxiety symptoms in people with BD only, where higher baseline circadian rhythm scores were associated with increased anxiety and lower baseline circadian rhythm scores were associated with decreased anxiety at follow-up.

Circadian rhythm was not associated with mental health symptoms; yet, it was positively associated with the size of participants’ social support networks. This association aligns with the study by Margraf et al [13], who found that greater social rhythm regularity was associated with better life satisfaction in healthy individuals. Greater circadian rhythm reflects individuals’ increased movement around various locations, suggesting that they may be more socially interactive. Social interaction also plays an important role in well-being [48], which could lead to improved mental health. Although

---

**Figure 3.** Baseline circadian rhythm moderating change in anxiety (Generalized Anxiety Disorder Scale [GAD-7]) at each study time point in (A) bipolar disorder and (B) depression. Following the convention suggested by Aiken and West [47], we used the mean value of the moderator (ie, circadian rhythm) as well as 1 SD below and above the mean value of the moderator to plot the effect of circadian rhythm on change in anxiety across time points. Error bars represent 95% CIs.
previous studies have established relationships between circadian rhythm and other measures of mental health functioning, the samples are generally made up of only participants with BD [12], healthy controls and participants with MDD [14], healthy individuals, or healthy adolescents [49]. Moreover, although Saeb et al [23] found a strong negative association (r = -0.63) between circadian rhythm and PHQ-9 scores, their limited analyzable sample size (n=28) may have underestimated the association. In this study, we found a nonsignificant association using a larger sample size (n=121), with the coefficient reported by Saeb et al [23] not found within our 95% CIs (r = -0.12, 95% CI -0.32 to 0.08). This finding is also in line with the follow-up study by Saeb et al [22], who found a nonsignificant association between circadian movement and PHQ-9 scores at baseline and a weaker correlation coefficient (r = -0.34, 95% CI -0.339 to 0.341). Taken together, our findings suggest that only social support may be associated with circadian rhythm at baseline, whereas little evidence is found for the association between circadian rhythm and mental health functioning.

Although both clinical groups showed a decrease in depressive symptoms over time, only those with BD showed a decrease in mania symptoms. These findings are in line with previous mental health studies that report fluctuations in mental health symptoms over time [28-30]. Moreover, given that BD is uniquely characterized by manic features [31,50], variation in these symptoms would be expected among people with BD. However, given the overall trajectories of symptom profiles, the study duration may not have been adequate to accurately capture the true symptom variability. This pattern of symptoms implies that future studies would benefit from longer durations and more incentives to increase study completion.

Circadian rhythm moderated the change in anxiety symptoms in people with BD only, where higher baseline circadian rhythm was associated with increased anxiety and lower baseline circadian rhythm was associated with decreased anxiety between baseline and time point 5. People with BD had significantly lower baseline anxiety symptoms and likely cases of anxiety than people with MDD, suggesting differing anxiety profiles between the diagnostic groups. Furthermore, increased anxiety in those with higher baseline circadian rhythm may reflect an initial decline in functioning that corresponds with poor mental health. Similarly, if circadian rhythm is lower during a period of poor mental health, it may then predict improvements in anxiety at future time points as one recovers. However, these effects seem to manifest dynamically in the symptom cycle because the effect disappears by the final time point. Together, these findings suggest that circadian rhythm may be a marker for anxiety symptom trajectories in people with BD that can be monitored for targeted treatment and intervention.

This study includes several limitations. Although the recruitment of clinically diagnosed patients is the strength of the study, the lack of healthy controls limits generalizability from clinical to nonclinical samples. Despite being recruited through a clinical setting, our participants reported low levels of symptoms, suggesting that they may be in a relatively stable symptomatic period. As such, future research should explore participants across different stages of the course of illness in longitudinal studies. Although GPS data offers an inexpensive and accessible data source for the tracking of circadian rhythm, the direct relationship to biological measures of circadian rhythm remains unclear and requires further investigation. Problems with data uploads were generally because of lost internet connection and disabled mobile phone settings (ie, GPS, internet, notifications, and data), which limited the analysis of GPS data to baseline only. Moreover, although enough participants expressed interest in terms of our target sample size, there was a considerable amount of participant dropout and missing data. Specifically, the number of participants with BD (n=42) was lower than those with MDD (n=76) and did not meet the target sample size (n=76). Future research may benefit from developing study apps that send push notifications to remind participants to enable certain settings and complete the required questionnaires.

Conclusions

In conclusion, our results suggest that circadian rhythm extracted from smartphone GPS data was associated with social support and change in anxiety in a clinical sample of adults with mood disorders. However, little evidence was found for the association between circadian rhythm and mental health functioning. Larger future studies examining a wider variety of biomarkers and sensor features for a longer duration of time are warranted. However, with the ubiquity of smartphones, there is an encouraging potential to shift the nature of identifying mood disorders. Contextual features translated directly from smartphone data may potentially detect mood disorders earlier and more easily.

Acknowledgments

The authors wish to thank all patients who participated in the study and the Black Dog Institute information technology team for providing the technical support to deliver the study.

Authors’ Contributions

TWB, MEL, HC, QJJW, and BO conceived and conducted the study. TWB analyzed the sensor data, and BO prepared the data set for analysis alongside MTZ. TAB and MTZ, with supervision from BO, analyzed the data. All authors interpreted the data, and TAB prepared the figures and tables. All authors contributed to the authorship of the manuscript.

Conflicts of Interest

None declared.
Recruitment flow of the study.

[DOCX File, 39 KB - mental_v9i5e35549_app1.docx]

Schedule and completion of the study questionnaires.

[DOCX File, 14 KB - mental_v9i5e35549_app2.docx]

Circadian rhythm.

[DOCX File, 24 KB - mental_v9i5e35549_app3.docx]

Supplementary analyses.

[DOCX File, 21 KB - mental_v9i5e35549_app4.docx]

Nonsignificant moderating effect of baseline circadian rhythm change in (A) depression severity (Patient Health Questionnaire-9) and (B) mania (Altman Self-Rating Mania Scale) across time points. Error bars represent 95% CIs.

[DOCX File, 48 KB - mental_v9i5e35549_app5.docx]

References


34. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009060.x] [Medline: 11556941]


46. emmeans: estimated marginal means, aka least-squares means. R project. URL: https://cran.r-project.org/web/packages/emmeans/index.html [accessed 2022-04-23]


Abbreviations
ASRM: Altman Self-Rating Mania Scale
BD: bipolar disorder
CONSORT: Consolidated Standards of Reporting Trials
FDR: false discovery rate
GAD-7: Generalized Anxiety Disorder Scale
MDD: major depressive disorder
PHQ-9: Patient Health Questionnaire-9
SCS-R: Social Connectedness Scale-Revised
SSQ: Social Support Questionnaire
SWLS: Satisfaction With Life Scale

©Taylor A Braund, May The Zin, Tjeerd W Boonstra, Quincy J J Wong, Mark E Larsen, Helen Christensen, Gabriel Tillman, Bridianne O’Dea. Originally published in JMIR Mental Health (https://mental.jmir.org), 04.05.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR...
Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on https://mental.jmir.org/, as well as this copyright and license information must be included.
Momentary Self-regulation: Scale Development and Preliminary Validation

Emily A Scherer, Sunny Jung Kim, Stephen A Metcalf, Mary Ann Sweeney, Jialing Wu, Haiyi Xie, Gina L Mazza, Matthew J Valente, David P MacKinnon, Lisa A Marsch

Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
Department of Health Behavior and Policy, School of Medicine, Virginia Commonwealth University, Richmond, VA, United States
Health Communication and Digital Innovation, Massey Cancer Center, Virginia Commonwealth University, Richmond, VA, United States
Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom
School of Media and Design, Shanghai JiaoTong University, Shanghai, China
Department of Computer Science, Dartmouth College, Hanover, NH, United States
Department of Psychology, Arizona State University, Tempe, AZ, United States
Department of Quantitative Health Sciences, Mayo Clinic, Scottsdale, AZ, United States
Center for Children and Families, Department of Psychology, Florida International University, Miami, FL, United States

* these authors contributed equally

Corresponding Author:
Emily A Scherer, PhD
Center for Technology and Behavioral Health
Geisel School of Medicine at Dartmouth
Suite 315
46 Centerra Pkwy
Lebanon, NH, 03766
United States
Phone: 1 603 646 7000
Email: blood.emily@gmail.com

Abstract

Background: Self-regulation refers to a person’s ability to manage their cognitive, emotional, and behavioral processes to achieve long-term goals. Most prior research has examined self-regulation at the individual level; however, individual-level assessments do not allow the examination of dynamic patterns of intraindividual variability in self-regulation and thus cannot aid in understanding potential malleable processes of self-regulation that may occur in response to the daily environment.

Objective: This study aims to develop a brief, psychometrically sound momentary self-regulation scale that can be practically administered through participants’ mobile devices at a momentary level.

Methods: This study was conducted in 2 phases. In the first phase, in a sample of 522 adults collected as part of a larger self-regulation project, we examined 23 previously validated assessments of self-regulation containing 594 items in total to evaluate the underlying structure of self-regulation via exploratory and confirmatory factor analyses. We then selected 20 trait-level items to be carried forward to the second phase. In the second phase, we converted each item into a momentary question and piloted the momentary items in a sample of 53 adults over 14 days. Using the results from the momentary pilot study, we explored the psychometric properties of the items and assessed their underlying structure. We then proposed a set of subscale and total score calculations.

Results: In the first phase, the selected individual-level items appeared to measure 4 factors of self-regulation. The factors identified were perseverance, sensation seeking, emotion regulation, and mindfulness. In the second phase of the ecological momentary assessment pilot, the selected items demonstrated strong construct validity as well as predictive validity for health risk behaviors.
**Conclusions:** Our findings provide preliminary evidence for a 12-item momentary self-regulation scale comprising 4 subscales designed to capture self-regulatory dynamics at the momentary level.

*(JMIR Ment Health 2022;9(5):e35273) doi:10.2196/35273*

**KEYWORDS**
self-regulation; momentary self-regulation; ecological momentary assessment; psychometric; health behavior change; health risk behaviors; mobile phone

**Introduction**

**Background**
Self-regulation refers to a person’s ability to manage emotions, cognition, and behavior to avoid immediate gratification, which may interfere with achieving long-term goals. People with self-regulatory competence tend to make goal-oriented decisions and inhibit impulsive behavior that is incompatible with their long-term goals [1]. Self-regulation lapses are linked to social and health problems, such as poor academic outcomes, obesity, substance use disorders, and preventable deaths [2-4].

An individual’s level of self-regulation is likely responsive to internal factors, such as negative affect or stress, which may lessen an individual’s determination or available resources to focus on self-regulatory behavior, as well as external environmental factors, such as being in a location with many temptations for risk behaviors. However, published measures of self-regulation are typically based on retrospective self-reports obtained through cross-sectional surveys or task-based methods. As in other assessments of psychological constructs (eg, self-esteem and personality), retrospective self-report methods collected at 1 time point are helpful in measuring dispositional states, allowing researchers to quantitatively describe an individual and examine interindividual variability, but are limited in measuring the changing states of such constructs within an individual [5,6]. Task-based assessments have similarly been developed for single or paired assessments before and after a defined stimulus and are typically delivered in a laboratory setting. Thus, traditional trait-level assessments may fall short of examining dynamic patterns reflecting intraindividual variability in self-regulation and understanding malleable processes of one’s self-regulation in a naturalistic setting (as aligned with the contextual model of self-regulation proposed by Roos and Witkiewitz [7]).

A methodologically and psychometrically sound metric that precisely and sensitively captures malleable processes involved in self-regulation in a real-world setting may enable a more contextually informed understanding of self-regulatory processes. Developing a valid assay for measuring changes in self-regulation in a nonlaboratory, everyday setting may help researchers better identify the construct’s responsiveness to internal and environmental factors and thus more effectively intervene in self-regulation as a putative mechanism that may play a causal role in facilitating health behavior change. Proliferation in information and communication technologies, combined with novel measurement methods, such as smartphone-based ecological momentary assessment (EMA), enables researchers to examine and assess self-regulatory processes and dynamics at a momentary level as people move through their lives in various real-world environmental contexts (ie, in contrast to laboratory settings or retrospective recall).

**Overall Study Objective**

**Overview**
This study aimed to develop a brief, psychometrically sound momentary self-regulation scale that can be practically administered through participants’ mobile devices at a momentary level. The objective was for the momentary scale to capture the constructs measured in existing self-regulation measures and capture both intra- and interindividual variability in self-regulation as it occurs in naturalistic settings. This work is part of a broader exploration of the ontology of self-regulation supported by the National Institutes of Health’s Science of Behavior Change initiative [8]. The work was conducted in 2 distinct phases.

**Phase 1: Measuring Self-regulation at the Individual Level Using Existing Scale Items**
Beginning with a broad representation of items putatively measuring self-regulation at the individual level, we aimed to understand the underlying dimensions measured by these items and then select a smaller subset of items that capture these dimensions well and that could be studied at the moment level in a naturalistic setting.

**Phase 2: Scale Development and Preliminary Validation**
Beginning with the items selected in phase 1, we aimed to modify the items for measurement at the moment level, pilot their use via momentary assessment methods, construct a momentary scale, and preliminarily assess its psychometric properties at the moment level.

**Methods**

**Phase 1**

**Overview**
Starting with a comprehensive set of existing scales that purport to measure self-regulation at the individual level, we aimed to confirm each scale’s factor structure and characterize its item characteristics. We then assessed the underlying constructs measured by the full set of items from all scales and determined their factor structure. Considering this as the underlying structure of self-regulation, we made an initial selection of the best-performing items from each factor. We aimed to confirm that the factor structure of self-regulation was preserved when using this limited set of items. Once we finalized a set of items that performed well and together measured all factors identified...
as part of the constructs measured at the individual level, we moved to study them further at the moment level in phase 2.

**Literature Review and Scale and Task Selection**

The larger self-regulation initiative began with a comprehensive review of the scientific literature of assessments (both survey based and cognitive task based) used in the domain of self-regulation research and related constructs (eg, impulsivity, mindfulness, behavioral disinhibition, and temporal discounting). This review outlined the origin of all assays and the conceptual and empirical associations between the data from each measure and health and social behaviors. This study identified 23 self-report surveys and 37 cognitive tasks that purport to measure some aspect of self-regulation. The process for selecting the scales has been described in detail elsewhere [9,10]. Briefly, these scales were chosen for their ability to measure underlying latent constructs of the umbrella construct of self-regulation. Self-regulation refers to a person’s ability to manage cognitive, motivational, and emotional resources to act in accordance with their long-term goals. The constructs were operationalized as cognitive functions that allow an individual to engage in effective self-regulatory behaviors. Measures that focused on aspects of self-regulation such as goal planning, self-regulation failures, impulsivity, cognitive control, and temporal discounting were sampled.

**Sample and Sample Partitioning**

Next, a sample of 522 adults was recruited through Amazon Mechanical Turk (MTurk), a crowdsourcing website, and the 594 items from the 23 surveys were administered. A subset of these individuals (150/522, 28.7%) was selected to complete the surveys again 3 months later to enable the assessment of test-retest reliability [11]. A description of the MTurk study design and sample recruitment procedures is described elsewhere, as is a description of the participants in the sample and their scores on the surveys and behavioral tasks [8,9]. For this project, survey data from this sample were used to perform a dimension analysis of a range of self-regulation measures. A full list of the surveys and their subscales is included in Multimedia Appendix 1 [6,12-31].

To support the dimension analysis via exploratory factor analysis (EFA), the observations (participants) were partitioned into a discovery set (200/522, 38.3%) and a validation set (322/522, 61.7%). The complete set (N=522) comprised the discovery and validation sets.

**Analytic Approach**

**Item Reduction to Solve an n<p Problem**

In statistics, an n<p problem describes the challenge of having more variables than observations on which the variables are measured. With the MTurk sample of 522, we encountered such a challenge in using all the variables (items) to perform a dimension analysis of self-regulation using all 594 items from the 23 self-regulation surveys. Therefore, a precursor to performing exploratory factor analyses was the reduction of the number of evaluated items while still maintaining items in each potential self-regulation domain. To facilitate this process, scale-level analyses were performed on each self-reported scale to first confirm the structure of the scale and then identify a subset of well-performing items to carry forward as candidates for further self-regulation dimension analysis. This process began with research on the scale’s derivation and a qualitative review of the scale to ensure self-regulation was indeed the scale’s target. Next, for each scale, correlated factor analysis was performed to confirm the scale’s factor structure in the sample. Finally, an item response theory (IRT) analysis was performed within each scale to identify a set of approximately 3 items per scale to inform the development of a measure that succinctly captures various dimensions of self-regulation. Items were considered to perform well in the factor analysis if they loaded primarily on one factor of the scale and did not have high loadings on other factors and were considered to perform well in the IRT analysis if they had high information and discrimination. The goal of the item reduction process was to keep items from each subscale or construct measured to retain full coverage of scales in the final candidate set while limiting them to well-performing items.

**Dimension Analysis and Factor Interpretation**

The MTurk data on the reduced set of items (116 items) were subjected to dimensional analyses. The discovery (n=200), validation (n=322), and complete data sets (N=522) were used to perform the EFA of the reduced set of items. The goal of the EFA was to identify the number of underlying factors in the sample and interpret the content of each factor where possible. EFA was performed using Mplus [32].

For each factor-based solution, a qualitative review of the results was performed to identify and describe the factors. This assessment was done by focusing on items that loaded onto a factor and then qualitatively reviewing the text of these items and naming the factor based on the content of all loading items.

**Test-Retest Results**

Test-retest information from each item was also considered in the item selection process. In the sample of 150 MTurk participants who completed the surveys at 2 time points, we computed item-level intraclass correlation coefficients (ICCs). The results of the scale and task ICCs and how they related to behaviors have been published elsewhere [11]. For this study, the ICCs provided further information on which items might be better for momentary measurement. A large ICC indicates that there is not a great deal of variability within individuals. Such an item likely measures an individual-level characteristic rather than a momentary characteristic that may vary over time and in different situations. Details on how the ICCs were incorporated into the selection of items for study at a momentary level are included in the item selection process described in the following sections.

**Item Selection for Study of Momentary Self-regulation**

Using the results of the dimension analyses, we aimed to select a set of items from each identified factor and select approximately 15 items in total. The goal for the number of items selected for the momentary study was based on the number thought to be reasonable to answer on a momentary basis (consistent with the broader literature on EMA), as well as provide a large enough sample of items so that further item
selection could be performed based on their performance in the planned pilot of the items (phase 2). To obtain a set of items from each factor, we selected items with the following characteristics:

1. Items that loaded primarily on one factor (did not load >0.5 on >1 factor) and loaded highly on that factor (>0.5)
2. Items whose ICC was not large enough to indicate a lack of variability within the individual
3. Items that did not refer to a specific activity that would not be present in a large proportion of moments in a real-world setting (eg, skiing or skydiving)

Confirming the Factor Structure
Once the items were selected for further study at the momentary level, we performed a confirmatory factor analysis (CFA) to determine whether the identified factors in the larger set of items were confirmed in the selected items. Adjustments to the item selection and further exploratory analyses and CFAs were performed until a solution with acceptable fit statistics was obtained.

Phase 2
Overview
Using the items selected in phase 1, we moved to study self-regulation at a momentary level. We piloted all selected items on a new sample of participants and then assessed within- and between-individual variability of items and examined the underlying factor structure at the within- and between-individual levels. The goal was to develop a momentary self-regulation measure and perform an initial evaluation of its validity. To ensure construct validity between the trait-level and momentary-level self-regulation measurements, we examined the association between nonmomentary self-regulation measured at baseline and the momentary self-regulation responses collected from phase 2 participants throughout a total of 42 time points over 14 consecutive days. We also assessed the psychometric properties of between- and within-individual momentary self-regulation items and subscales, as well as the predictive validity of the subscales and total scores for health risk behaviors (eg, smoking and overeating). Self-regulation has been implicated in many health risk behaviors, including substance use and disordered eating [33-37].

Sample
A new sample of participants was recruited through MTurk for phase 2. To be eligible for the momentary study, participants had to be aged between 18 and 50 years, US residents in states that included only the Eastern time zone (because of the manual process involved in implementing the text prompting and compensation procedures), fluent in English, and willing to receive text prompts on their smartphone to initiate and complete 3 surveys per day over 14 consecutive days. We recruited 60 participants to account for potential study attrition and meet a minimum analytic sample size of 50. We did not perform a formal power analysis or sample size calculation, given the exploratory nature of the study and the lack of preexisting data on the variability of momentary self-regulation. However, we expected that the sample size (n=50), along with 42 time points for each participant, would provide a rich data set for performing the proposed psychometric analyses.

Ethics Approval
The study procedure and survey materials were approved by the Dartmouth College Committee for the Protection of Human Subjects (STUDY00028975).

Data Collection
We leveraged the technology features available through mobile texting prompts and the Qualtrics survey platform to simulate EMA methods with 42 repeated microsurveys (3 times per day—morning, midday, and evening—for 14 days) and facilitate rapid compensation. Each microsurvey contained 20 survey items and took <3 minutes on average to complete. Participants were compensated within an hour of the completion of each microsurvey. We simulated the EMA method instead of developing an EMA mobile app to expedite data collection while enabling remote recruitment and data collection. Each text prompt contained a unique hyperlink that directed the participants to a given web-based microsurvey. Text prompts were sent at a random time within a predetermined time window. Participants were asked to complete microsurveys on their mobile devices (verified via an external website that tracks the devices used to answer surveys). To promote study retention, we sent a reminder with the same message to those who did not complete the survey within an hour of receiving the random prompt. Consistent with compensation models offered within the MTurk crowdsourcing environment, participants were compensated US $0.30 for each microassessment and US $5 daily bonuses for completing all 3 assessments per day.

Measures
Baseline Survey
We measured demographic characteristics (eg, age, gender, ethnicity, race, education, and income) and height and weight for BMI, as well as health behavior characteristics, such as the Drug Abuse Screening Test (DAST-10) [38,39], Cannabis Use Disorders Identification Test–Revised (CUDIT-R) [40], Alcohol Use Disorders Identification Test (AUDIT) [41], Three-Factor Eating Questionnaire-R18 (TFEQ-R18) [42], and smoking status. These behavioral characteristics were collected to examine the relationship between momentary self-regulation dynamics and health risk behaviors. In phase 1, a subset of items from the 23 self-report surveys was selected. In the phase 2 baseline survey, we included the 8 self-report surveys from which the momentary self-regulation scale items were selected. These scales were the functional and dysfunctional impulsivity...
subscapes from the Dickman Impulsivity Inventory [12]; the suppression subscale of the Emotion Regulation Questionnaire (ERQ) [13]; the nonjudging subscale of the Five Facet Mindfulness Questionnaire (FFMQ) [14]; the venturesomeness subscale of the Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire (I-7) [15]; the Mindful Attention Awareness Scale (MAAS) [16]; the Selection, Optimization, and Compensation Questionnaire [17]; the Short Self-Regulation Questionnaire (SSRQ) [18]; and the lack of premeditation and lack of perseverance subscales of the Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, and Positive Urgency (UPPS-P) Impulsive Behavior Scale [19,20]. These subscales were collected to facilitate validation analyses for the momentary items.

Microsurveys
Driven by the practice of momentary scale development and validation research in psychometric studies [43], the wording of the 20 items was modified to capture the momentary level of self-regulation. For example, the item “I keep my emotions to myself” was modified to “Since the last prompt, I kept my emotions to myself.” Similar modification methods with the leading phrase “Since the last prompt...” were applied to all other candidate items. Response options for all items were standardized to a 5-point Likert scale, ranging from not at all (1) to extremely (5). After phase 1 and data collection for phase 2, we noted that 2 of the 20 items originated from the Multidimensional Personality Questionnaire Control scale [21], which has a copyright restriction that prevents their use in the development and publishing of new measures. Therefore, we removed these 2 items from the originally selected items, verified that the other 22 self-regulation surveys did not have copyright restrictions, and proceeded with 18 items in the analysis. As detailed in the following sections, the 18 items did not have a different structure than that of the 20 items.

Analytic Approach
First, we evaluated the construct validity of the 20 individual momentary items using baseline self-regulation surveys. To do this, we fit generalized estimating equation models to examine the association between each momentary self-regulation item and its corresponding trait-level self-regulation subscale assessed at baseline. This was to ensure what we intended to measure at a momentary level (self-regulation at the moment of each assessment) was the same concept (self-regulation) captured by the original self-regulation survey.

Second, to assess the intra- and interindividual variability of the items, we examined ICCs estimated via univariate multilevel models with a probit link. We then performed multilevel EFA followed by CFA, allowing for correlated factors to identify the number of factors measured by the momentary set of items at the between- and within-individual levels. The results of these factor analyses, along with the information on construct validity and ICCs, were used to select approximately 3 items per factor to be included in the final momentary scale.

With the final set of momentary self-regulation items, we created subscale scores comprising the mean item response from all items from a factor and created overall scores computed as the mean of the 4 subscale scores. We evaluated the construct validity of the final momentary subscales and total scores via mixed-effects models examining the relationship between the momentary subscale and total score and the baseline self-regulation measures. Finally, we evaluated the predictive validity using mixed-effects models for health information (eg, alcohol, smoking, other substance use, food intake, and BMI) and explored the association between momentary self-regulation subscales and age, sex, education, and income. All analyses accommodated the multilevel structure of the data by modeling both the between- and within-individual variations in repeated assessments. Mplus [32] was used for the multilevel factor analyses. SAS software (version 9.4, SAS Institute Inc) was used for data merging and processing, generalized estimating equations, and mixed models.

Results
Phase 1
Item Reduction
Owing to space constraints, we do not present the results from each scale-level analysis used to select a reduced set of well-performing items. Instead, we briefly describe the steps taken for the UPPS-P Impulsive Behavior Scale as an example. The same process was followed for all 23 self-regulation surveys.

First, the 5-factor structure was confirmed through a factor analysis of all the items on the scale. The factors loaded onto the subscales that were previously defined in the literature. The selected items loaded strongly onto their designated factors and showed a minimal overlap. Separate scree plots for each subscale confirmed the 1-factor structure of the subscales.

The selected items from each subscale are described in Textbox 1. These items were selected based on a qualitative assessment to identify the items with the best discrimination and highest level of information.
Selected items from each subscale

**Premeditation**
- “I like to stop and think things over before I do them.” [UPP17]
- “I usually think carefully before doing anything.” [UPP49]
- “Before making up my mind, I consider all the advantages and disadvantages.” [UPP56]

**Perseverance**
- “I generally like to see things through to the end.” [UPP05]
- “I finish what I start.” [UPP28]
- “I almost always finish projects that I start.” [UPP43]

**Negative urgency**
- “When I feel bad, I will often do things I later regret in order to make myself feel better now.” [UPP18]
- “When I am upset I often act without thinking.” [UPP30]
- “I often make matters worse because I act without thinking when I am upset.” [UPP45]

**Positive urgency**
- “When I get really happy about something, I tend to do things that can have bad consequences.” [UPP41]
- “When overjoyed, I feel like I can’t stop myself from going overboard.” [UPP46]
- “I tend to act without thinking when I am really excited.” [UPP53]

**Sensation seeking**
- “I quite enjoy taking risks.” [UPP24]
- “I welcome new and exciting experiences and sensations, even if they are a little frightening and unconventional.” [UPP32]
- “I sometimes like doing things that are a bit frightening.” [UPP42]

**Dimension Analysis and Factor Interpretation**

**Number of Factors: EFA**
Scree plots of eigenvalues from the discovery, validation, and complete sets (reduced items) showed a decreased rate of change after 3 eigenvalues and were almost flat after 6 eigenvalues. Given this, 3- and 4-factor EFAs were completed.

**Interpretation of Factors**
Patterns of factor loadings were used to interpret the measured factors across several factor solutions. For each solution, all items loading on a factor were reviewed, and an attempt was made to identify the construct under study and assign a name to each factor through a qualitative assessment of the item text. Table 1 lists the names of the factors across multiple factor-based solutions.

In the 3-factor solution, in both the discovery and validation sets, as well as the complete observation set, the factors appeared to represent (1) perseverance or lack of impulsivity, (2) sensation seeking, and (3) inhibition or mind over matter. In the 4-factor solution, in both the discovery and validation sets, the factors appeared to represent (1) perseverance, (2) impulsivity or sensation seeking, (3) inhibition or mind over matter, and (4) negative rumination or self-judgment. However, in the complete observation set, the factors differed, appearing to represent (1) impulsivity, (2) sensation seeking, (3) inhibition or mind over matter, and (4) negative rumination or self-judgment. Overall, the 3-factor solution had the most support based on the scree plot, consistency of results across observation sets, and additional exploratory factor analyses not presented. Therefore, we moved to the next step, assuming a 3-factor solution.
Table 1. Phase 1: factor interpretation for the 3- and 4-factor exploratory factor analysis.

<table>
<thead>
<tr>
<th>Factor analysis and data set</th>
<th>3-factor</th>
<th>4-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete</td>
<td>Validation</td>
</tr>
<tr>
<td>Perseverance or lack of impulsivity</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Perseverance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sensation seeking</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Impulsivity or sensation seeking</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Inhibition or mind over matter</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Negative rumination or self-judgment</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Item Selection for Study of Momentary Self-regulation**

As 3 factors appeared to have the most support, and among the 3-factor solutions, the discovery, validation, and complete data sets yielded similar factors and results, we considered the complete set and 3-factor solution as the results from which we would identify items for further study at the momentary level. The selection process is detailed in Textbox 2.

Textbox 2. Item selection process.

<table>
<thead>
<tr>
<th>Item selection process based on the complete data set and 3-factor solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. From each factor, we removed items with factor loadings &lt;0.5 or with factor loadings &gt;0.5 on &gt;1 factor from further consideration.</td>
</tr>
<tr>
<td>2. Among the remaining items, the perseverance factor had many more items with factor loadings &gt;0.5 (49 items), whereas the other 2 factors (sensation seeking and inhibition or mind over matter) each had a smaller number of items with loadings &gt;0.5 (15 and 9 items, respectively). As we wanted to represent all 3 factors in our selected items, we only considered the first 21 items (approximately 40% of all items with loadings &gt;0.5) that loaded onto the perseverance factor (ordered by factor loading value). For sensation seeking and inhibition or mind over matter, we evaluated all items with loadings &gt;0.5 onto the factor.</td>
</tr>
<tr>
<td>3. Among the resultant 21+15+9 items, we aimed to select items in proportion to the number loading on the 3 factors; hence, we sought to select 10 perseverance or impulsivity items, 3 sensation-seeking items, and 2 inhibition or mind over matter items. The selection procedure was as follows:</td>
</tr>
<tr>
<td>• We removed items with a large intraclass correlation coefficient (ICC) value (&gt;0.7).</td>
</tr>
<tr>
<td>• We removed items referring to a specific activity (eg, skiing or skydiving), which is common among sensation-seeking items.</td>
</tr>
<tr>
<td>• We added items outside of the initial selection from each factor if they narrowly missed the selection but had available ICCs that were &lt;0.4.</td>
</tr>
<tr>
<td>• We then selected items that represented a variety of themes within the factor, favoring items with lower ICCs.</td>
</tr>
</tbody>
</table>

The selection resulted in 20 initially selected items: 10 items from the perseverance or impulsivity factor, 5 items from the sensation-seeking factor, and 5 items from the inhibition or mind over matter factor. Given that the items that appeared to represent impulsivity (eg, “I think before doing;” “do you generally do and say things without stopping to think?” and “I get in trouble because I don’t think before I act”) were not selected, we call the first factor perseverance rather than perseverence or lack of impulsivity.

**Confirming the Factor Structure**

We performed a CFA to determine whether the identified factors in the larger set of items were confirmed in the selected 20 items. The 3-factor confirmatory model did not fit the data well (root mean square error of approximation [RMSEA] 0.115, 95% CI 0.110-0.121; Tucker-Lewis Index [TLI] 0.801). Therefore, we performed a 3-, 4-, and 5-factor EFA of the selected items. In the exploratory analysis, the 4-factor solution best fit the data (RMSEA 0.051, 95% CI 0.043-0.059; TLI 0.961). Qualitative examination of items that had previously made up the third factor (inhibition or mind over matter) showed a split across the third and fourth factors in the EFA, suggesting 2 separate factors: emotion regulation and mindfulness. To accommodate this new structure, from the originally selected 20 items, 2 items were removed from the perseverance factor, and 1 item each was selected (based on the qualitative assessment of measurement and examination of factor loadings within the larger item set and item-level ICCs) to measure the emotion regulation and mindfulness factors. Furthermore, a 4-factor CFA was performed on the revised 20-item set, and the model fit the data well (RMSEA 0.094, 95% CI 0.088-0.100; TLI 0.917), and we considered our selection complete. Tables 2 and 3 show the items initially selected and the final item selection, respectively, that resulted from the confirmatory analyses. The factors identified were perseverance, sensation seeking, emotion regulation, and mindfulness.
Table 2. Phase 1: initial (before confirmatory analysis) selection of items that represent 3 underlying factors to be considered candidates for momentary measurement in phase 2.

<table>
<thead>
<tr>
<th>Item source</th>
<th>Item text</th>
<th>Factor name</th>
<th>Sensation seeking</th>
<th>Inhibition or mind over matter</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPPS-P(^a) Impulsive Behavior Scale [19,20]</td>
<td>• I finish what I start.</td>
<td>Perseverance</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Selection, Optimization, and Compensation Questionnaire [17]</td>
<td>• I keep working on what I have planned until I succeed. • When I do not succeed right away at what I want to do I do not try other possibilities for very long.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>• I generally like to see things through to the end.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Short Self-Regulation Survey [18]</td>
<td>• I set goals for myself and keep track of my progress.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Dickman Impulsivity Inventory [12]</td>
<td>• I often say and do things without considering the consequences.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>• I usually think carefully before doing anything.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Multidimensional Personality Questionnaire [21]</td>
<td>• I am careful in reasoning.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>10-Item Personality Questionnaire [6]</td>
<td>• Dependable, self-disciplined.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Multidimensional Personality Questionnaire [21]</td>
<td>• I value a rational approach.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Short Self-Regulation Survey [18]</td>
<td>• I am able to resist temptation.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>• Do you quite enjoy taking risks?</td>
<td>Positive Urgency</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>• Do you sometimes like doing things that are a bit frightening?</td>
<td>Positive Urgency</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Dickman Impulsivity Inventory [12]</td>
<td>• I am good at taking advantage of unexpected opportunities, where you have to do something immediately or lose your chance.</td>
<td>Positive Urgency</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>• I have a reserved and cautious attitude toward life.</td>
<td>Positive Urgency</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Dickman Impulsivity Inventory [12]</td>
<td>• I like to take part in really fast-paced conversations, where you don’t have much time to think before you speak.</td>
<td>Positive Urgency</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Five Facet Mindfulness Questionnaire [14]</td>
<td>• I think some of my emotions are bad or inappropriate and I shouldn’t feel them.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Emotion Regulation Questionnaire [13]</td>
<td>• When I am feeling negative emotions, I make sure not to express them.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Emotion Regulation Questionnaire [13]</td>
<td>• I control my emotions by not expressing them.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mindful Attention Awareness Scale [16]</td>
<td>• I find myself doing things without paying attention.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mindful Attention Awareness Scale [16]</td>
<td>• It seems I am “running on automatic” without much awareness of what I’m doing.</td>
<td>Perseverance</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

\(^a\)UPPS-P: Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, and Positive Urgency.
Table 3. Phase 1: final (after confirmatory analysis) selection of items that will be considered candidates for momentary measurement in phase 2.

<table>
<thead>
<tr>
<th>Factor name</th>
<th>Item text</th>
<th>Perseverance</th>
<th>Sensation seeking</th>
<th>Emotion regulation</th>
<th>Mindfulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection, Optimization, and Compensation Questionnaire [17]</td>
<td>I keep working on what I have planned until I succeed.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>When I do not succeed right away at what I want to do I do not try other possibilities for very long.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>I generally like to see things through to the end.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Self-Regulation Survey [18]</td>
<td>I set goals for myself and keep track of my progress.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dickman Impulsivity Inventory [12]</td>
<td>I often say and do things without considering the consequences.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>I usually think carefully before doing anything.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multidimensional Personality Questionnaire [21]</td>
<td>I am careful in reasoning.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multidimensional Personality Questionnaire [21]</td>
<td>I value a rational approach.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Self-Regulation Survey [18]</td>
<td>I am able to resist temptation.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>Do you quite enjoy taking risks?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>Do you sometimes like doing things that are a bit frightening?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dickman Impulsivity Inventory [12]</td>
<td>I am good at taking advantage of unexpected opportunities, where you have to do something immediately or lose your chance.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>I have a reserved and cautious attitude toward life.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five Facet Mindfulness Questionnaire [14]</td>
<td>I think some of my emotions are bad or inappropriate and I shouldn’t feel them.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion Regulation Questionnaire [13]</td>
<td>I control my emotions by not expressing them.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five Facet Mindfulness Questionnaire [14]</td>
<td>I tell myself I shouldn’t be thinking the way I am thinking.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindful Attention Awareness Scale [16]</td>
<td>I find myself doing things without paying attention.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindful Attention Awareness Scale [16]</td>
<td>It seems I am “running on automatic” without much awareness of what I’m doing.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Phase 2

Participants and Study Retention

Approximately 12% (7/60) of participants who enrolled in the study did not complete any of the 42 microsurveys, resulting in 53 participants with momentary data. Participants were mainly White (47/53, 88%), non-Hispanic (48/53, 90%), married (20/53, 37%), and female (30/53, 56%) and aged between 22 and 48 years at baseline. These 53 participants completed a median of 42 of the 42 microsurveys (range 1-42). The mean BMI was 27.18 (SD 6.08, range 18.30-44.30) kg/m², falling in the overweight category. Approximately 30% (16/53) of the participants reported that they smoked cigarettes every day, 40% (21/53) had used cannabis, and 36% (19/53) reported that they had not limited their alcohol intake to less than monthly in the past year. Approximately 28% (15/53) of participants reported that they had used drugs other than those required for medical reasons.

Momentary Item Examinations: Construct Validity and Intra- and Interindividual Variability

Analyses examining the construct validity of the individual items assessed the association between the baseline trait-level self-regulation item and momentary-level self-regulation item. They showed a significant association with the exception of 1 item (item 3), indicating that for all other items, the concept operationalized and measured through momentary items reflects the original construct (trait-level self-regulation) measured at the baseline survey (Table 4).

Unconditional multilevel models confirmed that there was significant variation at the individual item level for all items with ICCs ranging from 37% to 64%, indicating that a large proportion of the variability in the items was because of within-individual fluctuation rather than between-individual variation.
### Table 4. Phase 2 construct validity of momentary items: regression coefficients from multilevel models relating momentary self-regulation items to corresponding baseline trait-level self-regulation subscales.

<table>
<thead>
<tr>
<th>Item number</th>
<th>Source (baseline self-regulation scale)</th>
<th>Momentary self-regulation item</th>
<th>Regression coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dickman Impulsivity Inventory [12]</td>
<td>Since the last prompt, I’ve been good at taking advantage of unexpected opportunities.</td>
<td>0.65^a</td>
</tr>
<tr>
<td>2</td>
<td>Dickman Impulsivity Inventory [12]</td>
<td>Since the last prompt, I’ve said things without considering the consequences.</td>
<td>0.56^a</td>
</tr>
<tr>
<td>3</td>
<td>Emotion Regulation Questionnaire [13]</td>
<td>Since the last prompt, I’ve controlled my emotions by not expressing them.</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>Emotion Regulation Questionnaire [13]</td>
<td>Since the last prompt, if I’ve felt negative emotions, I’ve made sure not to express them.</td>
<td>0.19^a</td>
</tr>
<tr>
<td>5</td>
<td>Five Facet Mindfulness Questionnaire [14]</td>
<td>Since the last prompt, I’ve told myself I shouldn’t be thinking the way I am thinking.</td>
<td>0.28^b</td>
</tr>
<tr>
<td>6</td>
<td>Five Facet Mindfulness Questionnaire [14]</td>
<td>Since the last prompt, I’ve thought some of my emotions are bad or inappropriate and I shouldn’t feel them.</td>
<td>0.21^b</td>
</tr>
<tr>
<td>7</td>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>Since the last prompt, I’ve quite enjoyed taking risks.</td>
<td>0.50^a</td>
</tr>
<tr>
<td>8</td>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>Since the last prompt, I’ve enjoyed doing things that are a bit frightening.</td>
<td>-0.44^a</td>
</tr>
<tr>
<td>9</td>
<td>Mindful Attention Awareness Scale [16]</td>
<td>Since the last prompt, it has seemed I am running on “automatic” without much awareness of what I’m doing.</td>
<td>-0.42^a</td>
</tr>
<tr>
<td>10</td>
<td>Mindful Attention Awareness Scale [16]</td>
<td>Since the last prompt, I’ve rushed through activities without really being attentive to them.</td>
<td>-0.43^a</td>
</tr>
<tr>
<td>11</td>
<td>Mindful Attention Awareness Scale [16]</td>
<td>Since the last prompt, I’ve found myself doing things without paying attention.</td>
<td>-1.05^a</td>
</tr>
<tr>
<td>14</td>
<td>Selection, Optimization, and Compensation Questionnaire [17]</td>
<td>Since the last prompt, I’ve worked on what I planned until I succeeded.</td>
<td>0.49^a</td>
</tr>
<tr>
<td>15</td>
<td>Short Self-Regulation Questionnaire [18]</td>
<td>Since the last prompt, I have set goals and kept track of my progress toward goals.</td>
<td>0.29^a</td>
</tr>
<tr>
<td>16</td>
<td>Short Self-Regulation Questionnaire [18]</td>
<td>Since the last prompt, I’ve been able to resist temptation.</td>
<td>-0.47^a</td>
</tr>
<tr>
<td>17</td>
<td>UPPS-P^c Impulsive Behavior Scale [19,20]</td>
<td>Since the last prompt, my attitude toward life has been reserved and cautious.</td>
<td>-0.68^a</td>
</tr>
<tr>
<td>18</td>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>Since the last prompt, I’ve generally seen things through to the end.</td>
<td>-0.43^a</td>
</tr>
<tr>
<td>19</td>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>Since the last prompt, I’ve been able to finish projects I started.</td>
<td>-0.61^a</td>
</tr>
<tr>
<td>20</td>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>Since the last prompt, I have thought carefully before doing things.</td>
<td></td>
</tr>
</tbody>
</table>

^a p ≤ .01.

^b p ≤ .05.

^c UPPS-P: Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, and Positive Urgency.

### Examining Factor Structure in Momentary Items

Multilevel EFA models with between 3 and 5 within-individual factors and between 3 and 4 between-individual factors were explored. Table 5 presents the fit statistics for these models. Models with both 4 and 5 within-individual factors and with both 3 and 4 between-individual factors had good fit statistics, with the exception of the chi-square test, which is sensitive to sample size.

We used iterative processes to report the best factor solutions for the momentary self-regulation scale. For example, although the 5-factor within-individual structure seemed to fit the data well, having 5 factors resulted in separating the emotion regulation factor into 2 factors based only on the scales from which the items originated, which is unlikely to be a meaningful distinction. The 3-factor solution at the between-individual level appeared to fit the data well. The difference between the 3- and 4-factor solutions is that emotion regulation and mindfulness factors may be combined into 1 factor at the between-individual level. Given these results, we examined the 4-factor EFA results for the item selection steps that followed. The 4 factors identified were very similar to those present in phase 1: perseverance, sensation seeking, mindfulness, and emotion regulation.

As we aimed to retain only items that performed well at the individual and moment levels in a momentary self-regulation scale, we examined the factor loadings and communalities of...
the items in the multilevel EFA. We selected items that consistently showed high communalities at both within- and between-individual levels, indicating that variability in these items is reasonably explained by the underlying factors, as well as items that consistently loaded highly on the factors of interest and did not simultaneously load highly on the other factors being measured. This process yielded a set of 12 items—3 selected for each underlying factor.

A multilevel CFA was fit to these 12 items. The fit statistics from this multilevel CFA model were comparative fit index 0.989, standardized root mean squared residual (between) 0.083, standardized root mean square residual (within) 0.083, RMSEA 0.012, and TLI 0.985. The significant path coefficients and covariances between the factors are shown in Figure 1.

**Table 5.** Phase 2 fit statistics from multilevel exploratory factor analyses on 18 momentary items.

<table>
<thead>
<tr>
<th>Exploratory factor analysis model</th>
<th>Statistics</th>
<th>3W3B&lt;sup&gt;a&lt;/sup&gt;</th>
<th>4W3B&lt;sup&gt;b&lt;/sup&gt;</th>
<th>5W3B&lt;sup&gt;c&lt;/sup&gt;</th>
<th>4W4B&lt;sup&gt;d&lt;/sup&gt;</th>
<th>5W4B&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of parameters</td>
<td></td>
<td>192</td>
<td>207</td>
<td>221</td>
<td>222</td>
<td>236</td>
</tr>
<tr>
<td>Chi-square (&lt;i df&lt;/i&gt;)&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td>606.7 (204)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>322.6 (189)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>222.7 (175)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>317.6 (174)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>213.1 (160)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>RMSEA&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td>0.030</td>
<td>0.018</td>
<td>0.011</td>
<td>0.020</td>
<td>0.012</td>
</tr>
<tr>
<td>CFI&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.913</td>
<td>0.971</td>
<td>0.990</td>
<td>0.969</td>
<td>0.988</td>
</tr>
<tr>
<td>TLI&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td>0.869</td>
<td>0.953</td>
<td>0.982</td>
<td>0.945</td>
<td>0.978</td>
</tr>
<tr>
<td>SRMR&lt;sup&gt;j&lt;/sup&gt; within</td>
<td></td>
<td>0.130</td>
<td>0.091</td>
<td>0.065</td>
<td>0.091</td>
<td>0.065</td>
</tr>
<tr>
<td>SRMR between</td>
<td></td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0.042</td>
<td>0.042</td>
</tr>
</tbody>
</table>

<sup>a</sup>Model with 3 within-level factors and 3 between-level factors.

<sup>b</sup>Model with 4 within-level factors and 3 between-level factors.

<sup>c</sup>Model with 5 within-level factors and 3 between-level factors.

<sup>d</sup>Model with 4 within-level factors and 4 between-level factors.

<sup>e</sup>Model with 5 within-level factors and 4 between-level factors.

<sup>f</sup><i>P</i>&lt;.01.

<sup>g</sup>RMSEA: root mean square error of approximation.

<sup>b</sup>CFI: comparative fit index.

<sup>i</sup>TLI: Tucker-Lewis Index.

<sup>j</sup>SISRMR: standardized root mean square residual.
Subscale and Total Score Creation

To create subscale scores based on the 4 identified factors, we calculated the mean of the scores of all items loading onto a factor and reverse coded the items so that all items measured the construct in the same direction. For example, the sensation-seeking subscale comprised the mean of items 7 and 8, which measure sensation-seeking behavior, and reverse-coded item 17, which measures a lack of sensation-seeking behavior. In addition, we reverse coded all mindfulness items (items 9, 10, and 11) to measure mindfulness rather than lack of mindfulness.

We also created a total momentary self-regulation scale score that was a composite (mean value) of all subscale scores using 2 methods: literature-based total score and data-based total score. On the basis of the existing literature, we would expect that perseverance, mindfulness, emotion regulation, and lack of sensation-seeking factors would all have positive correlations. Therefore, we created a literature-based total by calculating the mean of the perseverance, mindfulness, emotion regulation, and reverse-coded sensation-seeking subscales (literature-based total).

In the CFA (and EFAs), we found that the emotion regulation and perseverance factors had a negative correlation. Therefore, we examined the items that make up the emotion regulation factor, and instead, they appeared to measure self-judgment to some extent. Indeed, the items were from the nonjudging scale of the FFMQ. Self-judgment could be negative rather than positive in terms of self-regulation; thus, we created another total score comprising the mean of the perseverance, mindfulness, reverse-coded emotion regulation, and reverse-coded sensation-seeking subscales (data-based total). Higher literature-based and data-based totals each indicate greater momentary self-regulation. Table 6 includes the final item set and details on how each item contributes to the subscale and how the subscales contribute to the total score.
Table 6. Phase 2 momentary-level items in momentary self-regulation scale.

<table>
<thead>
<tr>
<th>Item number</th>
<th>Source</th>
<th>Momentary self-regulation item</th>
<th>Momentary self-regulation subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Selection, Optimization, and Compensation Questionnaire [17]</td>
<td>Since the last prompt, I’ve worked on what I planned until I succeeded.</td>
<td>Perseverance</td>
</tr>
<tr>
<td>15</td>
<td>Short Self-Regulation Questionnaire [18]</td>
<td>Since the last prompt, I have set goals and kept track of my progress toward goals.</td>
<td>Perseverance</td>
</tr>
<tr>
<td>19</td>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>Since the last prompt, I’ve been able to finish projects I started.</td>
<td>Perseverance</td>
</tr>
<tr>
<td>7</td>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>Since the last prompt, I’ve quite enjoyed taking risks.</td>
<td>(-) Sensation seeking</td>
</tr>
<tr>
<td>8</td>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire [15]</td>
<td>Since the last prompt, I’ve enjoyed doing things that are a bit frightening.</td>
<td>(-) Sensation seeking</td>
</tr>
<tr>
<td>17</td>
<td>UPPS-P Impulsive Behavior Scale [19,20]</td>
<td>Since the last prompt, my attitude towards life has been reserved and cautious.</td>
<td>(-) Sensation seeking (R)</td>
</tr>
<tr>
<td>4</td>
<td>Emotion Regulation Questionnaire [13]</td>
<td>Since the last prompt, if I’ve felt negative emotions, I’ve made sure not to express them.</td>
<td>(-) Self-judgment</td>
</tr>
<tr>
<td>5</td>
<td>Five Facet Mindfulness Questionnaire [14]</td>
<td>Since the last prompt, I’ve told myself I shouldn’t be thinking the way I am thinking.</td>
<td>(-) Self-judgment</td>
</tr>
<tr>
<td>6</td>
<td>Five Facet Mindfulness Questionnaire [14]</td>
<td>Since the last prompt, I’ve thought some of my emotions are bad or inappropriate and I shouldn’t feel them.</td>
<td>(-) Self-judgment</td>
</tr>
<tr>
<td>9</td>
<td>Mindful Attention Awareness Scale [16]</td>
<td>Since the last prompt, it has seemed I am running on “automatic” without much awareness of what I’m doing.</td>
<td>Mindfulness (R)</td>
</tr>
<tr>
<td>10</td>
<td>Mindful Attention Awareness Scale [16]</td>
<td>Since the last prompt, I’ve rushed through activities without really being attentive to them.</td>
<td>Mindfulness (R)</td>
</tr>
<tr>
<td>11</td>
<td>Mindful Attention Awareness Scale [16]</td>
<td>Since the last prompt, I’ve found myself doing things without paying attention.</td>
<td>Mindfulness (R)</td>
</tr>
</tbody>
</table>

aIndicates the subscale should be reverse coded when combining to create the data-based total score.
bIndicates the item should be reverse coded when creating the subscale score.
cUPPS-P: Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, and Positive Urgency.

Construct Validity of Momentary Subscales and Total Scores

We examined the relationships between the momentary self-regulation subscales and the trait-level self-regulation subscales by modeling the associations between the momentary subscales and the literature-based and data-based total scores with the trait-level subscale scores assessed at baseline. The results are presented in Table 7. A regression coefficient that is significantly different from 0 would support construct validity, as we expected the momentary subscales to relate to existing scales that measure self-regulation. The strength of the association varied across the baseline measures and momentary subscales. This may be because of different subscales measuring different aspects of self-regulation than the particular baseline measure, or it may be because of a greater degree of individual variability in the momentary scale. The data-based total was associated with the suppression subscale of the ERQ, the nonjudging subscale of the FFMQ, the venturesomeness subscale of the I-7 (marginal, \( P=.05 \)), the MAAS, the SSRQ, and the lack of premeditation and lack of perseverance subscales of the UPPS-P. The literature-based total was associated with the venturesomeness subscale of the I-7, the MAAS, the SSRQ, and the lack of premeditation and lack of perseverance subscales of the UPPS-P.

The momentary mindfulness subscale was significantly associated with the trait-level suppression subscale of the ERQ, the nonjudging subscale of the FFMQ, the venturesomeness subscale of the I-7 (marginal, \( P=.05 \)), the MAAS, the SSRQ, and the lack of premeditation and lack of perseverance subscales in the UPPS-P, indicating various convergent features of the momentary mindfulness subscale with trait-level surveys.

Momentary emotion regulation or self-judgment demonstrated a significant association with the suppression subscale of the ERQ and a marginally significant association with the nonjudging subscale of the FFMQ, indicating the convergent properties of the momentary self-judgment subscale. The momentary perseverance subscale was related to the MAAS, the SSRQ, and the lack of perseverance and lack of premeditation subscales of the UPPS-P. The momentary sensation-seeking subscale was associated with the functional impulsivity subscale of the Dickman Impulsivity Inventory, the venturesomeness subscale of the I-7, and the lack of premeditation subscale of the UPPS-P.
### Table 7. Phase 2 regression coefficients from multilevel models relating baseline trait-level existing self-regulation subscales with momentary self-regulation subscales and total scores (based on 12 items).

<table>
<thead>
<tr>
<th>Trait-level scale (subscale, if applicable)</th>
<th>Momentary-level subscale</th>
<th>Perseverance</th>
<th>Sensation seeking</th>
<th>Data-based total</th>
<th>Literature-based total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dickman Impulsivity Inventory (functional impulsivity) [12]</td>
<td>Mindfulness</td>
<td>−0.21</td>
<td>−0.17</td>
<td>0.37</td>
<td>0.48&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Dickman Impulsivity Inventory (dysfunctional impulsivity) [12]</td>
<td></td>
<td>−0.29</td>
<td>0.27</td>
<td>−0.21</td>
<td>0.32</td>
</tr>
<tr>
<td>Emotion Regulation Questionnaire (suppression) [13]</td>
<td></td>
<td>−0.10&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.08&lt;sup&gt;c&lt;/sup&gt;</td>
<td>−0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Five Facet Mindfulness Questionnaire (nonjudging) [14]</td>
<td></td>
<td>−0.15&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.09</td>
<td>−0.08</td>
<td>−0.01</td>
</tr>
<tr>
<td>Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire (venturesomeness) [15]</td>
<td></td>
<td>−0.25&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.02</td>
<td>0.14</td>
<td>0.39&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mindful Attention Awareness Scale [16]</td>
<td></td>
<td>0.17&lt;sup&gt;c&lt;/sup&gt;</td>
<td>−0.03</td>
<td>0.11&lt;sup&gt;c&lt;/sup&gt;</td>
<td>−0.02</td>
</tr>
<tr>
<td>Short Self-Regulation Questionnaire [18]</td>
<td></td>
<td>0.12&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.03</td>
<td>0.24&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.01</td>
</tr>
<tr>
<td>UPPS-P&lt;sup&gt;e&lt;/sup&gt; Impulsive Behavior Scale (lack of premeditation) [19,20]</td>
<td></td>
<td>−0.20&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.01</td>
<td>−0.19&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.23&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>UPPS-P Impulsive Behavior Scale (lack of perseverance) [19,20]</td>
<td></td>
<td>−0.17&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.02</td>
<td>−0.33&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>Data-based total comprising the mean of mindfulness, perseverance, reverse-coded self-judgment, and reverse-coded sensation seeking.

<sup>b</sup>Literature-based total comprising the mean of mindfulness, perseverance, self-judgment, and reverse-coded sensation seeking.

<sup>c</sup>P ≤ .01.

<sup>d</sup>P ≤ .05.

<sup>e</sup>UPPS-P: Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, and Positive Urgency.

### Predictive Validity

The associations between the subscales and the total scores and behavioral characteristic variables are presented in Table 8. The mindfulness subscale was significantly negatively associated with the TFEQ-R18 total, TFEQ-R18 uncontrolled eating, TFEQ-R18 emotional eating, and having never smoked and positively associated with age, such that older individuals had higher scores on momentary mindfulness. The momentary self-judgment subscale was significantly positively associated with the DAST-10 total, TFEQ-R18 total, TFEQ-R18 uncontrolled eating, and TFEQ-R18 cognitive restraint.
Table 8. Regression coefficients from multilevel models relating demographics and health behaviors to momentary self-regulation subscales and total scores.

<table>
<thead>
<tr>
<th>Demographic and health behavior measures</th>
<th>Mindfulness</th>
<th>Self-judgment</th>
<th>Perseverance</th>
<th>Sensation seeking</th>
<th>Data-based total&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Literature-based total&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AUDIT&lt;sup&gt;c&lt;/sup&gt; total</strong></td>
<td>–0.03</td>
<td>0.03</td>
<td>0</td>
<td>0.06&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–0.03&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–0.02</td>
</tr>
<tr>
<td><strong>AUDIT categories</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.14</td>
<td>–0.9</td>
<td>0.23</td>
<td>–1.14&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.9&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.43</td>
</tr>
<tr>
<td>2</td>
<td>1.29&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–1.04</td>
<td>0.42</td>
<td>–0.72</td>
<td>0.92&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.37</td>
</tr>
<tr>
<td>3</td>
<td>–0.18</td>
<td>0.25</td>
<td>0.24</td>
<td>0.03</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>4 (reference)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>CUDIT-R&lt;sup&gt;f&lt;/sup&gt; total</strong></td>
<td>–0.06</td>
<td>0.05</td>
<td>0.02</td>
<td>0.06&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–0.04</td>
<td>–0.01</td>
</tr>
<tr>
<td><strong>DAST-10&lt;sup&gt;g&lt;/sup&gt; total</strong></td>
<td>–0.09</td>
<td>0.09&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.06</td>
<td>0.11&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–0.06</td>
<td>–0.02</td>
</tr>
<tr>
<td><strong>TFEQ-R18&lt;sup&gt;h&lt;/sup&gt; total</strong></td>
<td>–0.03&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.03&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–0.01</td>
<td>0.003</td>
<td>–0.02&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>TFEQ-R18 uncontrolled eating</strong></td>
<td>–0.01&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.004</td>
<td>–0.01</td>
<td>–0.0004</td>
<td>–0.01&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–0.004</td>
</tr>
<tr>
<td><strong>TFEQ-R18 emotional eating</strong></td>
<td>–0.01&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.01&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–0.01</td>
<td>0.001</td>
<td>–0.005&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–0.002</td>
</tr>
<tr>
<td><strong>TFEQ cognitive restraint</strong></td>
<td>–0.003</td>
<td>0.01&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.003</td>
<td>0.003</td>
<td>–0.004</td>
<td>0.003</td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>–0.36</td>
<td>0.19</td>
<td>–0.26</td>
<td>0.25</td>
<td>–0.27</td>
<td>–0.14</td>
</tr>
<tr>
<td>Never</td>
<td>–0.61&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.30</td>
<td>–0.39</td>
<td>0.09</td>
<td>–0.35&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–0.21</td>
</tr>
<tr>
<td>Past (reference)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>0.03&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–0.02</td>
<td>–0.003</td>
<td>–0.03&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.02&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Sex (male)</strong></td>
<td>–0.22</td>
<td>0.08</td>
<td>0.19</td>
<td>0.31</td>
<td>–0.11</td>
<td>–0.05</td>
</tr>
<tr>
<td><strong>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;)</strong></td>
<td>–0.02</td>
<td>–0.01</td>
<td>–0.02</td>
<td>–0.01</td>
<td>–0.01</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>–0.15</td>
<td>0.36</td>
<td>–0.27</td>
<td>0.28</td>
<td>–0.27</td>
<td>–0.05</td>
</tr>
<tr>
<td>High School or GED&lt;sup&gt;i&lt;/sup&gt;</td>
<td>–0.07</td>
<td>0.08</td>
<td>–0.25</td>
<td>0.22</td>
<td>–0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>–0.27</td>
<td>0.48</td>
<td>0.33</td>
<td>0.09</td>
<td>–0.13</td>
<td>–0.05</td>
</tr>
<tr>
<td>Some college (reference)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Income (US $)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30,000</td>
<td>0.02</td>
<td>0.16</td>
<td>–0.23</td>
<td>–0.01</td>
<td>–0.09</td>
<td>–0.06</td>
</tr>
<tr>
<td>30,000-50,000</td>
<td>–0.24</td>
<td>0.08</td>
<td>–0.98&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.41</td>
<td>–0.44&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–0.41&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>50,000-70,000</td>
<td>0.01</td>
<td>0.05</td>
<td>–0.63&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–0.05</td>
<td>–0.16</td>
<td>–0.16</td>
</tr>
<tr>
<td>&gt;70,000 (reference)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup>Data-based total comprises the means of mindfulness, perseverance, reverse-coded self-judgment, and reverse-coded sensation seeking.

<sup>b</sup>Literature-based total comprises the means of mindfulness, perseverance, self-judgment, and reverse-coded sensation seeking.

<sup>c</sup>AUDIT: Alcohol Use Disorders Identification Test.

<sup>d</sup>P ≤0.01.

<sup>e</sup>P ≤0.05.

<sup>f</sup>CUDIT-R: Cannabis Use Disorders Identification Test–Revised.

<sup>g</sup>DAST-10: Drug Abuse Screening Test.

<sup>h</sup>TFEQ-R18: Three-Factor Eating Questionnaire-R18 (revised version with 18 questions).

<sup>i</sup>GED: General Educational Development.
The momentary perseverance subscale was negatively associated with only specific income categories; perseverance scores tended to decrease among people with an income ranging between US $30,000 and US $70,000. The momentary sensation-seeking subscale was significantly positively associated with the AUDIT, CUDIT-R, and DAST-10 total scores and negatively associated with age.

The data-based total score (mean of perseverance, mindfulness, reverse-coded self-judgment, and reverse-coded sensation seeking) had the most associations with negative health behaviors and is therefore considered the optimal way of combining the subscales into a total score for the predictive ability for risk behaviors. This data-based total score was significantly negatively associated with the AUDIT total, TFEQ-R18 total, TFEQ-R18 uncontrolled eating, TFEQ-R18 emotional eating, and having never smoked and positively associated with low categories of the AUDIT. The data-based total also had a significant association with older age. Although the pairwise comparison between those in the US $30,000 to US $50,000 income category and those in the ≥US $70,000 income category had significantly different data-based total scores, income was not significantly associated with the data-based total score overall. In contrast, the literature-based total (mean of perseverance, mindfulness, emotion regulation, and reverse-coded sensation seeking) was not related to any negative health behaviors. The literature-based total had a significant association with income (type 3 $P$ value=.04), with those in the US $30,000 to US $50,000 income category reporting a significantly lower literature-based total score than those in the ≥US $70,000 income category.

Discussion

Principal Findings

We developed and tested a momentary self-regulation scale, starting with a broad literature review on the overarching concept of self-regulation. Using an empirically driven, iterative data analytic and refinement process with 23 self-regulation-related surveys, in phase 1 we conducted dimension-level and item-level analyses and reduction through EFA, CFA, and IRT to select the best candidate set of 20 survey items with which to develop the momentary self-regulation scale. The phase 1 results indicated that existing self-regulation scales measure several underlying constructs within self-regulation, including perseverance, sensation seeking, emotion regulation, and mindfulness. We demonstrated that these constructs could be measured at the individual level using a reduced set of items from these scales. In phase 2, we examined the factor structure, item loadings, and within- and between-individual variations to further reduce the 20 candidate items to 4 subscales with 3 items each. The final 12-item momentary self-regulation scale totals and subscales demonstrated construct validity relative to the trait-level scales from which they were derived, as well as predictive validity for health risk behaviors. The phase 2 results provide initial evidence that momentary self-regulation varies both at the individual level and within individuals across time in a real-world setting. The resulting metric may be useful for assessing factors that promote or fail to promote self-regulation, including within-individual variation in self-regulation. It may also be useful in assessing when interventions do or do not promote self-regulation, a putative mechanism of behavior change in many populations.

Research using EMA has illuminated processes that drive not only time-sensitive, environment-responsive health behaviors, such as substance use [44], but also psychological functioning, such as impulsivity [43]. Digital technology has enabled this research so that psychological and behavioral processes can be explored within and between individuals as well as within a nonlaboratory, naturalistic setting. A granular understanding of self-regulation as a mechanism of behavior change is key to understanding the conditions under which health behavior change interventions may produce replicable effects and inform the development and refinement of more effective behavior change interventions. To facilitate this understanding, new measures of known individual characteristics that can be captured on a momentary basis are needed. The current work demonstrates that self-regulation is one construct that can be explored at this level, as it varies both within and between individuals. These findings also indicate that this momentary self-regulation scale can be administered through mobile devices in a naturalistic setting. This scale may be useful in capturing the richness of self-regulatory function in vivo and in changing contexts and may help further inform contextual models of self-regulation.

Limitations

Although the final factor structure and items used in the momentary self-regulation scale have demonstrated evidence for construct validity and predictive validity, as well as intraindividual level variations, this study has several limitations. First, both samples were drawn from a population of MTurk workers whose representativeness is unknown relative to the broader US population. Basic demographic information indicated that the sample had limited racial and ethnic diversity. Future studies should examine whether these findings can be replicated and extended to other populations. Furthermore, the momentary (phase 2) study did not measure environmental contextual factors or social interactions of participants, which may have affected the responses. Future efforts may adapt the study methods and analytic procedures to capture these external factors as momentary self-regulatory dynamics are likely to be interactive with, and reactive to, environmental and internal factors.

In addition, in phase 2, we recruited a sample of 60 individuals who resided in states in the Eastern time zone in this study because of the operational capacity of the research team and the manual process involved in sending text prompts with microsurveys at randomized times. Although we do not have reason to believe that persons from states in the Eastern time zone have significantly different self-regulatory dynamics and characteristics than those in other areas, future studies should examine the validity and reliability of our scale among individuals recruited from geographically diverse areas and groups, including rural and urban areas.
Conclusions

Beginning with existing self-regulation scales, we examined the underlying constructs measured by the full set of items in the scales and then identified a reduced set of items that captured the underlying constructs measured across the scales. After confirming that the construct structure was retained within the reduced set of items, we further piloted the set of items in a momentary study in which we captured data from individuals in the moment, in their naturalistic contexts. Using the momentary study results, we further reduced the item set and demonstrated the initial validity in this sample of a momentary self-regulation scale comprising 12 items spanning 4 momentary self-regulation subscales. To further evaluate the validity and reliability more generally, the momentary self-regulation scale should be evaluated in other samples and contexts. This novel momentary self-regulation scale measures self-regulation on a momentary basis as individuals move through their daily lives and can be administered on mobile devices.

Acknowledgments

The authors wish to thank Dr Oscar Gonzalez at the University of North Carolina at Chapel Hill and Dr Russell Poldrack at Stanford University for their contributions to this work. This research was supported by the National Institutes of Health Science of Behavior Change Common Fund through an award administered by the National Institute on Drug Abuse (grant UH2DA041713) and the National Institutes of Health Science of Behavior Change Columbia University Resource and Coordinating Center (grant 5U24AG052175-02).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Self-report surveys list.
[DOCX File, 17 KB - mental_v9i5e35273_app1.docx ]

References


Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
CFA: confirmatory factor analysis
CUDIT-R: Cannabis Use Disorders Identification Test–Revised
DAST-10: Drug Abuse Screening Test
EFA: exploratory factor analysis
EMA: ecological momentary assessment
ERQ: Emotion Regulation Questionnaire
FFMQ: Five Facet Mindfulness Questionnaire
I-7: Eysenck I-7 Impulsiveness and Venturesomeness Questionnaire
ICC: intraclass correlation coefficient
IRT: item response theory
MAAS: Mindful Attention Awareness Scale
MTurk: Amazon Mechanical Turk
RMSEA: root mean square error of approximation
SSRQ: Short Self-Regulation Questionnaire
TFEQ-R18: Three-Factor Eating Questionnaire-R18 (revised version with 18 questions)
TLI: Tucker-Lewis Index
UPPS-P: Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, and Positive Urgency

©Emily A Scherer, Sunny Jung Kim, Stephen A Metcalf, Mary Ann Sweeney, Jialing Wu, Haiyi Xie, Gina L Mazza, Matthew J Valente, David P MacKinnon, Lisa A Marsch. Originally published in JMIR Mental Health (https://mental.jmir.org), 10.05.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on https://mental.jmir.org/, as well as this copyright and license information must be included.
Acceptability of Web-Based Mental Health Interventions in the Workplace: Systematic Review

Johanna Scheutzow¹, BSc, MSc; Chris Attoe¹,², BSc, OHP, MBA; Joshua Harwood², BSc, MSc, DClinPsych

¹Department of Psychosis, Institute of Psychiatry, Psychology and Neuroscience, School of Academic Psychiatry, King’s College London, London, United Kingdom
²Maudsley Learning, South London & Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Johanna Scheutzow, BSc, MSc
Department of Psychosis
Institute of Psychiatry, Psychology and Neuroscience, School of Academic Psychiatry
King’s College London
16 De Crespigny Park
London, SE5 8AB
United Kingdom
Phone: 44 020 7848 0002
Email: johanna.scheutzow2512@gmail.com

Abstract

Background: Web-based interventions have proven to be effective not only in clinical populations but also in the occupational setting. Recent studies conducted in the work environment have focused on the effectiveness of these interventions. However, the role of employees’ acceptability of web-based interventions and programs has not yet enjoyed a similar level of attention.

Objective: The objective of this systematic review was to conduct the first comprehensive study on employees’ level of acceptability of web-based mental health interventions based on direct and indirect measures, outline the utility of different types of web-based interventions for work-related mental health issues, and build a research base in the field.

Methods: The search was conducted between October 2018 and July 2019 and allowed for any study design. The studies used either qualitative or quantitative data sources. The web-based interventions were generally aimed at supporting employees with their mental health issues. The study characteristics were outlined in a table as well as graded based on their quality using a traffic light schema. The level of acceptability was individually rated using commonly applied methods, including percentile quartiles ranging from low to very high.

Results: A total of 1303 studies were identified through multiple database searches and additional resources, from which 28 (2%) were rated as eligible for the synthesis. The results of employees’ acceptability levels were mixed, and the studies were very heterogeneous in design, intervention characteristics, and population. Approximately 79% (22/28) of the studies outlined acceptability measures from high to very high, and 54% (15/28) of the studies reported acceptability levels from low to moderate (overlap when studies reported both quantitative and qualitative results). Qualitative studies also provided insights into barriers and preferences, including simple and tailored application tools as well as the preference for nonstigmatized language. However, there were multiple flaws in the methodology of the studies, such as the blinding of participants and personnel.

Conclusions: The results outline the need for further research with more homogeneous acceptability studies to draw a final conclusion. However, the underlying results show that there is a tendency toward general acceptability of web-based interventions in the workplace, with findings of general applicability to the use of web-based mental health interventions.

(JMIR Ment Health 2022;9(5):e34655) doi:10.2196/34655

KEYWORDS
acceptability; e-mental health; online mental health interventions; occupational online interventions; employees; mobile phone
Introduction

Background

There is an increasing level of awareness regarding the importance of health and well-being in the workplace [1]. Anxiety, stress, and depression are the dominant mental health issues for workers in the United Kingdom, with a prevalence of 1320 cases per 100,000 workers, causing close to 18 million lost working days per year [2]. Employers have a responsibility to take care of their employees and provide support for both their physical and mental health [3,4].

Web-based mental health interventions are increasingly being used in the work environment as they have the advantage of being cost-effective, efficient, anonymous, location-independent, flexible, and empowering. They are regularly used for both prevention and intervention [5-10].

Web-based interventions also have multiple flaws, including technical difficulties, ethical concerns, increased attrition rates, and low engagement in the absence of guidance by professional support [6,11]. Therefore, it is important to understand the barriers that reduce engagement and acceptability of web-based interventions. Multiple systematic studies provide evidence of the effectiveness of web-based mental health interventions at work [12,13]. Importantly, they outline the need to tailor interventions to populations’ needs, which requires greater insight into barriers and the acceptability of web-based mental health interventions in the workplace.

The acceptability of an intervention includes users’ emotional and cognitive responses to the intervention [14], including affective perceptions, burden and barriers, perceived benefits, understanding of the intervention, opportunity costs, and usability. In practice, this takes into account the individuals’ preferences for features and tools, their willingness to use web-based interventions, their engagement (eg, dropout and attrition rate), and users’ perceived utility or satisfaction with the intervention.

Studying users’ acceptability of new treatments has ethical, methodological (validity), and practical applications [15]. Specifically, ethical obligations include the exploration of reasons for acceptable or unacceptable treatments as perceived by the users. It is important to understand potential barriers to intervention engagement before introducing the intervention to employees. Awareness of intervention efficacy alone does not mean that employees accept web-based interventions as a useful tool for self-help.

Sekhon et al [14] outlined studies assessing interventions’ acceptability by using operational definitions in line with measurable acceptability data (dropout rate and satisfaction rating) and qualitative studies focusing on in-depth user experiences. Current research has been limited to studies on clinical populations. Clinical populations differ significantly in symptom severity, level of risks, functionality, and response to treatment; thus, the results might not be generalizable to occupational populations [13]. Therefore, it is relevant to explicitly assess employees’ acceptability of web-based interventions.

Objectives

This systematic review aimed to assess employees’ acceptability of web-based interventions to improve their mental health. The study aimed to inform intervention design and utility by evaluating user experience and barriers and facilitators to using web-based mental health interventions in the workplace.

Methods

This systematic review was conducted in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [16] and followed the ENTREQ (Enhancing the Transparency in Reporting the Synthesis of Qualitative Research) guidelines [17,18].

Eligibility Criteria

Eligible studies met the Population, Intervention, Comparison, Outcome criteria and included qualitative interviews, quantitative studies including scale measures of satisfaction and forms of attrition rates, and mixed methods studies. Acceptability was assessed by means of both direct (acceptability, satisfaction, and experience) and indirect (compliance, completion, adherence, attrition, and dropout rate) measures. Studies were included if they were available in English and published after 2005.

Population

The population was narrowed down to people aged ≥18 years as there is a difference between child and adult interventions. The participants could be employed part-time or full-time or self-employed. Studies were included if the participants were >60% employed. This threshold guaranteed that the main outcome could be generalized to the eligible population for the study’s purpose.

Intervention

Following the guidance of the meta-review by Joyce et al [19] on general workplace mental health interventions, web-based interventions were kept broad to include those that were conducted at work, had a work-related component, or aimed to treat work-related risk factors (eg, stress, depression, or anxiety). However, eligible interventions had to be exclusively web-based programs or interventions that targeted employed people or were applied in an occupational setting. Interventions or programs could be delivered via a computer program, app, or website. They could also differ in the device used to deliver the content (computer, laptop, or mobile phone) as well as include various forms of multimedia. All interventions aimed to change employees’ behavior or mental health. They could have the aim of preventing, treating, or rehabilitating mental health issues.

Comparison

This review compared randomized controlled trials (RCTs), nonrandomized comparative trials, noncomparative trials, explorative studies, and qualitative studies published between 2005 and 2019.

Outcome

Studies were included if they measured acceptability directly or indirectly by means of qualitative assessment of acceptability.
satisfaction, and experience or the indirect measure of acceptability through compliance, completion, adherence, attrition, or dropout rate. Studies were included that assessed the potential willingness to use interventions or the potential features of interventions that were preferred or addressed as disadvantageous for utility.

**Exclusion Criteria**
Studies were excluded if they did not meet the Population, Intervention, Comparison, Outcome criteria; that is, if they included guidance through coaches, therapists, or face-to-face interactions and were applied to participants who were retired or unemployed (>40%). In addition, studies were excluded if they did not measure acceptability or willingness of use as an outcome variable or used interventions that were not focused on the users’ mental health.

**Data Sources**
The search was conducted in July 2019 and included the following electronic databases: PsycINFO (Ovid), Embase, MEDLINE (Ovid), Global Health (Ovid), and the Cochrane Library Trials (CENTRAL). Backward searching was used to ensure that no key papers were missed.

**Search Strategy**
Databases were searched for studies published between 2005 and 2019. Duplicates were removed (Ovid search option). The Boolean system was applied using AND and OR (Textbox 1) to combine different terminologies of 4 key concepts included in a free-text and keyword search. Specific occupations were added to the general search of employees to increase the likelihood of finding studies on high stress–exposed work settings; for example, in the military or firefighter professions. The search terms used were categorized as occupational settings (employee), web-based interventions, mental ill health, and acceptability of interventions. All the key terms considered American and British spelling.

**Textbox 1.** Search terms organized into 4 concepts.

<table>
<thead>
<tr>
<th>Concepts (combined using <strong>AND</strong>) and corresponding search terms (combined using <strong>OR</strong>)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Employee:</strong></td>
</tr>
<tr>
<td><strong>Mental ill health:</strong></td>
</tr>
<tr>
<td><strong>Acceptance:</strong></td>
</tr>
</tbody>
</table>

**Study Selection**
Duplicates were removed, and titles, abstracts, and full texts were scanned for the inclusion criteria. After the assessment of the full texts’ eligibility by the first author (JS), all the included studies were summarized and synthesized. The study selection process is outlined in the PRISMA flowchart (Figure 1).
Data Collection Process

Data were collected according to the following criteria: reference, characteristics of the intervention, its aims and objectives, study design, population, setting and recruitment, results, acceptability, and—if available—reasons for dropout, as well as qualitative data.

Quality Assessment

The quality of the studies was assessed using the Critical Appraisal Skills Programme (CASP) checklist [20] for both qualitative studies and RCTs. Studies were evaluated based on research design, representativeness, recruitment procedure, presence of a comparison group, dropout rate, validity, reliability, and relevance of the measurement tools. Quality was graded with a traffic light system based on the 10 quality questions of the CASP and answered with yes (green) if the information was present, no information (red), and not available (yellow) if the information was not apparent or clearly outlined within the study (Multimedia Appendix 1 [21-48]). Studies were included in this systematic review irrespective of quality judgment.

Synthesis of Data

Similar to corresponding systematic reviews on the acceptability of web-based interventions [10], the level of acceptability was categorized into the following quartiles: low (−−), moderate (−), high (+), and very high (++). This was specifically used for studies that reported a satisfaction rating on a scale, the percentage results of which could then be transferred to the suggested levels of acceptability. In addition, studies reporting dropout rates and compliance percentages were organized according to the 4-quartile rating system for acceptability. If studies reported mixed results, including positive and negative outcomes on different acceptability factors, they were rated with a tilde (~). Qualitative studies were synthesized in an integrative, meta-agregative style following methodological guidance on the use of meta-aggregations [49] as well as using similar systematic reviews [50]. The key data were extracted and are outlined in Table 1 (see Multimedia Appendix 2 [31-33,35,38,39,45-48] for direct measures and qualitative data sources).
Table 1. Study characteristics.

<table>
<thead>
<tr>
<th>Author and country</th>
<th>Intervention, duration, and aim</th>
<th>Design and recruitment</th>
<th>Population</th>
<th>Results</th>
<th>Acceptance measure</th>
<th>Reasons for dropout</th>
<th>Level (acceptance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott et al [42], Australia</td>
<td>Internet-delivered CBT program for employees with tinnitus distress in industrial organizations; 6 weeks; effectiveness of the program</td>
<td>Clustered RCT comparing CBT intervention group with IOC; recruited in industrial organizations (BP Australia and BHP Billiton)</td>
<td>56</td>
<td>CBT: mean 50.5 (SD 9.5); IOC: mean 48.7 (SD 8.6)</td>
<td>_c</td>
<td>CBT program was similarly effective to the information program for treating tinnitus distress, depression, anxiety, stress, and quality of life</td>
<td>Unknown</td>
</tr>
<tr>
<td>Allexandre et al [21], United States</td>
<td>Web-based interactive educational stress management program (website), <em>Stress Free Now</em>, using mindfulness meditation; 8 weeks; effectiveness of the intervention</td>
<td>RCT comparing 4 groups, including no support, group support, group and expert clinical support, and waitlist control; recruitment via email in a corporate call center</td>
<td>161</td>
<td>Mean 40.0 (SD 12.6) women</td>
<td>49.1% full-time work shift (days); debt collectors and customer service or fraud representatives</td>
<td>Participants favored guided practices and showed low use of program. All groups decreased in perceived stress and improved in psychological and emotional well-being</td>
<td>Web-based use: 10% to 15% (intervention)</td>
</tr>
<tr>
<td>Beiwinkel et al [36], Germany</td>
<td>Web-based program, HelpID, for depression based on CBT and awareness training; 12 weeks; effectiveness of HelpID in reducing sickness absence and depression</td>
<td>RCT comparing intervention with control; recruitment via health insurance</td>
<td>180</td>
<td>Mean 48</td>
<td>68% women</td>
<td>HelpID effectively reduced depressive symptoms</td>
<td>Dropout: 45.5% after the assessment, 67.7% follow-up; satisfaction: 68.2% (mean 2.04, intervention)</td>
</tr>
<tr>
<td>Birney et al [37], United States</td>
<td>Mobile app intervention MoodHacker, CBT-based depression self-management; 6 to 10 weeks; effectiveness of a program to reduce stress and prevent depression, anxiety, and substance abuse among employees</td>
<td>RCT: MoodHacker group compared with alternative care with links to websites on depression; recruited via EAPs and other outlets</td>
<td>300</td>
<td>MoodHacker: mean 40.6 (SD 11.5); alternative care: mean 40.7 (SD 11.2)</td>
<td>56% full-time, 35.3% part-time, and 8.7% self-employed</td>
<td>MoodHacker caused significant effects on depression symptoms compared with alternative care</td>
<td>Attraction: 6.7% follow-up; satisfaction: 76% (mean 4.6/6); system usability: B+</td>
</tr>
</tbody>
</table>

Unknown: _i, _f, _g, _h, _c, _d, _e
<table>
<thead>
<tr>
<th>Author and country</th>
<th>Intervention, duration, and aim</th>
<th>Design and recruitment</th>
<th>Population</th>
<th>Results</th>
<th>Acceptance measure</th>
<th>Reasons for dropout</th>
<th>Level (acceptance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Billings et al [22], United States</td>
<td>Web-based stress and mood management multimedia program for employees based on CBT; 12-week access; effectiveness of the program to reduce depression and increase behavioral activation, knowledge of depression, and performance at work</td>
<td>RCT: experimental and control; recruited from a technology company via email and health fair</td>
<td>309</td>
<td>Most (51%) between 30 and 40</td>
<td>70.6% women</td>
<td>Decrease in stress, increase in knowledge of anxiety and depression as well as positive perception of treatment and improvement in the consumption of alcohol; most used it only once</td>
<td>Acceptance measure (measures of uptake and compliance)</td>
</tr>
<tr>
<td>Bolier et al [23], the Netherlands</td>
<td>Web-based health promotion programs (Colour Your Life, Don’t Panic Online, Drinking Less, Psyfit, and Strong at work) designed for the work setting aiming to decrease stress and prevent substance abuse, depression, and anxiety in health professionals; 6 to 12 weeks; measure effectiveness of the modules</td>
<td>Clustered RCT: web-based condition and waitlist control; recruited nurses and health professionals in a medical center via mail</td>
<td>1140</td>
<td>Mean 40</td>
<td>79.8% women 71.9% nurses</td>
<td>The intervention significantly enhanced positive mental health</td>
<td>Acceptance measure (measures of uptake and compliance)</td>
</tr>
<tr>
<td>Ebert et al [40], Germany</td>
<td>Web-based unguided recovery training, GET.ON, for teachers with insomnia and psychological detachment from the workplace; 6 weeks; psychological efficacy of GET.ON</td>
<td>RCT: intervention and waitlist control group; recruited via email at schools by the Ministry of Education (Germany, NRW1)</td>
<td>64</td>
<td>Mean 48.5 (SD 9.9)</td>
<td>74.2% women</td>
<td>Significant reduction in insomnia severity</td>
<td>Completion rate: 48.4% all sessions</td>
</tr>
</tbody>
</table>

Participant details:
- Sample size, N
- Age (years)
- Gender
- Employment details
- Results
- Acceptance measure
- Reasons for dropout
- Level (acceptance)
<table>
<thead>
<tr>
<th>Author and country</th>
<th>Intervention, duration, and aim</th>
<th>Design and recruitment</th>
<th>Population</th>
<th>Results</th>
<th>Acceptance measure</th>
<th>Reasons for dropout</th>
<th>Level (acceptance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ebert et al [24], Germany</td>
<td>Unguided web-based stress management program, GET.ON Stress, for employees using problem solving and emotional regulation; 7 weeks; efficacy of the program</td>
<td>RCT: intervention or waitlist control; recruited from general employees via the occupational health program of a health insurance company as well as via contacted HR departments in Germany</td>
<td>Sample size, N: 264</td>
<td>72% women</td>
<td>75% full-time; diverse sectors including economy, health, service, and social</td>
<td>Effectively reduced symptoms of mental and work-related stress among employees with stress</td>
<td>—</td>
</tr>
<tr>
<td>Hama-mura et al [25], Japan</td>
<td>Computer-delivered intervention (app), Self-Record, that facilitates cognitive restructuring for distress and alcohol consumption through self-monitoring of thought and activities; 4 weeks; effectiveness of the intervention on mental distress and consumption of alcohol</td>
<td>Pilot non-RCT, quasi-experiment with intervention and control groups; recruited via research marketing company</td>
<td>Mean 38.82 (SD 9.58)</td>
<td>71.6% employed by a company, 7.5% employed by the government or a non-profit organization, 6.3% self-employed, and 3.1% professionals</td>
<td>Intervention heightened participants’ perception of their pathological thoughts and alcohol consumption, whereas they only decreased face to face</td>
<td>Attrition rate: 42% (7 sessions); dropout: 90% provided follow-up data; satisfaction (high): 95% overall</td>
<td>—</td>
</tr>
<tr>
<td>Heber et al [26], Germany</td>
<td>Stress website, GET.ON, including psychoeducation and interactive exercises tailored through personalized feedback; 4 weeks; efficacy of the intervention</td>
<td>RCT: intervention and waitlist control group; recruited by the Ministry of Education from the general working population showing symptoms of stress and through newspaper articles</td>
<td>Mean 43.3 (SD 10.2)</td>
<td>73.1% women</td>
<td>77.3% full-time</td>
<td>Web-based interventions effectively decreased stress in employees</td>
<td>Dropout rate: 15.3% follow up; adherence: 64.8% (intervention) stopped after the first day</td>
</tr>
<tr>
<td>Ketelaar et al [27], the Netherlands</td>
<td>eMH interventions for health professionals—Pay-fit, Strain at work, Colour your life, Don’t panic online, and Drinking less: self-help on the internet (CBT and other); evaluate eMH approach targeting work functioning and psychological well-being</td>
<td>RCT, randomization at ward level with intervention and control groups; recruited nurses and health professionals employed at an academic hospital</td>
<td>Mean 39.5</td>
<td>80% men</td>
<td>—</td>
<td>eMH approach was not more effective than a control to increase work functioning and psychological well-being</td>
<td>Compliance rate: 6% started the intervention; dropout: 45% to follow-up</td>
</tr>
</tbody>
</table>

**Results Population**

**Design and recruitment**

**Acceptance measure**

**Reasons for dropout**

**Level (acceptance)**
<table>
<thead>
<tr>
<th>Author and country</th>
<th>Intervention, duration, and aim</th>
<th>Design and recruitment</th>
<th>Population</th>
<th>Results</th>
<th>Acceptance measure</th>
<th>Reasons for dropout</th>
<th>Level (acceptance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ly et al [28], Sweden</td>
<td>Mobile phone stress management intervention for managers including short audio lectures, information, and exercise focusing on acceptance and commitment therapy; 6 weeks; efficacy of the smartphone treatment</td>
<td>RCT: stress intervention and waitlist control group; recruitment took place after a presentation about the project at multiple organizations (Swedish or American) and via advertisements on the internet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevedal et al [43], United States</td>
<td>Digital health coaching program for chronic pain management using psychoeducation on self-management, coping, and stress; 4 weeks; effectiveness of the program on work interference, activity, stress, pain, quality of life, and health</td>
<td>Case report; 1-group design; recruited via mailings, emails, and posted communications within 37 American organizations or a member of 1 of 18 health care plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feicht et al [44], Germany</td>
<td>Happiness exercises to develop a positive psychological state; 7 weeks; examined the impact of the intervention on psychological and physiological parameters</td>
<td>Longitudinal design (2 groups—intervention and control); recruited via local insurance company in Germany (2 participating departments were chosen by the company)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thiart et al [29], Germany</td>
<td>Internet-based CBT-I intervention, GET.ON Recovery, for stress, work-related strain, and insomnia in teachers; 6 weeks; evaluate the efficacy of the intervention</td>
<td>RCT: intervention group and waitlist control; recruited via email sent to schools in Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author and country</td>
<td>Intervention, duration, and aim</td>
<td>Design and recruitment</td>
<td>Population</td>
<td>Results</td>
<td>Acceptance measure</td>
<td>Reasons for dropout</td>
<td>Level (acceptance)</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------------------</td>
<td>------------------------</td>
<td>------------</td>
<td>---------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Umanodan et al [30], Japan</td>
<td>Computer-based stress management training using self-paced behavioral, communication, and cognitive techniques; 7 weeks; effectiveness of the program in improving mental health and performance at work</td>
<td>Clustered RCT; recruited via informational posters and the supervisor during meetings in a manufacturing company</td>
<td>Sample size, N = 263</td>
<td>Mean Age 38.85</td>
<td>Gender: 92.6% men, 23% managers</td>
<td>Knowledge about stress management and coping skills increased (if participants had enough time)</td>
<td>Completion rate (intervention group): 89%; High baseline levels of distress increased the chance of dropout</td>
</tr>
<tr>
<td>Wood et al [41], United States</td>
<td>Resilience mobile app to decrease burnout (assessment tools); 4 weeks; assess usability, acceptability, and effectiveness</td>
<td>Pilot study; recruited mental health care professionals from a health care system</td>
<td>Sample size, N = 30</td>
<td>Mean Age 42.5 (SD 12)</td>
<td>Gender: 43% psychologists, 30% social workers, 13% psychiatric nurses, 7% psychiatrists, and 7% other</td>
<td>App reduced burnout and compassion fatigue in participants</td>
<td>System usability (overall): 79.4%</td>
</tr>
<tr>
<td>Bush et al [31], United States</td>
<td>T2 Mood Tracker mobile app to track symptoms associated with deployment-related behavioral health issues (well-being, anxiety, stress, PTSD, injury, and depression); 1.4 weeks; assessment of the utility of the app</td>
<td>Mixed methods design; qualitative and quantitative (Likert-style and open-ended questions); recruited via posters and flyers distributed by WTU</td>
<td>Sample size, N = 8</td>
<td>Mean Age 45 (based on the sample N=82)</td>
<td>Gender: 78% women</td>
<td>The app was perceived as easy to use, helpful, and beneficial</td>
<td>Useful rating: 88%; qualitative: utility rating positive but could incorporate additional factors to make it more manifold</td>
</tr>
<tr>
<td>Carolan et al [32], United Kingdom</td>
<td>Web-based stress management intervention, WorkGuru (CBT, mindfulness, and problem solving); 8 weeks; employees’ attitude toward digital mental health interventions at work</td>
<td>Qualitative study: 18 semistructured interviews (taken from previous RCT with and without access to a web-facilitated discussion group); recruited from 6 UK-based organizations and invited via mail (universities, local authorities, third sector, and telecommunication)</td>
<td>Sample size, N = 18</td>
<td>Mean Age 45 (based on the sample N=82)</td>
<td>Gender: 78% office and 22% mixture of office and client work</td>
<td>Outlined advantages of digital mental health interventions, but high barriers appeared with the application in the workplace</td>
<td>Engagement: 39%; qualitative: preference for short, interactive, easy to use, personalized, and anonymous interventions and access via computer or mobile phone</td>
</tr>
<tr>
<td>Author and country</td>
<td>Intervention, duration, and aim</td>
<td>Design and recruitment</td>
<td>Population</td>
<td>Results</td>
<td>Acceptance measure</td>
<td>Reasons for dropout</td>
<td>Level (acceptance)</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------------------</td>
<td>------------------------</td>
<td>------------</td>
<td>---------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Deady et al [48], Australia</td>
<td>Emergency service workers’ attitude toward mobile mental health apps</td>
<td>Cross-sectional study; recruited from 4 metropolitan Fire and Rescue stations</td>
<td>106</td>
<td>Participants showed positive perception and interest in using mental health apps but had preferences regarding language, features, and therapeutic techniques</td>
<td>—</td>
<td>—</td>
<td>+</td>
</tr>
<tr>
<td>Deady et al [38], Australia</td>
<td>Acceptance and effectiveness study on HeadGear, an app-based program aiming to decrease depressive symptoms and increase well-being; 5 weeks</td>
<td>2-stage pilot study; recruited via email and Facebook from industrial organizations (agriculture, freight or postage, and mining)</td>
<td>Stage 1: 21; stage 2: 84</td>
<td>HeadGear was effective and reduced symptoms significantly. However, attrition rate was high</td>
<td>Utility: 40% to 50% would use it; qualitative: most appreciated the utility, helpfulness, overall ease, and accessibility but complained about engagement and navigation issues</td>
<td>—</td>
<td>–</td>
</tr>
<tr>
<td>Eklund et al [33], Sweden</td>
<td>University staff’s experiences of a customized, interactive, web-based program that aims to change behavior in stress management as well as explore intervention adjustments</td>
<td>Explorative qualitative study: semistructured interviews; recruitment via 3 departments at the university</td>
<td>9</td>
<td>Staff accepted a web-based program for stress-related problems</td>
<td>Acceptance was positive as long as it was short in time and applied in a transparent and tailored way</td>
<td>—</td>
<td>+</td>
</tr>
<tr>
<td>Author and country</td>
<td>Intervention, duration, and aim</td>
<td>Design and recruitment</td>
<td>Population</td>
<td>Results</td>
<td>Acceptance measure</td>
<td>Reasons for dropout</td>
<td>Level (acceptance)</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------</td>
<td>------------------------</td>
<td>------------</td>
<td>---------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Henne mann et al [34], Germany</td>
<td>Employees’ acceptance of organizational eMH interventions focusing on work-related distress</td>
<td>Longitudinal cohort study; self-administered questionnaire; recruited employees showing health problems and previous sickness absence</td>
<td>1829</td>
<td>Mean 49.93 (SD 4.06)</td>
<td>Attitudes toward organizational eMH interventions were disadvantageous</td>
<td>Acceptance (low): 89.1%; suggestions for improvement of acceptance: previous education (awareness and attitudes regarding efficacy and usability)</td>
<td>Higher scores in men and high-education group, those with previous experience with eHealth, and mentally demanding work types; lower scores in those diagnosed with a mental health disorder and non–internet users</td>
</tr>
<tr>
<td>Peters et al [45], Australia</td>
<td>Explorative workshop of perceptions, thoughts, and preferences of employees in male-dominated workplaces to build and adapt a mental health mobile app</td>
<td>Exploratory qualitative study; recruited via emails distributed to 2 organizations (state fire and rescue service and a freight transport organization)</td>
<td>60</td>
<td>Between 26 and 65</td>
<td>Relevance of considering language use and preferred features and balancing preferences with the need for evidence-based interventions</td>
<td>Men preferred unstigmatized language use, a simple mood management app, and guidance involvement</td>
<td>— —</td>
</tr>
<tr>
<td>Schneider et al [39], United Kingdom</td>
<td>Views and acceptance of 2 self-help applications for depression: MoodGYM (cCBT) and informational websites applied at work; 5 weeks</td>
<td>Mixed methods; recruited from 3 organizations: 2 private enterprises (telecommunications and transport) and 1 health organization</td>
<td>637</td>
<td>Mean 42 (SD 9.6)</td>
<td>Evidence-based computerized approaches supported acceptability, which could be increased by taking care of barriers and users’ expectations</td>
<td>Dropout: 63%; positive rating: 24%; various intrinsic and extrinsic barriers that lead to a high unacceptance; acceptance increases with interactive support</td>
<td>Intrinsic: intrapersonal problems; extrinsic: technical problems; generic: perception of cCBT</td>
</tr>
<tr>
<td>Wang and Ho [46], Canada</td>
<td>Explorative study on barriers and preferences for specific features among male workers in a mental health tool</td>
<td>Cross-sectional study; recruited by random digit-calling method to households collecting data from 511 men with risk of depression</td>
<td>841</td>
<td>Mean 44.3 (SD 13.7)</td>
<td>Overall positive results, but men’s preferences and perceived barriers should be taken into account to increase acceptability</td>
<td>Acceptance in men was good, but apps should be mobile and tailored to preferences, including various topics and designs</td>
<td>Having high risk of depression at baseline increased the chance to see the utility of the intervention compared with low-risk individuals (83.4% vs 75%)</td>
</tr>
<tr>
<td>Author and country</td>
<td>Intervention, duration, and aim</td>
<td>Design and recruitment</td>
<td>Population</td>
<td>Results</td>
<td>Acceptance measure</td>
<td>Reasons for dropout</td>
<td>Level (acceptance)</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------</td>
<td>------------------------</td>
<td>------------</td>
<td>---------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Williams et al [35], United States</td>
<td>Feasibility of a web-enhanced behavioral self-management program, Stress Gym, in a military setting built on the model of cognitive appraisal by Lazarus and Folkman</td>
<td>Cross-sectional study; recruited and invited all active-duty members at Naval Medical Center, Portsmouth, Virginia</td>
<td>Sample size, N: 142</td>
<td>Mean age: 41.1 (SD 9.2)</td>
<td>Gender: 55% women</td>
<td>Employment details: 24% officers and 76% enlisted sailors</td>
<td>Supported the feasibility of Stress Gym as being a web-based CBT-based self-help intervention accepted by the users and demonstrated reduction in stress</td>
</tr>
<tr>
<td>Wilson et al [47], United States</td>
<td>Soldiers’ attitude toward technology-based approaches to mental health care</td>
<td>Cross-sectional study; recruited from pre- and postdeployment clinic (in the waiting room for screening visits)</td>
<td>Sample size, N: 352</td>
<td>Mean age: 25.9 (SD 5.8)</td>
<td>Gender: 92% men</td>
<td>—</td>
<td>Feasibility of technology-based approaches was supported</td>
</tr>
</tbody>
</table>

\[a\] Sorted from indirect to more direct measures.  
\[b\] CBT: cognitive behavioral therapy.  
\[c\] RCT: randomized controlled trial.  
\[d\] IOC: information-only control.  
\[e\] Data missing or not relevant.  
\[f\] Mixed results.  
\[g\] Low.  
\[h\] Moderate.  
\[i\] EAP: employee assistance program.  
\[j\] Very high.  
\[k\] High.  
\[l\] NRW: North Rhine-Westphalia.  
\[m\] HR: human resources.  
\[n\] eMH: e-mental health.  
\[o\] CBT-I: cognitive behavioral therapy for insomnia.  
\[p\] PTSD: posttraumatic stress disorder.  
\[q\] WTU: Warrior Transition Unit.  
\[r\] cCBT: computerized cognitive behavioral therapy.  

### Results

#### Review Process

The characteristics of the studies are outlined in Table 1 as well as in Multimedia Appendix 2 including more details. Within the review process, 1303 papers were identified, of which 363 (28%) were duplicates, not published in English, or published before 2005. Of the 1303 papers, titles and abstracts were then scanned, a process that identified 940 (72%) and 379 (29%) papers, respectively. Papers were excluded if the interventions were independent of work environments, did not include most employees, and did not focus on mental health issues. Most studies were excluded owing to the involvement of face-to-face or telephone guidance by a coach or therapist. Ultimately, 28 studies were identified for further analysis, which either reported indirect measures of the acceptability of web-based interventions (n=17, 61%) or provided qualitative data on acceptability (n=11, 39%; Figure 1).
Study Characteristics

The 28 included studies had an overall sample of 9739 participants, with sample sizes ranging from 8 to 1140. The mean age of the participants was 40.7 years, and most participants were White and employed full-time.

The studies had various differing methodological designs and study characteristics, which are outlined in Table 1 as well as in Multimedia Appendix 2 including more details on direct measures. Interventions were heterogeneous in type, application outcome focus, length, setting (within a specific organization or random employees), and characteristics of the participants (type of profession and demographics). In addition, they differed in the level of support potentially provided in web-based format (email or message) versus unguided. Most interventions (18/28, 64%) were focused on reducing stress [21-35] and depressive symptoms [23,31,36-39] in employees. Other interventions focused on insomnia (2/28, 7%) [29,40], anxiety (2/28, 7%) [23,31], panic (2/28, 7%) [23,27], psychological detachment from work (1/28, 4%) [40], resilience (burnout; 1/28, 4%) [41], mood (1/28, 4%) [22], tinnitus distress (1/28, 4%) [42], chronic pain (1/28, 4%) [43], substance misuse (2/28, 7%) [23,25], and well-being or happiness (3/28, 11%) [31,38,44]. Studies and interventions often included more than one focus of mental illness, used various treatment techniques, and assessed the acceptability of general web-based mental health interventions [34,45,46]. Cognitive behavioral therapy (CBT) was the most used form of intervention (9/28, 32%). Other interventions applied mindfulness (3/28, 11%), psychological education (3/28, 11%), cognitive appraisal or restructuring (2/28, 7%), emotional regulation (1/28, 4%), acceptance and commitment therapy (1/28, 4%), problem solving (2/28, 7%), exercise (1/28, 4%), or communication strategies (1/28, 4%). Approximately 7% (2/28) of the studies used tracking or assessment tools for burnout and mood (eg, depression, stress, or well-being). The studies were predominantly RCTs (13/28, 46%). CBT was mostly used in interventions that tackled depression or stress at work. For example, the CBT interventions focused on depression were HelpID, Mood Hacker, Colour Your Life, and MoodGym. The interventions that used CBT and aimed to reduce stress and mental strain were Psyfit, Strong at work, GetON Recovery, and Work Guru.

Of the 28 studies, 13 (46%) were RCTs, 4 (14%) were cross-sectional studies, 3 (11%) were qualitative studies, 3 (11%) were pilot studies, 2 (7%) were longitudinal studies, 2 (7%) were mixed methods studies, and 1 (4%) was a case report. The studies mostly used waitlist control groups, internet-based information website groups, or variations of intervention type groups as comparators. The studies originated in the United States (8/28, 29%), Germany (7/28, 25%), Australia (4/28, 14%), the Netherlands (2/28, 7%), Japan (2/28, 7%), Sweden (2/28, 7%), the United Kingdom (2/28, 7%), and Canada (1/28, 4%).

Methodological Quality

The methodological quality of the studies is summarized in Multimedia Appendix 1. The studies were assessed for quality using the CASP [20] qualitative and quantitative templates and reported in the form of a traffic light schema. Various quality flaws were outlined in the studies, and no study met all 10 criteria marked by the CASP [20]. Independent of quality, all studies (28/28, 100%) were included in the final synthesis. Allocation bias appeared to be low in the quantitative studies as participants were mostly randomly distributed to their condition (15/17, 88%). However, qualitative studies often indicated performance and detection bias as studies often missed reporting on blinding status or researchers’ awareness of the participants’ condition. Selection bias was generally high as participants repeatedly originated from specific population samples (eg, male-dominated industry workers or female educational staff). Attrition bias was predominantly high as various studies reported a high dropout rate, which weakened their generalizability. Several studies missed reporting on the specific demographics of their samples and, thus, might risk the presence of confounders, whereas other studies (4/28, 14%) clearly outlined their risk of confounding [25,26,30,42]. The analysis of quantitative studies was generally good as all studies used data from all participant groups in their final analysis. Qualitative studies showed generally good quality in the guidance of clear questions, taking care of ethical considerations, and the provision of clear information on methodology. However, various studies missed accounting for the potential bias caused by the relationship between the researchers and the participants.

Setting and Types of Employees

The recruitment setting and included characteristics of the employees were very diverse (Table 1). Some studies (10/28, 36%) used the whole working population, recruiting samples via local insurance companies, occupational programs or employee assistance programs, random digit calling, or advertisements [24,26,28,34,36-38,43,44,46]. Alternatively, some studies used specific population samples originating from 1 type of profession. In particular, samples were recruited from the military (2/28, 7%) [35,47], telecommunications (2/28, 7%) [21,39], transport (2/28, 7%) [39,45], the public sector (1/28, 4%) [39], state fire and rescue services (2/28, 7%) [33,45], office and client employees (1/28, 4%) [32], university staff and teachers (3/28, 11%) [29,33,40], clinical staff and health professionals (4/28, 14%) [23,27,31,41], manufacturing and industrial workers (2/28, 7%) [30,42], marketing (1/28, 4%) [25], and technology (1/28, 4%) [22]. Most studies were conducted in the United States (8/28, 29%) and Germany (7/28, 25%). Germany primarily recruited from the general working population [26,34,36,44], whereas the United States mainly recruited from multiple specific locations, including larger organizations (eg, corporate call centers and technology companies), health centers, and military-related workplaces. Their acceptability results were mixed in outcome, ranging from very high in the study by Beiwinkel et al [36] to very low in the study by Hennemann et al [34]. The synthesis did not outline any pattern of setting and participant characteristics that was associated with the acceptability level of web-based interventions.

Intervention Characteristics and Country of Conduct

As outlined in the Study Characteristics section, most studies used CBT in their administered interventions (9/28, 32%). CBT was relatively equally distributed across Western countries,
including the United States (2/9, 22%), Germany (2/9, 22%), the Netherlands (2/9, 22%), the United Kingdom (2/9, 22%), and Australia (1/9, 11%). Summarizing the CBT studies, the acceptability level indicated that 44% (4/9) of the studies had a low to moderate level of acceptability, whereas 33% (3/9) of the studies showed a high to very high acceptability level. Approximately 11% (1/9) of the studies had a mix of moderate and high acceptability levels. Other analyzed intervention types (mindfulness, psychological education, cognitive appraisal, emotional regulation, acceptance and commitment therapy, problem solving, cognitive strategies, exercise, and tracking tools) did not indicate any pattern of acceptability level. Broadly speaking, the intervention type, country of conduct, and outcome of the study did not indicate any notable patterns. However, most of the studies (26/28, 93%) were conducted in Western countries.

**Measure of Acceptability**

Relevant studies measured acceptability in different ways. They used direct measures of acceptability, which included qualitative data through questionnaires and interviews, or indirect quantitative measures by means of take-up, dropout, compliance, adherence, attrition, or completion rate. Some studies used both direct and indirect measures. All measures of acceptability are outlined in either Table 1 or Multimedia Appendix 2 (qualitative synthesized data) in the context of the reference, intervention, sample, study design, recruitment, outcome, indirect and direct acceptability measures, available reasons for dropout, example quotations from interviews, and an individually rated acceptability level.

**Direct Measure of Acceptability**

Table 1 and Multimedia Appendix 2 present the direct outcome of employees’ acceptability of web-based therapy in the workplace. When categorizing the qualitative outcome into key themes, the following topics commonly emerged: (1) general interest in or willingness to use web-based interventions, (2) employees’ satisfaction rating of the utility of the interventions, and (3) preferred features of the design and application style of the interventions. Most participants reported a generally positive interest in and acceptability of web-based interventions [31,33,35,38,46]. However, there were mixed results and negative opinions in other studies [32,33,39,48]. Common preferred features of web-based mental health interventions were the use of nonstigmatized language [45,48], the preference for interventions with interactive support [39,45], and broad application spectrum as well as short mobile and interactive multimedia interventions [31,35,38,48]. The synthesized outputs of the studies were written in descriptions of each theme as well as provided within the context of the setting and intervention type. To deliver a deeper insight into common themes, Multimedia Appendix 2 provides quotations of interviewees in primary studies. As this systematic review synthesized key themes in an integrative, meta-aggregative way, quotations aid primary studies. As this systematic review synthesized key themes, the following topics commonly emerged: (1) general interest in or willingness to use web-based interventions, (2) employees’ satisfaction rating of the utility of the interventions, and (3) preferred features of the design and application style of the interventions. Most of the studies reported details on the satisfaction rating on a scale associated with web-based interventions. Employees were mostly satisfied with the interventions and rated their utility positively.

In addition, multiple studies assessed acceptability using satisfaction, usability, or interest ratings of the intervention (Table 1). Satisfaction ratings were frequently used in the studies [22,24,26,29,31,36-38,41,43]. The average satisfaction score was 82.6%, which is similar to a very high individual-defined acceptability level (++) of web-based interventions. Moreover, 14% (4/28) of the studies reported a score of 0.85% for practical use [22,31,38,41], equivalent to a high (+) acceptability. In particular, Wilson et al [47] reported a rate of 75% in “comfortability” of using a mental health program on the computer and an 84% rate in “willingness to use.” In contrast, Hennemann et al [34] reported that 89.1% of participants rated low on the “acceptability” of general occupational web-based mental health interventions. Both studies were very heterogeneous in intervention specificity and sample population. Hennemann et al [34] explained the negative outcome by the direct predictor variables of acceptability with “social influence,” “effort” and “performance expectancy,” and “time spent in the web” as well as the “frequency of searching online for health information.”

**Indirect Measure of Acceptability**

This systematic review included hypothetical measures of acceptability characterized by dropout, attrition, compliance, adherence, uptake, and completion rate. The indirect or hypothetical measures of the acceptability of web-based interventions in the workplace are summarized in Table 1. The mean percentage of dropout rates from the included studies was 50.9% with a range of 15.3% [25] to 67.7% [36], which is equivalent to a moderate individual-defined level of acceptability [23-25,27,36,39,44]. A few studies reported the reasons for dropout or termination of the interventions. Repeated reasons were lack of time [21,26,40], technical difficulties [26,27,39,40], younger age [23,27,36], lower education [36], lack of motivation [26,40], no need for help [40], ability to manage stress personally [22], dissatisfaction with the intervention [26,39], higher initial level of psychological distress [30], and privacy concerns [31]. Other measures of acceptability included an average attrition rate of 32% [24,37,42], an average adherence rate of 54% [25,28], an uptake and intervention start rate of 11% [23,27], and a completion rate of 68% [30,40]. As visible in the outcome, there was no clear consensus in acceptability level, and the comparison of studies was difficult as they were heterogeneous in study design, sample, and methodology. However, the most frequently reported indirect measure of acceptability was the dropout rate, supported by a moderate (−) level of acceptability.

**Discussion**

**Principal Findings**

This systematic review assessed the levels of employees’ acceptability of web-based interventions aimed at improving mental health. The findings showed a generally positive level of acceptability and highlighted various factors to be considered in making interventions acceptable, engaging, and useful for employees. Themes to be addressed with caution when introducing interventions are the use of stigmatized terminology, including words of ill health and mental illness. In terms of

https://mental.jmir.org/2022/5/e34655

Scheutzow et al

JMIR Mental Health 2022 | vol. 9 | iss. 5 | e34655 | p.49

(page number not for citation purposes)
implementation, applications are recommended to be short and use interactive multimedia tools.

Results were obtained from 28 separate studies. Satisfaction ratings and feedback appeared positive, particularly when the interventions included multimedia and nonstigmatizing language. In particular, 79% (22/28) of the studies showed acceptability measures from high to very high, and 54% (15/28) of the studies reported acceptability levels from low to moderate (overlap when studies reported both quantitative and qualitative results). The average satisfaction rating was >80%, and the employees rated the interventions’ utility as good overall. However, quantitative measures contradicted the universal positive perspective of web-based interventions by means of the common measured dropout rate of approximately 50%. Hence, the attrition rate was very high in multiple studies, which questions the efficacy of unguided self-applied interventions.

Collectively, these results are in line with other acceptability studies that supported the general acceptability of web-based interventions in clinical settings [51]. Various studies have outlined barriers to assessing acceptability; for example, negative results from indirect measures. In addition, complications in synthesis owing to the heterogeneity of the interventions have been repeatedly reported.

Stigma and attitudes toward mental health at work were an emerging theme. Acceptability levels may relate to the web-based interventions themselves or to the fact that the intervention relates to mental health. This is supported by other studies showing that there is fear of stigmatization when seeking support [52]. It may also be difficult to successfully implement web-based interventions within an organization as employees prefer to separate health matters and their workplace [32]. Hence, the issue around mental health and stigma, especially at work, may be strongly influenced by the organizational culture that influences the use of mental health interventions [45].

The relationship between dropout and acceptability requires further assessment to interpret the current evidence. Although dropout for web-based workplace interventions was high (the mean percentage score of the included studies was 50.9%), explorations of the reasons for this were limited. Indeed, studies have outlined that dropout rates might not be the result of disinterest in occupational web-based interventions for mental health issues but appear to be generally high in computerized interventions [53], suggesting that these interventions are not as engaging as guided or face-to-face sessions and people might not feel committed enough to complete the treatment or program. Consequently, web-based interventions should be tailored and made as interactive and attractive as possible by using animation tools, pictures, and videos, as well as made as short and simple as possible to increase engagement and decrease the likelihood of technical issues [12]. Furthermore, the findings of this study suggest that, before applying interventions in organizations, people’s needs, the environment, and the culture should be assessed; the interventions should be tailored accordingly; and awareness of the benefits and understanding of the use should be addressed.

**Strengths and Limitations**

The generalizability of the findings across workplaces may be limited because of the diversity of individual workplaces; for example, their organizational culture and stigma or attitudes toward mental health. In addition, assessing for confounding variables, including recruitment, setting, intervention characteristics, and country of conduct, did not reveal significant information. However, most of the included studies were conducted in Western countries and used CBT-based interventions, which may further limit generalizability.

Assessing acceptability using indirect measures may be flawed as there could be multiple reasons for employees to stop the intervention. Specifically, dropping out of interventions could be the result of feeling rehabilitated and seeing no further benefit of using the intervention. Nevertheless, dropouts provide great insight into the acceptability of interventions, but more in-depth analyses of the reasons for dropping out should be conducted.

Analysis of the specific assessment of acceptability of occupational web-based interventions was limited because of the heterogeneity of the study designs, intervention types, sample characteristics, and conditions under which the interventions were provided to employees. The studies used data assessment techniques, including cross-sectional self-report methods, whereas the qualitative studies used small samples. Data collection and analysis biases may be observed based on the role of the researchers [54]. As qualitative acceptability results were generally higher compared with indirect measures, this further raises the question of the role of researcher bias. In addition, limitations regarding the consistent and objective measurement of acceptability in the wider literature prevent robust conclusions from being drawn. However, the inclusion and critical appraisal of qualitative studies may have added depth to the factors within the acceptability capture in this study [55].

Despite these limitations, this study offers a comprehensive insight into multiple forms of acceptability measures [56]. Using both qualitative and quantitative as well as direct and indirect measures of acceptability provided a deeper insight into the options for assessing the acceptability of interventions in general. Although this study focused on the workplace, it examined the acceptability of web-based interventions that could be applied more generally to support people’s mental health. For example, the findings could support the implementation of interventions outside of the workplace (eg, as part of clinical mental health treatments). These results might help clinicians, developers, researchers, and the health technology industry create effective and engaging tools in the future.

**Implications**

In relation to workplace practice, before applying interventions, it would be beneficial to increase people’s knowledge of web-based interventions as well as assess their needs in general to improve their attitude toward interventions [13,34]. This is supported by Murray et al [57], whose study found that participants who rejected computerized treatments had significantly lower expectations of the usefulness of self-help
and had general concerns, anxiety, and misunderstandings about computerized treatments. Hence, acceptability may be increased by identifying and correcting misperceptions before participation. Similarly, tailoring interventions to the environment and employees’ needs could increase their general interest and willingness to use them [13]. In other words, web-based interventions for employees should be adapted to the specific environment applied as well as to the users’ needs to increase engagement and acceptability levels. Generally, the acceptability of interventions might increase if employees and organizations are made aware of evidence-based web-based interventions that have multiple practical benefits and the potential to increase individuals’ mental health and well-being in the long run. Finally, the ability of web-based interventions to engage and retain users is critical for ensuring reduced dropout and increased acceptability.

Regarding future research, the results of acceptability studies could be influenced by the general stigma on mental health topics and interventions. Therefore, future research should incorporate acceptability measures of mental health issues into their analysis to assess for confounding variables. Second, regarding quantitative data on acceptability, it would be beneficial if future research included a more in-depth analysis of the reasons for dropout or attrition rates. Third, future research should also address the conceptual and methodological limitations of the research in the field. If there were more organizations using mental health interventions from various settings, the research analysis could be more homogeneous. Organizations might lack the knowledge on how to apply personal health support but could provide their employees with interventions that range from broader aspects of stress management to specified apps that tackle specific mental health issues (eg, depression). Finally, this research was conducted before the COVID-19 pandemic, which changed work styles and environments and affected how people sought and received mental health support. Further research should analyze changes in acceptability as a result of the pandemic to examine shifts in use and acceptability of mental health interventions both within and outside of the workplace.

Conclusions
This study assessed the area of acceptability of web-based workplace interventions for mental health. In general, workers are open to web-based mental health interventions. However, qualitative and quantitative studies suggested varying levels of acceptability, raising the possibility of bias. The importance of stigma, organizational culture, and the implementation of the intervention were highlighted, the latter relating to the engaging design and quality of the intervention as well as the approach to delivery in the workplace itself. Several factors were identified that need to be considered to ensure the effective implementation of web-based interventions in the workplace, some aspects of which may also apply to the general use in supporting people’s mental health. Interventions should be tailored to the respective individual needs and cultural context, use nonstigmatized language, and be made interactive and easy to use. It is also recommended to foster an understanding of the potential value of an intervention to increase its acceptability. Methodological limitations were highlighted to guide the cautious interpretation and generalization of early evidence in this area along with the need to improve methodological rigor in emerging research.

Acknowledgments
The author would like to thank their supervisor, Chris Attoe, for his inspiring feedback and guidance. The author would also like to express gratitude to Dr Joshua Harwood for his invaluable advice.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Quality rating.
[ PNG File , 245 KB - mental_v9i5e34655_app1.png ]

Multimedia Appendix 2
Direct measures.
[ DOCX File , 36 KB - mental_v9i5e34655_app2.docx ]

References


44. Feicht T, Wittmann M, Jose G, Mock A, von Hirschhausen E, Esch T. Evaluation of a seven-week Web-based happiness training to improve psychological well-being, reduce stress, and enhance mindfulness and flourishing: a randomized


Abbreviations

CASP: Critical Appraisal Skills Programme
CBT: cognitive behavioral therapy
ENTREQ: Enhancing the Transparency in Reporting the Synthesis of Qualitative Research
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edits by J Torous; submitted 04.11.21; peer-reviewed by A Richert-Kaźmierska, O Chelidoni; comments to author 08.12.21; revised version received 31.01.22; accepted 08.02.22; published 11.05.22.

Please cite as:
Scheutzow J, Attoe C, Harwood J
Acceptability of Web-Based Mental Health Interventions in the Workplace: Systematic Review
JMIR Ment Health 2022;9(5):e34655
URL: https://mental.jmir.org/2022/5/e34655
doi:10.2196/34655
PMID:35544305
Abstract

Background: The COVID-19 pandemic has shifted mental health care delivery to digital platforms, videoconferencing, and other mobile communications. However, existing reviews of digital health interventions are narrow in scope and focus on a limited number of mental health conditions.

Objective: To address this gap, we conducted a comprehensive systematic meta-review of the literature to assess the state of digital health interventions for the treatment of mental health conditions.

Methods: We searched MEDLINE for secondary literature published between 2010 and 2021 on the use, efficacy, and appropriateness of digital health interventions for the delivery of mental health care.

Results: Of the 3022 records identified, 466 proceeded to full-text review and 304 met the criteria for inclusion in this study. A majority (52%) of research involved the treatment of substance use disorders, 29% focused on mood, anxiety, and traumatic stress disorders, and >5% for each remaining mental health conditions. Synchronous and asynchronous communication, computerized therapy, and cognitive training appear to be effective but require further examination in understudied mental health conditions. Similarly, virtual reality, mobile apps, social media platforms, and web-based forums are novel technologies that have the potential to improve mental health but require higher quality evidence.

Conclusions: Digital health interventions offer promise in the treatment of mental health conditions. In the context of the COVID-19 pandemic, digital health interventions provide a safer alternative to face-to-face treatment. However, further research on the applications of digital interventions in understudied mental health conditions is needed. Additionally, evidence is needed on the effectiveness and appropriateness of digital health tools for patients who are marginalized and may lack access to digital health interventions.

(JMIR Ment Health 2022;9(5):e35159) doi:10.2196/35159

KEYWORDS
digital health; telepsychology; computer-assisted therapy; online therapy; mobile applications; mobile apps; telemedicine; telepsychiatry; virtual reality exposure therapy; mental health; COVID-19
Introduction

Patients with mental health conditions often experience long-term disability, resulting from challenges in accessing mental health services, including low treatment availability and long wait times [1]. Moreover, the COVID-19 pandemic has exposed crucial gaps in mental health care systems, which significantly impact the well-being of many people globally [2-4]. Increased fears of contracting SARS-CoV-2, the burden of quarantine requirements, social distancing, social isolation, rising economic inequities, unemployment, and new workplace requirements are additional stressors brought on by the pandemic, which can exacerbate the symptoms of mental health conditions [5-14]. The pandemic is thought to account for recent increases in mood, anxiety, trauma, and substance use disorders [10-16]. Similar trends in mental illness were observed during the 2003 severe acute respiratory syndrome outbreak, other previous pandemics [10-12,17,18], and recent economic crises [10-12,17,18]. The rise in mental health issues due to the COVID-19 pandemic creates substantial pressures on an already strained mental health care system [12,19], with evidence pointing to a silent mental health crisis as resources are prioritized for stemming the spread of SARS-CoV-2 infections [12].

Consequently, interest in web-based health service delivery has been growing in recent years. These include synchronous and asynchronous therapist contact via messaging, phone call, and videoconferencing; computer, web-based, and mobile delivery of therapy programs; virtual or augmented reality–based programs; computerized or web-based cognitive training, and web-based peer and social support groups (defined below). The global reach of digital health care potentially extends to billions of people with internet access. Web-based and mobile delivery of therapy programs may save practitioner time owing to efficient and effective delivery of treatments at lower associated cost [20]. Digital health interventions may also offer a way to reduce or avert care interruptions while allowing practitioners to adhere to safe social distancing measures [20]. At the onset of the COVID-19 pandemic, health care providers rapidly transitioned to web-based health care delivery to limit the risk of COVID-19 transmission. However, the state of the evidence on the effectiveness of digital interventions is unclear, and the implications for health outcomes of such a drastic shift to digital health platforms are difficult to predict [12,21-25]. Whether clinicians can provide effective and reliable treatment, perform assessments [26,27], identify ailments and symptoms [28], manage suicidal behaviors [26,28,29], and provide personable, compassionate services [26,30,31] remains uncertain. Furthermore, digital delivery of services may be complicated by the symptomatology of some mental health conditions [26,29], concurrent medical conditions [29], and socioeconomic factors [31-42]. A lack of information, resources, and understanding of complex patient-related factors could negatively affect care delivery and overall patient health.

Mobile apps are increasingly used by the public for the treatment of mood and anxiety disorders, sometimes without professional referral or guidance [13,43,44]. There is also some evidence that web-based forums and resources are increasingly common [45-54]. Similarly, over the past decade, there have been noticeable shifts in the provision of cognitive and behavioral training for developmental disorders and dementia to computer and other web-based platforms [55-64]. There are also significant developments in the application of virtual reality tools in health care settings [65-64]. The need for professional guidance in the use of web-based or mobile services and forums is subject to controversy [68-77], and more evidence is needed on optimal ways to integrate these tools into a comprehensive approach to mental health care.

This review is motivated primarily by questions from health care stakeholders in a Canadian setting, who were required to rapidly shift to digital delivery of mental health services during the COVID-19 pandemic. However, to date, there has been no comprehensive review on the use of digital interventions for the treatment of a representative range of mental health conditions. With the present meta-review, we seek to fill this gap and summarize existing evidence on the use of digital health interventions in mental health care. Our hope is that our review will be used by health care stakeholders to inform their consideration of mental health care options for digital delivery.

Methods

Literature Search

We conducted a review of peer-reviewed literature examining the application of digital health interventions for the treatment of mental health conditions described below. We searched Medline on November 1, 2021, for research published after January 1, 2010. We used Medline filters to restrict retrieved records to meta-analyses, systematic reviews, and other types (narrative and conceptual) of literature reviews. We used broad term definitions to maximize the types of digital health interventions and mental health conditions captured in the search. The search strategy consisted of combinations of Medical Subject Headings (MeSH) words and other keywords including the following: virtual reality; telemedicine; computer-assisted therapy; digital health; videoconferencing; mental health; mental health services; psychotherapy; attention deficit and disruptive behavior disorders; anxiety disorders; trauma and stressor related disorders; mood disorders; bipolar and related disorders; dementia; disruptive, impulse control and conduct disorders; dissociative disorders; feeding and eating disorders; neurodevelopmental disorders; neurotic disorders; pain; personality disorders; schizophrenia spectrum and other psychotic disorders; sleep wake disorders; and substance-related disorders (see search query in Multimedia Appendix 1).

Inclusion and Exclusion Criteria

This review is restricted to other reviews that assessed the use of digital interventions for the treatment of mental health conditions. Studies that did not report on the effectiveness of digital interventions on mental health outcomes or did not outline a study protocol were excluded from this review.

Data Extraction, Analysis, and Quality Assessment

Once records were retrieved and deduplicated, TJP, NS, and AJ conducted title and abstract screening where any disagreements were resolved through consensus. Team members...
then proceeded with mutually exclusive full-text screening to identify articles that qualified for inclusion in the review. As with previous meta-reviews, adherence to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [78] was considered to assess risk of bias in selected studies (maximum score of 1: completely adheres to PRISMA Guidelines). To assess the quality and reliability of research within the field, one reviewer conducted data extraction using a standardized and iterative data extraction form (Multimedia Appendix 2). Extracted data included study details (author, date, and type), participant characteristics (mental health conditions), intervention details (intervention type and effectiveness), number of participants, and controls used (treatment as usual, waitlist, placebo, or not applicable). Quality and bias scores describing the primary literature reported in included studies were averaged and faithfully converted (when necessary) to a consistent 3-point scale (1=low, 2=moderate, and 3=high). Owing to the significant heterogeneity in research approaches and findings, we selected a qualitative and semiquantitative approach to summarize and present research findings.

Organization of Mental Health Conditions

To provide a clearer picture of how digital interventions are used in treating various mental health conditions, we separated mental health conditions based on the parent MeSH terms and Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria. Where we identified dissimilarities or similarities in treatment, we either added a subcategory or combined categories together. We removed attention-deficit/hyperactivity disorder (ADHD) from developmental disorders and added it as a separate category owing to dissimilarities in the treatment and management of this condition versus other developmental disorders. We combined the frequently comorbid anxiety, mood, and trauma disorders owing to similar treatment approaches, effectiveness, and reporting in the literature. Additionally, patients with chronic pain, chronic medical illnesses, and chronic disabilities (shortened to chronic illnesses) often experience mental health issues that are underrecognized, receive little attention within digital health intervention literature [79-83], and have unreliable treatment efficacies [79-83]. Despite the use of similar psychological treatments anxiety, mood, and trauma disorders [84-87], chronic illness treatments also involve acceptance, remediation, music, and virtual reality [79-82,84-86]. We therefore retain chronic illness as a category related to but separate from anxiety, mood, and trauma disorders. Similarly, caregivers are often untrained family members who face significant stress and anxiety in the process of providing care for loved ones. Caregivers also benefit from mental health services such as cognitive behavioral therapy and specific psychoeducation, which overlap with some mental health conditions, but also benefit from peer support, training, and acceptance therapy [70,88-95]. Substance use disorder was included since treatments include therapies based on psychological principles [32,78,96-115], and this disorder is often comorbid with other mental health disorders and is considered a mental health condition by medical associations (eg, Canadian Medical Association, American Medical Association, and World Health Organization) and diagnostic manuals (eg, DSM-5).

This review adheres to PRISMA guidelines [116] (Multimedia Appendix 1).

Results

Included Studies

The PRISMA flowchart of the screening process is presented in Figure 1. We identified 3051 records and used Medline selection tools to exclude primary articles (n=2510), and studies published before 2010 (n=42). Of the remainder, 4 were inaccessible and authors did not respond to copy requests; thus, 466 studies proceeded to full-text review where 159 were excluded for not reporting on intervention effectiveness and 3 were proposals. This selection resulted in 77 meta-analyses, 84 systematic reviews, and 143 literature reviews examining the use of digital health interventions for the treatment of mental health conditions.
**Mental Health Conditions**

A summary of metadata extracted from database searches, curated secondary literature, curated primary literature, and participant numbers is provided in Table 1. Per participant, studies on substance use disorders account for a majority (n=241,377, 52%) of digital mental health research, followed by mood, anxiety, and trauma disorders (136,121, n=29%), and >5% for other mental health conditions (pain: n=24,327, schizophrenia: n=20,500, dementia: n=10,823, feeding and eating: n=10,441, developmental: n=8736, bipolar: n=3573, sleep-wake: n=3333; and ADHD: n=2428). Additionally, limited research has examined the use of digital health to provide psychological support to caregivers of people with dementia and developmental disorders. Lastly, we retrieved no records examining the use of digital health interventions to treat antisocial, avoidant, dependent, histrionic, and narcissistic personality, dissociative identity, paraphilic, and sexual health disorders. To demonstrate how the amount and reliability of research can be estimated from metadata, we correlated elements in Table 1 and report a 4D correlation (P<.05; see Table S2 and 4D illustration in Figure S1 in Multimedia Appendix 1). Overall, this illustrates a significant need for the development and testing of digital interventions for other mental health conditions. Nevertheless, the number of research publications has steadily increased since 2015 (Figure 2)—a trend that will likely continue with greater interest in digital mental health research.
### Table 1. Metadata per mental health condition examining article and participant numbers.

<table>
<thead>
<tr>
<th>Mental health conditions</th>
<th>Total literature(^a), n</th>
<th>Secondary literature(^b), n</th>
<th>Primary literature(^c), n</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>90</td>
<td>8</td>
<td>35</td>
<td>2428 (0.5)</td>
</tr>
<tr>
<td>Anxiety, mood, stress, trauma</td>
<td>1205</td>
<td>123</td>
<td>923</td>
<td>136,121 (29.5)</td>
</tr>
<tr>
<td>Bipolar and related disorders</td>
<td>65</td>
<td>9</td>
<td>42</td>
<td>3573 (0.8)</td>
</tr>
<tr>
<td>Dementia</td>
<td>246</td>
<td>24</td>
<td>180</td>
<td>10,823 (2.3)</td>
</tr>
<tr>
<td>Developmental disorders (excluding attention-deficit/hyperactivity disorder)</td>
<td>326</td>
<td>24</td>
<td>349</td>
<td>8736 (1.9)</td>
</tr>
<tr>
<td>Feeding and eating disorders</td>
<td>154</td>
<td>23</td>
<td>117</td>
<td>10,441 (2.3)</td>
</tr>
<tr>
<td>Pain</td>
<td>147</td>
<td>23</td>
<td>348</td>
<td>24,327 (5.3)</td>
</tr>
<tr>
<td>Schizophrenia and psychotic disorders</td>
<td>263</td>
<td>30</td>
<td>304</td>
<td>20,500 (4.4)</td>
</tr>
<tr>
<td>Sleep-wake disorders</td>
<td>145</td>
<td>8</td>
<td>29</td>
<td>3333 (0.7)</td>
</tr>
<tr>
<td>Substance-related disorders</td>
<td>555</td>
<td>59</td>
<td>466</td>
<td>241,377 (52.3)</td>
</tr>
</tbody>
</table>

\(^a\)Total number of articles from Medline searches.  
\(^b\)Selected secondary literature.  
\(^c\)Primary literature curated by secondary sources.

### Figure 2. Number of included articles per year.

Digital Health Interventions

To more precisely measure the amount of research available to treat specific mental health conditions using a specific digital health intervention, we superimposed the primary digital health interventions on study characteristics and conclusions drawn from our search results (Table 2). All digital health interventions are supplementary to synchronous or real-time communication.
**Table 2.** Digital health interventions used to treat mental health conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Therapist contact</th>
<th>Online peer support</th>
<th>Web-based therapy</th>
<th>Mobile therapy</th>
<th>Virtual reality</th>
<th>Cognitive training</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention-deficit/hyperactivity disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45/1</td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>45/1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>363/4</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>968/6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>929/14</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>45/4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2/1</td>
<td>36/5</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>Not reported</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>Medium</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Emerging</td>
<td>Inconclusive</td>
</tr>
<tr>
<td>Effective as assessment modality?</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Emerging</td>
<td>[55,56,117-119]</td>
</tr>
<tr>
<td><strong>Anxiety-, mood-, stress-, and trauma-related disorders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>7156/60</td>
<td>—</td>
<td>19,803/105</td>
<td>1333/6</td>
<td>2842/65</td>
<td>42/3</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>5460/38</td>
<td>73/1</td>
<td>51,074/279</td>
<td>3905/22</td>
<td>2974/69</td>
<td>222/6</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>1567/41</td>
<td>—</td>
<td>31,461/93</td>
<td>45/1</td>
<td>305/21</td>
<td>—</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>2.33 (SD 0.57)</td>
<td>Not reported</td>
<td>2.38 (SD 0.68)</td>
<td>1.64 (SD 0.64)</td>
<td>1.78 (SD 0.68)</td>
<td>1.00</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>Emerging</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Emerging</td>
</tr>
<tr>
<td><strong>Bipolar and Related Disorders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>—</td>
<td>—</td>
<td>992/14</td>
<td>132/1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>14/1</td>
<td>286/3</td>
<td>1499/7</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>—/3</td>
<td>156/1</td>
<td>273/5</td>
<td>51/1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>Not reported</td>
<td>1.00</td>
<td>1.00</td>
<td>2.75</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>Emerging</td>
<td>No</td>
<td>Emerging</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

[1-117,117-118,121,122]

[56,120,121]

[55,56,117-119]
<table>
<thead>
<tr>
<th>Condition</th>
<th>Therapist contact</th>
<th>Online peer support</th>
<th>Web-based therapy</th>
<th>Mobile therapy</th>
<th>Virtual reality</th>
<th>Cognitive training</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dementia and Related Disorders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dementia and Related Disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>483/6</td>
<td>—</td>
<td>—</td>
<td>3981/7</td>
<td>331/11</td>
<td>590/16</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>486/4</td>
<td>—</td>
<td>—</td>
<td>—/30</td>
<td>30/1</td>
<td>1222/19</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>1695/29</td>
<td>—</td>
<td>—</td>
<td>—/30</td>
<td>273/10</td>
<td>282/16</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>1.75 (SD 0.75)</td>
<td>—</td>
<td>—</td>
<td>Not reported</td>
<td>1.50 (SD 0.50)</td>
<td>1.63 (SD 0.41)</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>Medium</td>
<td>—</td>
<td>—</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>No conclusive</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Dementia: caregiver support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dementia: caregiver support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>773/8</td>
<td>11/1</td>
<td>1054/10</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>1019/10</td>
<td>384/4</td>
<td>2852/17</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>78/3</td>
<td>54/2</td>
<td>176/6</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>1.50</td>
<td>1.50</td>
<td>1.50</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>Emerging</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Developmental disorders (excluding attention-deficit/hyperactivity disorder)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental disorders (excluding attention-deficit/hyperactivity disorder)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>535/13</td>
<td>—</td>
<td>—</td>
<td>107/4</td>
<td>222/7</td>
<td>984/20</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>327/10</td>
<td>—</td>
<td>80/3</td>
<td>69/3</td>
<td>877/24</td>
<td>1298/55</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>3330/88</td>
<td>—</td>
<td>7/1</td>
<td>7/2</td>
<td>212/24</td>
<td>392/37</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>1.56 (SD 0.77)</td>
<td>—</td>
<td>1.00</td>
<td>1.50</td>
<td>2.50 (SD 0.50)</td>
<td>1.00 (SD 0)</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>High</td>
<td>—</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>—</td>
<td>Inconclusive</td>
<td>Inconclusive</td>
<td>Inconclusive</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Feeding and eating disorders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeding and eating disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>833/15</td>
<td>—</td>
<td>2497/23</td>
<td>276/3</td>
<td>628/9</td>
<td>—</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>97/1</td>
<td>275/3</td>
<td>3361/29</td>
<td>143/3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Condition</td>
<td>Therapist contact</td>
<td>Online peer support</td>
<td>Web-based therapy</td>
<td>Mobile therapy</td>
<td>Virtual reality</td>
<td>Cognitive training</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>296/8</td>
<td>—</td>
<td>1928/18</td>
<td>107/5</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>2.00</td>
<td>2.00</td>
<td>1.83 (SD 0.69)</td>
<td>1.00</td>
<td>1.00</td>
<td>—</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>—</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>Emerging</td>
<td>Yes</td>
<td>Emerging</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Chronic pain, disability, and other medical illness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>4350/54</td>
<td>—</td>
<td>1339/10</td>
<td>0</td>
<td>3583/43</td>
<td>—</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>3203/46</td>
<td>—</td>
<td>5666/37</td>
<td>—</td>
<td>2642/35</td>
<td>—</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>2066/66</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1478/56</td>
<td>—</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>1.42 (SD 0.55)</td>
<td>—</td>
<td>2.0 (SD 0.71)</td>
<td>—</td>
<td>1.21 (SD 0.37)</td>
<td>—</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>High</td>
<td>—</td>
<td>High</td>
<td>—</td>
<td>High</td>
<td>—</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Schizophrenia and psychotic disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>—</td>
<td>—</td>
<td>/4</td>
<td>—</td>
<td>9</td>
<td>1580/6</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>3287/5</td>
<td>—</td>
<td>/2</td>
<td>101/3</td>
<td>5837/4</td>
<td>1735/28</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>404/5</td>
<td>—</td>
<td>/2</td>
<td>20/2</td>
<td>1891/23</td>
<td>1267/32</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>1.50 (SD 0.87)</td>
<td>1.67 (SD 0.94)</td>
<td>2.00 (SD 1.00)</td>
<td>1.83 (SD 0.62)</td>
<td>2.25 (SD 1.0)</td>
<td>2.50 (SD 0.41)</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Inconclusive</td>
<td>Yes</td>
</tr>
<tr>
<td>Sleep-wake disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>—</td>
<td>—</td>
<td>1779/9</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>—</td>
<td>—</td>
<td>1220/13</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>—</td>
<td>—</td>
<td>334/6</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>—</td>
<td>—</td>
<td>2.00 (SD 0.63)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Digital health interventions can be separated into 7 primary categories. **Synchronous and Asynchronous Therapist Contact**

Synchronous contact refers to methods where providers and patients communicate at the same time (eg, phone call and

<table>
<thead>
<tr>
<th>Condition</th>
<th>Therapist contact</th>
<th>Online peer support</th>
<th>Web-based therapy</th>
<th>Mobile therapy</th>
<th>Virtual reality</th>
<th>Cognitive training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall strength of evidence</td>
<td>—</td>
<td>—</td>
<td>Medium</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>—</td>
<td>—</td>
<td>No</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Effective as assessment modality?</td>
<td>—</td>
<td>—</td>
<td>No studies</td>
<td>[182,322-328]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Substance use disorders</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>8151/21</td>
<td>—</td>
<td>61,896/93</td>
<td>4650/8</td>
<td>11/1</td>
<td>—</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>15,610/31</td>
<td>—</td>
<td>97,802/180</td>
<td>12,385/22</td>
<td>219/5</td>
<td>—</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>984/5</td>
<td>—</td>
<td>14,603/35</td>
<td>5231/30</td>
<td>181/8</td>
<td>—</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>2.13 (SD 0.74)</td>
<td>—</td>
<td>2.07 (SD 0.63)</td>
<td>2.26 (SD 0.74)</td>
<td>1.33 (SD 0.47)</td>
<td>—</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
<td>Inconclusive</td>
<td>—</td>
</tr>
<tr>
<td>Effective as assessment modality?</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>Emerging</td>
<td>No studies</td>
<td>—</td>
</tr>
<tr>
<td>Substance use disorders</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RCT-TAU (patients/studies), n/n</td>
<td>22,326/178</td>
<td>11/5</td>
<td>89,360/273</td>
<td>11,927/34</td>
<td>7909/140</td>
<td>3474/76</td>
</tr>
<tr>
<td>RTC-other (patients/studies), n/n</td>
<td>30,471/152</td>
<td>1018/13</td>
<td>163,655/569</td>
<td>22,521/63</td>
<td>8477/162</td>
<td>5494/125</td>
</tr>
<tr>
<td>Observational (patients/studies), n/n</td>
<td>10,465/252</td>
<td>210/5</td>
<td>48,802/166</td>
<td>7332/92</td>
<td>3718/152</td>
<td>796/63</td>
</tr>
<tr>
<td>Reported study quality (1=low to 3=high)</td>
<td>1.69</td>
<td>1.54</td>
<td>1.86</td>
<td>1.83</td>
<td>1.64</td>
<td>1.45</td>
</tr>
<tr>
<td>Overall strength of evidence</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Effective as treatment modality?</td>
<td>Yes</td>
<td>Emerging</td>
<td>Inconclusive</td>
<td>Inconclusive</td>
<td>Inconclusive</td>
<td>Yes</td>
</tr>
<tr>
<td>Effective as assessment modality?</td>
<td>Yes</td>
<td>No studies</td>
<td>Yes</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Yes</td>
</tr>
</tbody>
</table>

---

**a**RCT-TAU control: Randomized controlled trials with a treatment-as-usual control.

**b**—: not determined.

**c**RCT-other control: Randomized controlled trials with a waitlist or placebo control.

**d**High confidence based on >30 randomized controlled trials with >2000 participants in total; Medium confidence owing to <30 randomized controlled trials with <2000 participants; Low confidence owing to <500 participants; N/A: not applicable.

**e**Yes=positive treatment outcomes and low drop-out rates; Inconclusive=mixed findings, may be effective; Emerging=novel area of research with insufficient evidence; No=no significant difference in outcomes between intervention and controls.

**f**Web-based programs developed for bipolar disorders only address depression symptoms but not mania symptoms.

**g**Patients with schizophrenia and psychotic disorders or symptoms may not be willing to use any digital modalities owing to paranoia about technology, which stems from the underlying psychopathology.
With better technology practitioners have gravitated toward videoconferencing, but a telephone call is used in the event of technical issues [232,360,361]. Delivery of assessment or treatment (eg, prescribing medication, parent and caregiver training, and various therapies) are usually provided using synchronous forms of communication. Nonetheless, it may also be more difficult to deliver time-dependent neurological tests [238]. Synchronous contact remains the primary form of treatment where other forms of treatment described below are only supplementary [68,69,84,98,109,111,112,124,135,151,159,169,170,178,179,199,273,348,349,362].

From a patient’s perspective, most felt that synchronous contact with a therapist afforded greater accessibility, independence, and made it easier for them to express themselves, others felt that it was impersonal [231,363], and some patients with schizophrenia and psychosis disorders were not comfortable with the technology, felt monitored or recorded, and refused care [147,232,308,360,361].

For asynchronous communications, there is a time delay between responses (eg, email and text). These methods can be useful in encouraging patients to attend their appointments, take their medications, exercise, relax, complete daily life tasks, and reduce relapse following remission [74,110,157,198,232,274,279,330,361,364-367]. However, asynchronous forms of communication were not as effective as synchronous forms of communication before remission [123,129,142,147,148]. Furthermore, asynchronous communications are rarely tested in emergency situations with patients who are potentially suicidal or violent [29,125,149,179].

**Web-Based Peer Support**

Mental health support provided by people with lived experience of mental health issues took place via web-based discussion groups (video calls, forums, and social media) where patients with similar disorders can interact. Treatment programs rarely include web-based groups, and few studies explore their role in treatment and adherence; therefore, we retained it as a separate category. Online communities and forums (eg, specific subreddits, forums, discord, and Facebook groups) are prevalent for all mental health conditions since patients can learn more about others’ experiences, learn about their condition, receive peer support, and accept their condition [87,232,304,305]. Online communities have also formed on YouTube where people living with mental health conditions are able to share their lived experience and insights. While there are examples of evidence-based forums and media content on mental health conditions, web-based content is not usually moderated, which may lead to the spread of misinformation. Indeed, unmoderated online communities have lower retention rates [306], suggesting that moderation by a practitioner may be required to reduce potential problems. Nevertheless, patient involvement and interaction on social media platforms provide significant insights, alternative perspectives, and fortitude to the general population, other patients, health care providers, and researchers. Despite their prominence, their use and effectiveness are rarely evaluated.

**Web-Based or Computer-Based Therapy Programs**

Various types of content delivered on the internet included psychoeducation, self-help therapy, journaling, assessments, topics traditionally covered in workbooks and paper format, reminders to take medication, motivational interventions, and web-based peer support. Web-based and mobile programs delivered with administrative or therapist guidance are as effective as treatment as usual (TAU), while those without guidance show significantly lower effectiveness and variable dropout rates [68,69,84,98,109,111,112,124,135,151,159,169,170,178,179,199,273,308,348,349,362]. These are well developed for substance use-, mood-, anxiety-, and trauma-related disorders but not bipolar, personality, and sleep-wake disorders (Table 2). Indeed, for the latter disorders these interventions yield mixed results since they primarily treat anxiety and mood symptoms, but not mania or other symptoms [233,323,326].

**Mobile-Based Therapy Programs**

Mobile apps are a novel way to deliver therapy programs on mobile devices and share similarities to web-based or computer-based therapy programs. Over 2200 mobile apps claim to deliver therapy for several mental health conditions but lack rigorous validation, are not necessarily based on therapeutic principles, are gamified and addictive, or harm recovery [1,43,74,154,180,185,187,194-196,198,200,231,237,280,308,333,356,368]. Furthermore, 38% of trials for mobile apps were uncontrolled (Table 2). Mobile apps were therefore separated from web-based and computerized therapy (Table 2). We also urge caution when selecting mobile apps and provide a list of web-based tools and apps that have previously been validated (Table S1 in Multimedia Appendix 1).

**Virtual and Augmented Reality**

Virtual and augmented reality provide realistic and immersive experiences with a sense of presence for participants. It is a promising tool for new forms of assessment, treatment, and research to understand psychological processes (eg, psychosis and paranoia) [310]. Virtual reality is easier to implement, perform, and more realistic, motivating, and enjoyable than traditional exposure therapy [65-67,98,115,134,183,205,206,208-210,212-215,219,222,311,369]. Virtual reality can be used to deliver psychotherapy, education, cognitive therapy, and exposure therapy [65-67,98,115,134,183,205,206,208-210,212-215,217,219,222,248,257,292,310,311,369,370]. Experiential cognitive therapy, a combination of virtual and cognitive therapy, has also been successful in treating eating- and weight-related disorders [284,371]. Lastly, virtual reality is valuable as a distraction tool, which leads to reduced pain perception, improved functional ability, and lower stress in patients with various acute and chronic illnesses [81,83,290,291,294,295]. Virtual reality could also provide otherwise inaccessible experiences to individuals with a disability, older individuals, or those living with a chronic illness or disability. Nevertheless, virtual reality should be part of a comprehensive treatment strategy [67].

https://mental.jmir.org/2022/5/e35159
Initial concerns that virtual reality could induce nausea, headaches, and other negative side effects, which could ultimately worsen phobias and attrition [207] have been assuaged by several improvements in the technology [67,206]. Practitioners should nevertheless use caution and test participants for susceptibility to motion sickness [218]. Some of these concerns may be addressed by using augmented reality where 3D representations of elements are imposed on the user’s native world, but more research is necessary for conclusive evidence of treatment efficacy between virtual reality and augmented reality [184]. Therapists also need to carefully assess for signs of cognitive avoidance in patients where they might treat virtual environment and stimuli as a “game” instead of cognitive immersion [218]. Mobile-based virtual reality treatments may provide new treatment avenues for patients who cannot attend in-person therapy owing to disability, transportation, or health concerns [137,220].

**Cognitive Training**

Cognitive training includes training exercises, neurofeedback, and games provided over mobile, web-based, or computer devices or virtual reality. These provide greater flexibility and development than pen-and-paper methods. Evidence suggests broad cognitive training is more effective than a narrow focus on a single cognitive modality [61,244,252,316,318]. Additionally, these must also be combined with tailored remediation to extract the greatest benefits in everyday life [58,92,244,252,258,316-318,372]. Cognitive declines are also reported in anxiety, mood, bipolar, and personality disorders, where similarly broad cognitive training could be useful to alleviate cognitive decline, reduce premature brain aging [59-62,64,150,251,252], and increase remission [166,230], and where the success of cognitive training in disorders such as schizophrenia, ADHD, developmental disorders, and dementia could be applied. Cognitive training and virtual reality could also improve broad motor and cognitive functions in patients with neurological disorders such as stroke, traumatic brain injury, Parkinson disease, and multiple sclerosis [61,247]. Attention bias modification appears to be successful in treating negative cognitive and attentional biases in patients with mood and anxiety disorders [163].

**Other Technologies**

Monitoring technologies (eg, breathalyzer, pill dispenser Wisepill, mobile apps, smart watches) are used to regularly monitor psychological symptoms, heart rate, blood pressure, location, and sleep and to alert practitioners to early signs of relapse, missed doses, or to flag early warning signs of disease [60,232,236,243,251,302,305,306,333,334]. Security systems, call screening technology (for scams), and chatbots can also improve quality of life, but more research is needed [60,243,245,251]. Lastly, transcranial direct current stimulation (tDCS) and similar treatments can be delivered remotely for dementia and schizophrenia [244].

**Discussion**

**Principal Findings**

This review found that a majority of studies on digital health interventions are focused on substance use–, anxiety-, mood-, and trauma-related disorders. For patients with these conditions, the greater flexibility, comfort, and routine associated with digital health offered a favorable substitute for in-person visits and retained therapeutic utility. Given this finding, we expect the use of digital health interventions to persist during and after the pandemic owing to the relaxation of insurance and administrative regulations [12,373-377]. The volume and quality of research for these disorders has enabled the discovery of new treatment methods and the refinement of existing digital health tools to improve treatment efficacy.

We also found that the sudden onset of the COVID-19 pandemic led to a rapid shift toward the use of new technology and interventions without the necessary time to train or prepare practitioners and posed challenges for many health care providers. To remedy this, governments, professional organizations, and academics, have created region-specific digital health toolkits [12,378-383] to facilitate and encourage the provision of digital health services. These toolkits are extensive and provide examples of ways in which digital health can be delivered in a meaningful and effective way.

Evidence from this review also suggests that digital health interventions have implications for combatting the dual public health emergencies across North America: the COVID-19 pandemic and the ongoing overdose crisis [11,384]. Findings indicate that there is significant potential for digital health interventions in reducing the harms experienced by people who use substances [32,75,78,96-114,329-339,346-353,385]. Research into digital health interventions for substance use disorders is relatively new and demonstrates the promising use of web-based programs and social media to reach participants instead of relying solely on referrals from practitioners [100,102,111-114,171,173,349,385]. These interventions may offer timely and cost-effective solutions, where texting, moderated forums, validated web-based or computer-based programs, or mobile apps may be used for treatment, psychoeducation, managing ongoing symptoms, and preventing relapse [12,20,25,30,79,80,85,147,157,175,351] (see Table S1 in Multimedia Appendix 1 for a list of validated tools). Nevertheless, it is important to acknowledge that there are certain instances where in-person contact with a service provider is most suitable. This is particularly important given that many homeless and street-involved populations lack access to and knowledge of technology [386-388].

Similarly, this review found indications that web-based programs in anxiety-, mood-, and trauma-related disorders are poised for similar expansion. Since anxiety and depression symptoms have risen in the general population during the pandemic [10-16], several interventions can be useful for short-term symptom management, such as synchronous communication (videoconferencing or telephone calls) with a therapist [68,96,123-126,129,130,133-139,141,142] and referral to validated web-based [1,30,68,69,84-86,96,118,124,351].
While this review also identified promising developments in digital programs for ADHD, developmental, dementia, eating, schizophrenia, and chronic illness, we found that digital health interventions for these conditions are nascent. Negative findings in sleep-wake and bipolar disorders suggest that significant retooling is necessary for treating these conditions. Furthermore, no reviews on the use of digital health tools for dissociative, elimination, sexual, and personality disorders were identified. The positive outcomes reported for digital health interventions in a wide range of mental health conditions suggest that there may be merit to exploring these interventions in additional clinical contexts during and after the COVID-19 pandemic. Caution is also warranted with patients with schizophrenia, psychosis, or bipolar disorder as technology may be triggering or could exacerbate existing symptoms [147,232,360,361].

Review findings also suggest that synchronous digital contact is an effective substitution for in-person treatment and assessment for many mental health conditions. Considering successes in most mental health conditions, these findings can be generalized to other conditions where less research is available, such as bipolar, sleep-wake–related, and personality disorders. While some health care providers have expressed concerns regarding their ability to build a therapeutic alliance with their patients, research shows that this is not significantly affected by synchronous communication [26,389,390]. Interestingly, synchronous digital health may be beneficial for autism spectrum disorders [269] and social anxiety since it reduces social interaction–related stress, need for eye contact, oversensitivity, and overstimulation. Evidence from this review indicates that synchronous platforms are associated with significant cost and time savings. First, this transition is also beneficial by reducing commutes to work, the ability to organize one’s working day and tasks [391-393], and protects therapists from the risk of physical confrontations [394,395].

Digital health tools have also been found to allow practitioners to reduce the time they spend with each patient, where evidence suggests that spending 10 minutes with patients through synchronous platforms, and providing referrals to asynchronous platforms (eg, web-based, mobile-, or computer-based therapy and cognitive training) is sufficient [30,69,73,76,77,155,157,252,308]. Some patients (eg, children and elderly) may face other barriers to using or accessing technology [396,397], which can be resolved by specific training on using the application [59,62], obtaining help from a caregiver, and could even be accomplished through remote desktop applications (such as Microsoft Teams: Remote Desktop Protocol). Nonetheless, transferring this responsibility to a family member increases caregiver burden and may lead to suboptimal results over the long term. However, the proliferation of untested applications (especially mobile apps) raises concerns around the quality of existing platforms [1,74,154,180,185,187,194-196,198,200,231,237,280,333].

More specifically, these applications often lack validation, reliability, and are not always built on sound psychotherapeutic principles [1,74,154,180,185,187,194-196,198,200,231,237,280,308,333].

Digital health interventions are also less effective at mitigating the impacts of social isolation, particularly in the context of the COVID-19 pandemic, where public health orders and the requirement of physical distancing is expected to drastically impact peoples’ mental health. Human connection contributes significantly to one’s mental health; therefore, it is important that digital health interventions maintain their human aspect as this is associated with increased efficacy [68,69,84,98,109,111,112,124,135,151,159,169,170,178,179,199,273,348,349,362]. Findings demonstrate that asynchronous platforms, such as web-based forums, social media, and other digital communities, likely increase patient engagement and adherence to treatment across all mental health conditions [87,232,304,305]. Additionally, preventative education can be disseminated via asynchronous platforms (eg, social media, groups, forums, and schools) for all mental health conditions, as seen in substance use disorders [100,102,111-114,171,339,349]. Owing to increased demand and lack of availability of services during the COVID-19 pandemic, many patients have transitioned to mobile apps and web-based programs without the guidance of a practitioner [13]. Hence, the absence of sufficient research into these venues, their impact on mental health, and the lack of practitioner guidance and support [1,74,154,180,185,187,194-196,198,200,231,237,280,333] raise concerns that these platforms may cause harm. Indeed, government intervention to increase the prominence of validated region-specific tools and resources in web-based and app-related searches may be required.

Another emerging asynchronous technology that can be used for the treatment of mental health conditions are virtual reality tools. Greater accessibility, comfort, and normalcy of the technology will encourage the development of virtual reality interventions on site or at home. Nevertheless, there are also barriers to providing and expanding virtual reality tools. For example, the high cost of equipment acts as a significant barrier, however, lower priced equipment or mobile phones can be used as substitutes [137,220,247]. Additionally, virtual reality tools are based on recent technological advancements, and there is little quality research on the use of industry-standard equipment and even less so for low-cost virtual reality options. Despite these limitations, virtual reality addresses a particular niche of therapeutic tools (eg, exposure therapy) [65-67,98,115,134,183,205,206,208-210,212-215,219,222,311,369] and is an effective tool for pain management [81,83,290,291,294,295], indicating that as technology and research advances, it may become a central component of any comprehensive mental health treatment strategy.

Owing to the social distancing and quarantine requirements posed by the pandemic, patients with mental health disorders already face social isolation in addition to increased stress and anxiety [5-14]. Additionally, patients surviving COVID-19 may experience lingering symptoms and post–intensive care syndrome long after discharge from intensive care units.
Mental health challenges for these patients include anxiety, depression, posttraumatic stress disorder (PTSD), cognitive decline, and chronic illness [398,399]. Along with previously mentioned interventions to deal with symptoms of anxiety, depression, and posttraumatic stress, virtual reality can be used to reduce stress, distract from pain, and retrain functional movement in patients who experience chronic illness after COVID-19.

Health care providers are also at risk of feeling social and professional isolation as well as burnout [26,394,400], which must be properly managed by managers, the professional organization, and practitioners themselves. Given the anticipated impact of the pandemic on the mental health of health care providers [11,18,401-403], health care organizations will benefit from specialized synchronous, web-based, and mobile therapy and moderated discussion forums to alleviate this burden. Similar interventions have been used with family caregivers [70,88-90] and health care providers [18,401,402,404,405] to treat anxiety, depression, PTSD, and burn out. Therefore, such interventions can help manage health care providers’ mental health.

Future Directions

Over the last two decades, research on the use of digital health interventions to deliver mental health care has increased significantly. Lessons learned from highly studied fields (eg, substance use–, anxiety–, mood–, and trauma-related disorders) can guide the implementation of digital health interventions to treat other mental health conditions. Starting at the most basic level, where practitioner guidance for 10 minutes was essential and often sufficient for the treatment of anxiety–, mood–, and trauma-related disorders, web-based, computer-based, or mobile programs or apps developed for these conditions could be adapted, improved upon, and evaluated to treat other conditions with overlapping symptomatology. For example, one could consider the overlap in symptomatology among mood-related, anxiety, bipolar, sleep-wake–related, and some personality disorders [406]. Thus, digital interventions for the former two conditions could be adapted to include journaling, behavioral modification prompts, and other psychotherapeutic treatments akin to these conditions, and finally be re-evaluated. Nonetheless, for conditions where no treatments exist, the development and digitization of novel treatment strategies is required [233,323,326]. Indeed, the digital nature of these programs enables the collection of regular assessment data, input from patients, and evaluation by health care providers to develop decision trees and machine learning algorithms to instantly improve and personalize treatment plans, require less practitioner time, and provide greater flexibility in treatment delivery.

The rapid pace of technological advancements also poses significant challenges. For example, treatment program implementation has evolved from computerized delivery with CDs to web platforms to mobile apps in the last two decades. Significant technological shifts have forced researchers to completely rebuild the programs despite apparent similarities between these modes of delivery. Many validated programs identified (Table S2 in Multimedia Appendix 1) are outpaced by technological advancements and lack recent updates. First, easy-to-use development and cross-platform tools (eg, React Native and Xamarin) will enable researchers to make, evaluate, and maintain programs despite rapid technological advancements. Second, health care policies and evaluation may need to be modified so that validated tools can evolve over time and across platforms when the underlying therapeutic principles remain consistent.

Existing research on digital interventions rarely covers comorbid conditions, emergency situations, or complex socioeconomic factors. For example, research on people experiencing homelessness is limited to commentaries and policy recommendations based on available research in the general population [32-42]. This is of particular concern when considering that those of lower socioeconomic status or with complex life circumstances show reduced benefits from digital health interventions [148,165,407]. Additional research considering individuals experiencing various psychosocial complexities or comorbid conditions is required.

Further research must endeavor to use appropriate controls and more rigorous design to improve overall study quality assessed in Table 2. Blinding patients to the digital nature of the treatment is difficult, but creative solutions (eg, unrelated cognitive tasks in lieu of treatment) are recommended. Additionally, standardized rating scales (ie, DSM-V criteria) should be used instead of nonstandard assessments or a participant’s opinion on the treatment. Most studies are restricted to treatment duration and lack long-term follow-up (>6 months). Considering digitization of treatments and records, practitioners can automatically request follow-up surveys and assessments via email or text. Follow-up surveys must also consider whether patients have pursued other treatment programs, as these could confound any pertinent treatment effects. Lastly, following successful remission, there is limited research on the use of digital health interventions (eg, email, text, social media, and forums) to prevent relapse, which can be accomplished via email or text [74,110,157,198,232,274,279,330,361,364-367].

Web-based peer support is dependent on human interaction, which can be unpredictable and include uncontrolled variables. For example, since any large number of people can participate in forums for intermittent periods of time, the inevitable turnover can cause cultural shifts. This would therefore require moderation by practitioners. Evaluation is further complicated by the lack of objective and quantifiable pre-post measures in open social media groups and forums. Indeed, practitioner moderated forums or groups may fare better and could automatically request participants to fill out monthly surveys. Further research is needed to address these hypotheses.

Limitations

There are several limitations to this review. First, to rapidly inform health care stakeholders responsible for managing treatment of a broad range of mental health disorders, we took a comprehensive approach. As a result, we limited the scope to secondary literature sources and utilized a systematic methodology designed for meta-reviews [167]. Since we primarily report on the effectiveness, feasibility, and reliability of digital delivery in lieu of face-to-face treatment, we did not.

https://mental.jmir.org/2022/5/e35159

JMIR MENTAL HEALTH

Philippe et al

XSL-FO

RenderX
Although digital delivery of mental health treatment has been in clinical use for a long time, the available research on the topic is far from comprehensive or consistent. New guidelines to increase reliability and consistency of reporting, evaluation, and quality and bias assessments would enable faster literature synthesis and increase confidence. Living systematic reviews for bipolar, personality, developmental, dementia, and sleep-wake disorders would also be very useful to guide and organize novel digital treatment strategies.

Overall, digital treatment strategies paired with synchronous practitioner contact are as effective as nondigital alternatives. However, in offering digital treatments, it is essential to consider feasibility of treatment, caregiver burden, patient-specific symptoms (eg, paranoia), and patient-specific parameters. More research is especially needed in marginalized populations who face greater barriers to mental health treatment access. Thus, to maintain treatment quality and efficacy, patients should have the option for face-to-face interventions, despite the challenges posed by the COVID-19 pandemic. Nonetheless, digital treatments offer many benefits such as increased patient engagement, accessibility, and availability paired with reduced practitioner workload. Additionally, the drastic shift to digital health is likely to encourage further developments in treatments for many mental health disorders and expansion into other digital modalities, such as virtual reality, social media, and web-based forums. These developments promise significant advances in mental health treatment via global collaboration and investment.

Acknowledgments
This review was initially conducted to provide comprehensive information about the use of digital health to treat mental health conditions in complex populations for a community clinic and was not funded. Publication of this article was cosponsored by the Journal of Medical Internet Research, the University of British Columbia (Graduate Student Travel and Research Dissemination Fund to TJP), and Simon Fraser University (SFU Central Open Access Fund to KV). We thank the VCHRI Rapid Evidence Synthesis Team and VCH Operations Director, Andy Day, for valuable feedback. All authors discussed and refined the direction and development of this review. TJP, AJ, and MEK carried out research for the Introduction section, which TJP wrote. TJP with guidance from EL and Evan Sterling developed the methodology and search strategy. TJP, NS, and AJ screened articles for inclusion. TJP, NS, and AJ designed the tables and figures and wrote the Results and Discussion sections. AP and KV supervised all stages in equal capacity. All authors approved of and reviewed the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary.
[PDF File (Adobe PDF File), 392 KB - mental_v9i5e35159_app1.pdf]
Multimedia Appendix 2
Extracted data for digital health interventions per mental health condition.
[XLSX File (Microsoft Excel File), 468 KB - mental_v9i5e35159_app2.xlsx ]

References


174. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression? A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and


**Abbreviations**

- ADHD: attention-deficit/hyperactivity disorder
- DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
- MeSH: Medical Subject Headings
- PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- PTSD: post-traumatic stress disorder
- TAU: treatment as usual

https://mental.jmir.org/2022/5/e35159

JMIR Ment Health 2022 | vol. 9 | iss. 5 | e35159 | p.88 (page number not for citation purposes)
The Role of Emotion Regulation and Loss-Related Coping Self-efficacy in an Internet Intervention for Grief: Mediation Analysis

Jeannette Brodbeck¹,², PhD, Prof Dr; Thomas Berger¹, PhD, Prof Dr; Nicola Biesold¹, MSc; Franziska Rockstroh¹,³, MSc; Stefanie J Schmidt¹, PhD, Prof Dr; Hansjoerg Znoj¹, PhD, Prof Dr

¹Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland
²School of Social Work, University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland
³University Hospital of Child and Adolescent Psychiatry and Psychotherapy, Bern, Switzerland

Corresponding Author:
Jeannette Brodbeck, PhD, Prof Dr
Department of Clinical Psychology and Psychotherapy
University of Bern
Fabrikstrasse 8
Bern, 3012
Switzerland
Phone: 41 316843805
Email: jeannette.brodbeck@unibe.ch

Abstract

Background: Internet interventions for mental disorders and psychological problems such as prolonged grief have established their efficacy. However, little is known about how internet interventions work and the mechanisms through which they are linked to the outcomes.

Objective: As a first step in identifying mechanisms of change, this study aimed to examine emotion regulation and loss-related coping self-efficacy as putative mediators in a randomized controlled trial of a guided internet intervention for prolonged grief symptoms after spousal bereavement or separation or divorce.

Methods: The sample comprised older adults who reported prolonged grief or adaptation problems after bereavement, separation, or divorce and sought help from a guided internet intervention. They were recruited mainly via newspaper articles. The outcome variables were grief symptoms assessed using the Texas Revised Inventory of Grief and psychopathology symptoms assessed using the Brief Symptom Inventory. A total of 6 module-related items assessed loss-focused emotion regulation and loss-related coping self-efficacy. In the first step, path models were used to examine emotion regulation and loss-related coping self-efficacy as single mediators for improvements in grief and psychopathology symptoms. Subsequently, exploratory path models with the simultaneous inclusion of emotion regulation and self-efficacy were used to investigate the specificity and relative strength of these variables as parallel mediators.

Results: A total of 100 participants took part in the guided internet intervention. The average age was 51.11 (SD 13.60) years; 80% (80/100) were separated or divorced, 69% (69/100) were female, and 76% (76/100) were of Swiss origin. The internet intervention increased emotion regulation skills (β=.33; P=.001) and loss-related coping self-efficacy (β=.30; P=.002), both of which correlated with improvements in grief and psychopathology symptoms. Path models suggested that emotion regulation and loss-related coping self-efficacy were mediators for improvement in grief. Emotion regulation showed a significant indirect effect (β=.13; P=.009), whereas coping self-efficacy showed a trend (β=.07; P=.06). Both were confirmed as mediators for psychopathology (β=.12, P=.02; β=.10, P=.02, respectively). The path from the intervention to the improvement in grief remained significant when including the mediators (β=.26, P=.004; β=.32, P≤.001, respectively) in contrast to the path from the intervention to improvements in psychopathology (β=.15, P=.13; β=.16, P=.10, respectively).

Conclusions: Emotion regulation and loss-related coping self-efficacy are promising therapeutic targets for optimizing internet interventions for grief. Both should be further examined as transdiagnostic or disorder-specific putative mediators in internet interventions for other disorders.
**Introduction**

**Background**

Several studies and meta-analyses have established the efficacy of traditional face-to-face interventions for grief counseling [1,2]. In addition, guided internet interventions for prolonged grief symptoms have demonstrated their efficacy with effect sizes in the moderate to large range [3-6]. A recent review and meta-analysis concluded that internet interventions for bereaved individuals with higher levels of disturbed grief showed medium effect sizes, suggesting that further research should focus on the moderators and underlying mechanisms of treatments [7]. The components of grief counseling mostly provided in palliative care settings were psychoeducation, enhancing communication and social support, stress reduction/relaxation skills, cognitive reframing, and identifying and modifying maladaptive behaviors [8]. Psychological interventions for severe, prolonged grief focus on cognitive restructuring, exposure, interpersonal elements, and behavioral activation [9]. A component of most internet interventions for grief is expressive writing assignments, which can foster the emotional processing of the loss, may lead to new perspectives on the loss, and might foster the process of sense making [7].

Although considerable evidence has established the efficacy of cognitive behavioral (internet) interventions for psychiatric disorders and several emotional problems, including grief, little is known about how they lead to an improvement in symptoms or behavior [10,11]. The identification of these mechanisms of change would be useful for tailoring interventions that specifically target these mechanisms and thus may be more potent or efficient [12]. It may also contribute to the development of more parsimonious interventions with fewer but equally effective components [13], which reduce the burden for clients as well as save time and cost [14].

Potential mechanisms of change can be specific factors posited in the theoretical background of the intervention (eg, changes in maladaptive thinking or behavior in cognitive behavioral therapy [CBT]) or common factors, such as the therapeutic alliance, empathy, expectations, or a rationale that provides credibility to the intervention [15]. Self-efficacy is a central variable in social cognitive theories [16]. Self-efficacy as a belief in the ability to exercise control over events that affect one’s life to manage one’s personal functioning and environmental demands plays an important role in stress reactions and adaptive coping in threatening situations [17]. Bereavement coping self-efficacy (CSE) predicted lower emotional distress, higher psychological and spiritual well-being, and better physical health in widows whose husbands had died of cancer [18]. Self-efficacy also predicted lower grief symptoms in students who lost a close person in a university campus shooting [19]. In addition, Benight and Bandura [17] concluded that CSE was a mediator in the recovery from traumatic experiences. In line with this notion, a change in CSE predicted a decrease in posttraumatic stress symptoms in an eHealth intervention for survivors of trauma [20]. Moreover, self-efficacy was a mediator between psychopathology symptoms and disabilities in activities and participation [21].

Emotion regulation (ER) has been established as a transdiagnostic risk factor for different psychological disorders [22] and is a central intervention target in psychotherapy [23]. A review concluded that face-to-face ER interventions had positive short- and long-term effects on emotion process outcomes, affect and mood states, and medical and psychiatric disorders [24]. Improvements in the ability to modify, accept, and tolerate negative emotions were consistent predictors of treatment outcomes in patients with various mental disorders [23]. Moreover, CBT enriched with ER training resulted in a greater reduction in depression and negative affect and increased well-being than routine CBT [25].

Although some studies have examined ER training as a predictor or outcome of treatment, few studies have investigated ER as a mechanism of change. For example, the modification of negative emotions was found to mediate the link between ER skills and psychopathological symptoms assessed using the Brief Symptom Inventory [26]. Furthermore, ER was a mediator and putative mechanism of change in an internet intervention for stress management [27].

ER and loss-related CSE can be integrated as putative mechanisms of change in the existing models of coping with grief. The dual process model of coping with bereavement posits that loss-oriented tasks, such as grief work, experiencing the pain of loss, expressing emotions toward the deceased, and transforming bonds with the lost person, are necessary for positive adaptation to the loss [28]. Similarly, the task model of mourning specifies tasks such as accepting the reality of the loss and experiencing the pain of grief [29]. Improvement in ER may be especially important for these loss-related tasks [30,31]. ER skills may make these processes more tolerable by facilitating the modulation of overwhelming or more persistent painful emotions.

Furthermore, the dual process model describes the importance of restoration-oriented tasks such as engaging in new activities and finding new social roles and identities. Restoration-oriented tasks can be perceived as very stressful, and a high level of loss-related CSE and the belief in the ability to achieve these goals may facilitate tackling these tasks and increase the sense of autonomy, self-determination, purpose in life, and perceived environmental mastery, leading to less avoidant behavior and...
less aversive ruminatiuon [18]. High loss-related CSE may decrease the appraisal of these restoration-oriented tasks as threatening, alleviate stress and anxiety, promote engagement in coping behavior, and sustain coping efforts [17]. Thus, loss-related CSE could promote positive adaptation to a life without the partner and decrease loss-related symptoms.

As a third element, the dual process model highlights the importance of oscillating between loss- and restoration-oriented tasks. ER skills may not only foster coping with loss-oriented tasks but also the alternation between loss and restoration orientation by eventually limiting grief work, rumination, and pain and facilitating distraction, soothing, and cheering oneself up.

Objectives
This study examines ER and loss-related CSE as potential mechanisms of change in an internet intervention, called LIVIA, for prolonged grief symptoms after spousal bereavement, separation, or divorce [32,33]. LIVIA addressed older adults who had experienced spousal bereavement, separation, or divorce and sought help for coping with prolonged grief symptoms, psychological distress, or adaptation problems in daily life. Thus, LIVIA is the first intervention that focuses on grief after bereavement, as well as grief after separation or divorce. Both events require similar adaptation and mourning tasks identified by Worden [29]; that is, accepting the reality of the loss, processing the pain of grief, adjusting to a life without the spouse, and remembering the lost spouse while reinvesting emotional energy into a new life. We assume that the dual process model is also applicable for separation or divorce from a spouse insofar as these events, similar to bereavement, imply breaking the bond and necessarily lead to the reorganization of one’s life circumstances. Furthermore, we hypothesize that the effect of the intervention is based on the same mechanisms of change.

The comparison of baseline characteristics and the efficacy of LIVIA for widowed and divorced participants, as well as the stability of the effects over 3 months, have been described elsewhere [33]. LIVIA improved grief, depression symptoms, psychopathological distress, embitterment, loneliness, and life satisfaction compared with the waitlist group. The between-group effect sizes were $d=0.81$ for grief and $d=0.39$ for psychopathology symptoms.

Building on these results, the present post hoc analyses aim to investigate whether gains in ER skills and loss-related CSE mediated the intervention effects as a first step in elucidating mechanisms of change in an intervention for grief after spousal bereavement, separation, or divorce. This study is one of the few to investigate mediators in internet interventions and the first to examine emotional and cognitive processes as mediators in a grief intervention. We hypothesize that both gains in ER skills and loss-related CSE mediated the effect of the intervention on improvements in grief and psychopathology symptoms.

Methods

Recruitment
The data presented in this study were based on a randomized controlled trial that evaluated the efficacy of LIVIA compared with a waitlist control group (ClinicalTrials.gov NCT02900534). Participants were mainly recruited via newspaper articles and web-based self-help forums. The main inclusion criteria were the experience of spousal bereavement, separation, or divorce >6 months before enrolling in the study and seeking help to cope with prolonged grief symptoms, psychological distress, or psychosocial adaptation to a life without a partner. The main exclusion criteria were severe psychological or somatic disorders that needed immediate treatment, acute suicidality (Beck Depression Inventory suicide item >1 or suicidal ideation in the telephone interview), concomitant psychotherapy, and/or prescribed drugs against depression or anxiety if prescription or dosage had changed in the month before or during the internet intervention.

Participants
Of the total sample of 110 individuals, 9 (8.2%) individuals did not start the internet intervention and were excluded from the present analyses. One of the participants was excluded because of being a multivariate outlier, which affected the mediation analyses. Therefore, the analysis sample comprised 100 German-speaking participants who lost their spouse through bereavement (20/100, 20%), separation, or divorce (80/100, 80%) and who were randomly allocated to the intervention group or the waitlist control group. The waitlist control group received access to the treatment after 12 weeks. The participants first provided electronic and then oral informed consent in a telephone screening interview.

Ethics Approval
This study was approved by the Cantonal Ethics Committee of the Canton of Bern, Switzerland (BASEC2016-00180).

Measures
The severity of grief symptoms was assessed using the 16-item Texas Revised Inventory of Grief–German Version (TRIG) [34]. The answer categories ranged from 1=completely true to 5=completely false. Cronbach $\alpha$ was .86 in the preintervention measurement and .90 in the postintervention measurement. The TRIG includes items that are applicable after divorce and bereavement and has proven to have good factorial validity that was temporally invariant over 1 year [35].

Psychopathology symptoms were measured using the German version of the widely used Brief Symptom Inventory [36]. The 53 items assessed a broad range of somatic and psychopathological symptoms within 7 days before completing the questionnaire. Answer categories ranged from 0=not at all to 4=very much. Cronbach alpha was .90 in the preintervention measurement and .96 in the postintervention measurement. A total of 6 module-related items assessed loss-focused ER and loss-related CSE. The response categories ranged from −3=not at all to 3=yes, exactly. A confirmatory factor analysis, including the 6 items, supported a 2-factor model compared with a 1-factor
model (1-factor model: Comparative Fit Index=0.84, Tucker-Lewis Index=0.74, root mean square error of approximation=0.273; 2-factor model: Comparative Fit Index=0.95, Tucker-Lewis Index=0.91, root mean square error of approximation=0.159). The details on the development of these measures and the results of the exploratory and confirmatory factor analyses are presented in Multimedia Appendix 1. Loss-focused ER was assessed with the following module-related items: “I can cheer myself up,” “I can have a positive influence on my thoughts and feelings,” and “I can take care of my own well-being.” Loss-focused CSE contained 3 items: “I am convinced that I can cope with the loss of my spouse/with the separation or the divorce,” “I am ready to do what is necessary to overcome my loss,” “I have a strong influence on the coping with my loss.” Cronbach α for ER was .90 in the preintervention measurement and .91 in the postintervention measurement, and Cronbach α for self-efficacy was .76 in the preintervention measurement and .85 in the postintervention measurement. To measure gains in ER and self-efficacy, we subtracted the presum score from the postsum score. Thus, a positive value indicated a gain during the intervention. All self-report questionnaires were web-based using Qualtrics (QualtricsXM) [37] at baseline (ie, before the intervention) and after the intervention 12 weeks after receiving access to the program.

In addition to the self-report questionnaires, the initial screening process included a telephone call, in which trained email supporters assessed the criteria for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition diagnosis of persistent complex bereavement disorder. This required an adaptation of the criteria to the purpose of our study; that is, we assessed the persistence of the symptoms 6 months after the loss instead of 12 months and also used interviews with individuals who lost their spouse through separation or divorce.

**Statistical Analyses**

As the first step, we computed correlations using the pooled data of the intervention and waitlist control groups, who received access to the intervention after a 12-week waiting period. The pooled data set represented a more comprehensive sample and provided a bigger sample size. In the second step, we computed mediation models for improvements in grief and psychopathology symptoms. We used the model indirect command to specify and estimate the specific indirect effects for both mediators and the total indirect effect. Regarding effect sizes for direct effects, we considered standardized regression coefficients of 0.1 as small, 0.3 as medium, and 0.5 as large [39]. For indirect effects in the mediation models, we considered 0.01 as small, 0.09 as medium, and 0.25 as large effects [40].

As data were missing at random (see the following sections), we used multiple imputation to deal with missing data [41]. Multiple imputation using the Bayes estimator yielded inconsistent estimates depending on the number of iterations. Therefore, we used a robust maximum likelihood estimator to impute missing data in 100 data sets. Sensitivity analyses showed robust findings for analyses with complete cases and imputed data.

**LIVIA Intervention**

The dual process model of coping with bereavement and the task model of mourning provided the theoretical background for a guided internet-based self-help intervention called LIVIA [32,33]. It comprised 10 text-based modules and a weekly email as guidance. The modules contained writing tasks for exposure to loss and assignments for practice in daily life. Several modules directly targeted ER processes: 3 modules focused on cognitive behavioral techniques fostering positive emotions, self-care, and social relationships, whereas 2 modules focused on exposure and loss-oriented interventions (ie, writing tasks for accepting memories and pain as well as addressing unfinished business). Loss-related CSE was a direct target in the modules, including information about grief or separation reactions, coping strategies, and restoration-oriented interventions for creating a life without the partner.

**Results**

**Sample Characteristics**

The mean age of the participants was 51.11 (SD 13.60, range 20-85) years, and 69% (69/100) were female. Most participants were of Swiss origin (76/100, 76%) and went to a vocational school (37/100, 37%) or university (34/100, 34%). The average time since the bereavement, separation, or divorce was 2 (SD 3.0, range 0.5-25) years, and 25% (25/100) of the participants fulfilled the B, C, and D criteria of a persistent complex bereavement disorder in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Table 1 presents the baseline characteristics of the intervention group, the waitlist control group, and the pooled data sets, as well as the results of the comparison between the intervention and the control groups.

On average, participants completed 8 of the 10 modules (mean 8.03, SD 2.79), and 57% (57/100) of participants completed all modules. Completers and individuals who did not fill out the postquestionnaires did not significantly differ in terms of baseline characteristics such as demographics or level of distress (P>.21). However, participants who did not fill out the postquestionnaires completed significantly fewer modules than completers (meanD=4.19, SDD=2.81 vs meaning=8.76, SCD=2.12; tD=7.49; P<.001; d=2.04) and were significantly younger (meanD=43.25, SD=14.58 vs mean=52.63, SCD=12.95; tD=2.60; P<.001; d=0.71). This suggests a missing at random mechanism (ie, that missingness is related to measured variables in the analysis model [42]).
### Table 1. Demographics and sample characteristics at baseline and means of grief and psychopathology at the postmeasurement time point (N=100).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Pooled</th>
<th>Intervention (n=58)</th>
<th>Control (n=42)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>51.11 (13.60)</td>
<td>50.85 (12.90)</td>
<td>51.48 (14.68)</td>
<td>.83</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>69 (69)</td>
<td>43 (74)</td>
<td>26 (62)</td>
<td>.19</td>
</tr>
<tr>
<td>Male</td>
<td>31 (31)</td>
<td>15 (26)</td>
<td>16 (38)</td>
<td>.19</td>
</tr>
<tr>
<td>Event, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spousal bereavement</td>
<td>20 (20)</td>
<td>11 (19)</td>
<td>9 (21)</td>
<td>.76</td>
</tr>
<tr>
<td>Separation or divorce</td>
<td>80 (80)</td>
<td>47 (81)</td>
<td>33 (79)</td>
<td>.76</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compulsory school</td>
<td>2 (2.)</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>.35</td>
</tr>
<tr>
<td>Apprenticeship</td>
<td>19 (19)</td>
<td>14 (24)</td>
<td>5 (12)</td>
<td>.35</td>
</tr>
<tr>
<td>Secondary 2</td>
<td>7 (7)</td>
<td>3 (5)</td>
<td>4 (10)</td>
<td>.35</td>
</tr>
<tr>
<td>Vocational school</td>
<td>37 (37)</td>
<td>19 (33)</td>
<td>18 (43)</td>
<td>.35</td>
</tr>
<tr>
<td>University</td>
<td>34 (34)</td>
<td>21 (36)</td>
<td>13 (31)</td>
<td>.35</td>
</tr>
<tr>
<td>Nationality, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swiss</td>
<td>76 (76)</td>
<td>46 (79)</td>
<td>30 (71)</td>
<td>.66</td>
</tr>
<tr>
<td>German-speaking countries</td>
<td>18 (18)</td>
<td>9 (16)</td>
<td>9 (21)</td>
<td>.66</td>
</tr>
<tr>
<td>Other countries</td>
<td>6 (6)</td>
<td>3 (5)</td>
<td>3 (7)</td>
<td>.66</td>
</tr>
<tr>
<td>Time since event (years), mean (SD)</td>
<td>2.21 (3.0)</td>
<td>2.16 (3.47)</td>
<td>2.27 (2.26)</td>
<td>.85</td>
</tr>
<tr>
<td>Persistent complex bereavement disorder, n (%)</td>
<td>25 (25)</td>
<td>15 (26)</td>
<td>10 (24)</td>
<td>.82</td>
</tr>
<tr>
<td>Grief, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before treatment</td>
<td>3.39 (0.78)</td>
<td>3.48 (0.74)</td>
<td>3.26 (0.82)</td>
<td>.18</td>
</tr>
<tr>
<td>After treatment</td>
<td>2.84 (0.89)</td>
<td>2.80 (0.86)</td>
<td>2.90 (0.93)</td>
<td>.60</td>
</tr>
<tr>
<td>Psychopathology, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before treatment</td>
<td>0.86 (0.57)</td>
<td>0.95 (0.63)</td>
<td>0.73 (0.44)</td>
<td>.06</td>
</tr>
<tr>
<td>After treatment</td>
<td>0.62 (0.50)</td>
<td>0.61 (0.51)</td>
<td>0.64 (0.48)</td>
<td>.81</td>
</tr>
</tbody>
</table>

<sup>a</sup>Comparison between intervention and control groups; t tests were 2-tailed.

### Correlations Between Gains in ER and CSE, Outcomes, and Demographics

Table 2 presents the correlation matrix of predictors and outcome variables based on the pooled data set.

Baseline levels of grief and psychopathology symptoms correlated at baseline ($r=0.48; P<.001$), and changes in grief symptoms correlated significantly with changes in psychopathology symptoms ($r=0.35; P<.001$). The pre-post correlation for grief was $r=0.68$, and the pre-post correlation for psychopathology symptoms was $r=0.64$ (both $P<.001$). Gains in ER correlated with gains in loss-related CSE ($r=0.45; P<.001$). Gains in ER and CSE correlated with improvements in grief symptoms and psychopathology symptoms. Gains in self-efficacy and ER, improvement in grief, and improvement in psychopathology symptoms did not correlate significantly with the event, time since the event, age, or gender. Therefore, because of the rather small sample size, we did not include covariates in the mediation models. Regression analyses showed that the intervention predicted a significant increase in ER ($\beta=.33; P=.001$) and self-efficacy ($\beta=.30; P=.002$; Table 3).
Table 2. Correlations between improvements in grief and psychopathology, gains in coping self-efficacy, gains in emotion regulation, and baseline variables.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>TRIG change</th>
<th>BSI change</th>
<th>TRIG before</th>
<th>BSI before</th>
<th>CSE change</th>
<th>ER change</th>
<th>Event</th>
<th>Time since event</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIG change</td>
<td>—^f</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>BSI change</td>
<td>0.35^f</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>TRIG before</td>
<td>0.29^b</td>
<td>0.17</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>BSI before</td>
<td>0.04</td>
<td>0.50^g</td>
<td>0.48^g</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CSE change</td>
<td>0.28^b</td>
<td>0.36^g</td>
<td>0.01</td>
<td>0.05</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ER change</td>
<td>0.48^g</td>
<td>0.38^g</td>
<td>0.24^i</td>
<td>0.14</td>
<td>0.45^g</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Event</td>
<td>0.02</td>
<td>−0.11</td>
<td>0.15</td>
<td>−0.08</td>
<td>−0.15</td>
<td>−0.11</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Time since event</td>
<td>0.04</td>
<td>0.02</td>
<td>−0.12^j</td>
<td>−0.04</td>
<td>0.14</td>
<td>0.18</td>
<td>−0.13^i</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sex</td>
<td>−0.15</td>
<td>−0.14</td>
<td>−0.07</td>
<td>−0.14</td>
<td>0.12</td>
<td>0.02</td>
<td>−0.12</td>
<td>−0.08</td>
<td>—</td>
</tr>
<tr>
<td>Age</td>
<td>0.06</td>
<td>0.01</td>
<td>0.05</td>
<td>−0.02</td>
<td>−0.06</td>
<td>0.01</td>
<td>0.48^g</td>
<td>0.16^i</td>
<td>0.10</td>
</tr>
</tbody>
</table>

^aTRIG: Texas Revised Inventory of Grief.
^bBSI: Brief Symptom Inventory.
^cCSE: coping self-efficacy.
^dER: emotion regulation.
^eEvent: 0=separation or divorce, 1=spousal bereavement.
^fNot applicable.
^gP<.001.
^hP<.01.
^iP<.05.
^jP<.10.
Table 3. Results of the mediation analyses.

<table>
<thead>
<tr>
<th>Analyses</th>
<th>$B$</th>
<th>SE</th>
<th>$P$ value</th>
<th>95% CI</th>
<th>$\beta$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gains in emotion regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>1.83</td>
<td>0.55</td>
<td>.001</td>
<td>0.76 to 2.91</td>
<td>.33</td>
<td>0.107</td>
</tr>
<tr>
<td><strong>Gains in coping self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>1.99</td>
<td>0.66</td>
<td>.002</td>
<td>0.71 to 3.28</td>
<td>.30</td>
<td>0.089</td>
</tr>
<tr>
<td><strong>Models with one mediator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Improvement in grief (TRIG) and emotion regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.37</td>
<td>0.13</td>
<td>.004</td>
<td>0.12 to 0.63</td>
<td>.26</td>
<td>0.305</td>
</tr>
<tr>
<td>Gains ER</td>
<td>0.10</td>
<td>0.02</td>
<td>&lt;.001</td>
<td>0.06 to 0.15</td>
<td>.41</td>
<td>0.305</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gains ER</td>
<td>0.19</td>
<td>0.07</td>
<td>.009</td>
<td>0.05 to 0.33</td>
<td>.13</td>
<td>0.305</td>
</tr>
<tr>
<td><strong>Improvement in grief (TRIG) and coping self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.46</td>
<td>0.14</td>
<td>&lt;.001</td>
<td>0.19 to 0.73</td>
<td>.32</td>
<td>0.209</td>
</tr>
<tr>
<td>Gains CSE</td>
<td>0.05</td>
<td>0.02</td>
<td>.02</td>
<td>0.01 to 0.09</td>
<td>.24</td>
<td>0.209</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gains CSE</td>
<td>0.10</td>
<td>0.05</td>
<td>.06</td>
<td>−0.005 to 0.21</td>
<td>.07</td>
<td>0.209</td>
</tr>
<tr>
<td><strong>Improvement in psychopathology (BSI) and emotion regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.12</td>
<td>0.08</td>
<td>.13</td>
<td>−0.04 to 0.28</td>
<td>.15</td>
<td>0.183</td>
</tr>
<tr>
<td>Gains ER</td>
<td>0.05</td>
<td>0.02</td>
<td>&lt;.001</td>
<td>0.02 to 0.08</td>
<td>.35</td>
<td>0.183</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gains ER</td>
<td>0.09</td>
<td>0.04</td>
<td>.02</td>
<td>0.02 to 0.17</td>
<td>.12</td>
<td>0.183</td>
</tr>
<tr>
<td><strong>Improvement in psychopathology (BSI) and coping self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.13</td>
<td>0.08</td>
<td>.10</td>
<td>−0.02 to 0.29</td>
<td>.16</td>
<td>0.181</td>
</tr>
<tr>
<td>Gains CSE</td>
<td>0.04</td>
<td>0.01</td>
<td>&lt;.001</td>
<td>0.02 to 0.07</td>
<td>.35</td>
<td>0.181</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gains CSE</td>
<td>0.08</td>
<td>0.04</td>
<td>.02</td>
<td>0.02 to 0.16</td>
<td>.10</td>
<td>0.181</td>
</tr>
<tr>
<td><strong>Models with 2 parallel mediators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Improvement in grief (TRIG)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.37</td>
<td>0.13</td>
<td>.005</td>
<td>0.11 to 0.63</td>
<td>.26</td>
<td>0.297</td>
</tr>
<tr>
<td>Gains ER</td>
<td>0.10</td>
<td>0.03</td>
<td>&lt;.001</td>
<td>0.04 to 0.15</td>
<td>.39</td>
<td>0.297</td>
</tr>
<tr>
<td>Gains CSE</td>
<td>0.01</td>
<td>0.02</td>
<td>.73</td>
<td>−0.04 to 0.05</td>
<td>.04</td>
<td>0.297</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gains ER</td>
<td>0.19</td>
<td>0.07</td>
<td>.008</td>
<td>0.03 to 0.33</td>
<td>.13</td>
<td>0.297</td>
</tr>
<tr>
<td>Gains CSE</td>
<td>0.02</td>
<td>0.05</td>
<td>.73</td>
<td>−0.08 to 0.11</td>
<td>.01</td>
<td>0.297</td>
</tr>
<tr>
<td>CSE and ER</td>
<td>0.18</td>
<td>0.08</td>
<td>.02</td>
<td>0.05 to 0.34</td>
<td>.14</td>
<td>0.297</td>
</tr>
</tbody>
</table>
Mediation Analyses

Overview

To investigate whether gains in ER and loss-related CSE were mechanisms of change, we used mediation models for improvements in grief and psychopathology symptoms. Table 3 presents the results of the mediation analyses, including the indirect effects. Figure 1 depicts the path models with direct paths for the models with the simultaneous inclusion of both mediators.

Figure 1. Path models for improvement in grief (A) and psychopathology (B); direct effects, standardized coefficients. *P < .05; **P < .01; ***P < .001; t: P = .054.

Mediation Models With a Single Mediator

Improvement in grief was associated with gains in ER ($\beta$=.41; $P \leq .001$) and gains in CSE ($\beta$=.24; $P = .02$). Only ER showed a significant indirect path with a medium effect size ($\beta$=.13; $P = .009$). The indirect effect for CSE showed a small to medium-sized effect but did not reach the significance level ($\beta$=.13; $P = .06$).

Improvement in psychopathology symptoms was related to gains in ER and loss-related self-efficacy (ER: $\beta$=.35, $P \leq .001$; CSE: $\beta$=.35, $P \leq .001$). ER and loss-related self-efficacy functioned as
mediators with medium effect sizes (indirect effects: ER: $\beta=0.12$, $P=0.02$; CSE: $\beta=0.10$, $P=0.02$).

**Models With Simultaneous Parallel Mediators**

Including ER and loss-related CSE simultaneously in the analyses, the relative strength of both mediators was investigated in exploratory models. Improvement in grief was associated with gains in ER ($\beta=0.39$; $P \leq 0.001$) but not with gains in self-efficacy ($\beta=0.04$; $P=0.73$; Figure 1). Only ER mediated the association between taking part in the intervention and improvement in grief (indirect effect: $\beta=0.13$; $P=0.008$). The path from intervention to improvement in grief remained significant ($\beta=0.26$; $P=0.005$). The total indirect effect was significant and showed a medium effect size ($\beta=0.14$; $P=0.02$).

**Improvement in psychopathology symptoms** was significantly predicted by gains in loss-related CSE ($\beta=0.23$; $P=0.04$; Figure 1). ER showed a similar effect size but did not reach the significance level ($\beta=0.24$; $P=0.05$). The total indirect effect was significant ($\beta=0.15$; $P=0.003$) but not for the specific indirect paths. However, post hoc Monte Carlo power analysis for indirect effects indicated that the models with 2 parallel mediators did not have enough power to detect specific indirect effects apart from the indirect path from ER on the improvement of grief (Multimedia Appendix 1).

**Discussion**

**Principal Findings**

This study examined ER and loss-related CSE as putative mediators for improvement in grief and psychopathology symptoms in an internet intervention for older adults after spousal bereavement, separation, or divorce. The results suggested that the cognitive behavioral intervention called LIVIA increased both ER and CSE, which correlated with improvements in grief and psychopathology symptoms. Mediation models confirmed ER and loss-related CSE as mediators of improvements in psychopathology. For improvements in grief, only ER showed a significant indirect effect, whereas CSE showed a trend. A model that simultaneously included both predictors suggested that only gains in ER mediated the association between participating in LIVIA and improvement in grief. Only the total indirect effect was significant in the model for improvement in psychopathological symptoms.

Our findings are in line with previous studies showing that ER was linked to treatment outcomes such as depression, negative affect, and other psychological disorders [23-25]. Our results are also consistent with findings that CSE was associated with lower emotional distress, higher psychological well-being in widows, and lower grief symptoms in students who had lost a close person in a university campus shooting [18,19].

Extending existing knowledge, our study investigated ER and loss-related CSE as putative mediators of improvements in grief and psychopathological symptoms after the loss of a spouse. The mediation models with single mediators confirmed a specific direct and indirect effect of gains in ER on improvement in grief and psychopathology symptoms. This suggests that ER was a mediator and potential mechanism of change. Gains in CSE showed significant direct effects on grief and psychopathology symptoms and a significant indirect effect on psychopathology symptoms. However, there was only a marginal indirect effect on grief symptoms. A power analysis revealed a power of 0.59 for finding a significant indirect path via CSE. This effect may have been significant in a larger sample.

Exploratory mediation models with the simultaneous inclusion of ER and loss-related CSE aimed to investigate the specificity and relative strength of these variables as parallel mediators. The results confirmed the importance of ER for improvement in grief and suggested that loss-related CSE was less relevant. However, the models were underpowered to detect any significant indirect effects. This also impedes a clear interpretation of the model for improvement in psychopathology symptoms.

Referring to the dual process model of coping with bereavement [28], in addition to loss-related CSE, ER skills may be especially important for loss-oriented work, such as experiencing the pain of loss. An improvement in ER skills such as self-soothing or cheering oneself up may make these processes more tolerable as individuals can regulate overwhelming or more persistent emotions. In addition, oscillating between loss- and restoration-oriented tasks may be fostered by better ER skills.

Our findings underline the crucial importance of ER for grief interventions and corroborate previous studies suggesting that ER was a mediator in an internet intervention for stress management [27]. Nevertheless, prolonged grief assessed with the TRIG has a strong separation distress component, which can be regarded as an ER problem. Therefore, these results do not necessarily generalize to persistent complex bereavement disorders, which also include avoidance symptoms and impairment in social, occupational, and other areas of life.

ER and loss-related CSE resulted in significant specific indirect effects on improvements in psychopathology symptoms, suggesting that both may be mediators for improvement in psychopathology. This is in line with Benight et al [20], who found that a change in CSE predicted a decrease in posttraumatic stress. High loss-related CSE may facilitate mainly restoration-oriented tasks such as addressing all the changes caused by the loss and creating a new life without the spouse, which may be perceived as very demanding. The belief in the ability to cope with these tasks may render them less threatening and alleviate stress and anxiety, as well as promote engagement in coping behavior and sustain coping efforts [17]. In addition, CSE may foster a sense of autonomy, self-determination, purpose in life, and perceived environmental mastery, which leads to less avoidant behavior and less aversive rumination [18]. Thus, loss-related CSE could promote a positive adaptation to life without a partner and decrease grief.

A further difference between the mediation models for grief and psychopathology was the significant direct effect of taking part in LIVIA and improvement in the outcome variables. Interestingly, and in contrast to the model for grief, no significant direct path existed from the intervention to improvements in psychopathological symptoms. In addition, improvement in grief was greater than that in psychopathology.
symptoms ($d=0.81$ vs $d=0.39$) [33], and the amount of explained variance was larger for grief than for psychopathology symptoms (30% vs 13%). This may reflect the fact that grief was a major focus of LIVIA, which was specifically developed for older adults who had lost their spouse through bereavement or divorce. For example, a module focused on positive social relationships, which may be more important for improvement in grief than for psychopathology symptoms. Thus, improvement in social support might be an additional mediator for the improvement in grief.

Exploratory mediation models with the simultaneous inclusion of ER and CSE suggested that ER was more important for improvement in grief than CSE, whereas both mediators showed similar effects for improvement in psychopathology. Regarding the interplay between ER and self-efficacy in grief processing, one can speculate that ER may be a more fundamental process than loss-related CSE as a social cognitive variable. The ability to modify negative emotions seems to have a positive effect on loss-related CSE. As Bandura [16] pointed out, emotional arousal can reduce self-efficacy in threatening situations as high arousal can debilitate performance. For example, among combat veterans, ER difficulties had only an indirect effect on a lower quality of life and higher posttraumatic stress symptom severity via lower CSE [43].

Limitations
Our study has several limitations. A meta-analysis demonstrated that web-based interventions increased disease-specific but not general self-efficacy, and context-specific measures are assumed to be more predictive of adjustment to stress [16]. As there are no psychometrically validated scales for assessing ER and loss-related CSE after bereavement and separation or divorce, we created 3 contextualized items to measure ER and loss-related CSE, specifically in the context of spousal bereavement and separation or divorce and the content of LIVIA. Furthermore, we aimed to keep the questionnaires short in order to not overburden the participants, which could have increased the attrition rate.

As a further limitation, we only had 2 measurement points during the 10-module intervention (ie, preintervention and postintervention) and then computed the difference scores for gains in ER and CSE. Thus, we assessed only changes in ER and CSE over the whole intervention, and thus, there was some temporal overlap in the measurement of the mediators and outcome variables. Therefore, results of the mediation analyses should be interpreted with caution. A fine-grained temporal design may also be able to more accurately detect the temporal sequence of change and the interplay of mediators and thus disentangle the mechanisms of change. However, the speed and shape of change are not necessarily linear, and sudden gains or losses may occur (for more details, see the study by Aderka et al [44]). Thus, the appropriate time point for assessing the mediators for capturing these changes may be difficult to determine, and the temporal associations between changes in the mechanism and changes in outcomes may be hard to disentangle [45]. Moreover, the sample size of 100 participants limited the number of variables in the models, precluded a more detailed analysis of the interplay between potential moderators and mediators, and led to power issues for the models with 2 parallel mediators. Moreover, the sample included only 20% (20/100) of widowed individuals, which precluded separate models for widowed participants.

Considering these limitations, the results of this study must be replicated and extended by using larger samples and more measurement points. Further research should use validated measures for ER and loss-related CSE and investigate whether the greater relative importance of ER compared with CSE is specific to prolonged grief symptoms or whether it also generalizes to distress-related disorders and other psychological disorders such as anxiety disorders. In addition, other potential mediators such as social support could be examined together with ER and CSE.

Despite these limitations, the findings of this study have several clinical implications. Gains in ER and loss-related CSE are promising targets for improving internet interventions and probably also face-to-face interventions for coping with the consequences of spousal bereavement, separation or divorce. Gains in ER and self-efficacy could be conceptualized as common factors in psychotherapy related to improved behavioral regulation and changing expectations of personal effectiveness [15]. Thus, one could hypothesize that ER and CSE are also mechanisms of change in interventions for different psychological disorders and in the promotion of psychological well-being in general.

Depending on the problems of the participants, more specific modules for ER or CSE could be added to the intervention. Techniques aimed at strengthening ER include, for example, emotional skills training. CSE could be bolstered by training in adaptive coping strategies, mastery experiences, or reappraising emotional and physiological reactivity [16]. These techniques could be integrated as additional modules or replace less effective modules. Alternatively, users of existing interventions could be advised to spend more time and effort on their respective modules.

Conclusions
This study is one of the few to investigate mediators in internet interventions and the first to examine emotional and cognitive processes as mediators in grief processing after spousal bereavement, separation, or divorce. Our findings suggest that ER and loss-related CSE mediated treatment outcomes and are promising therapeutic targets for improving grief and psychopathology symptoms in internet interventions. ER and CSE should be examined as transdiagnostic or disorder-specific putative mediators in internet interventions for other disorders.
Acknowledgments
The authors thank Franziska Linder, Gabriella Weiss, Diana Gsponer, Vera Bergamaschi, Rebekka Strub, and Laura Durrer for their contribution to the email support of the participants, data collection, and preliminary analyses of the data. The authors thank Timo Stolz for his help with the programming of the intervention and all the participants in this trial for their involvement. This work was supported by the Swiss National Science Foundation (grant 51NF40-160590 granted to Dario Spini). The Swiss National Science Foundation had no role in the design of the study; collection, analysis, and interpretation of data; or in writing the manuscript.

Authors’ Contributions
HZ, TB, and JB conceived the study and the design. TB set up the intervention. JB, NB, and FR conducted the statistical analyses. NB was in charge of data collection. JB wrote the first draft of the manuscript. SJS, TB, and HZ revised the manuscript critically for important intellectual content. All authors contributed to the manuscript and approved the submitted version.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Exploratory and confirmatory factor analyses for emotion regulation and loss-related coping self-efficacy and Monte Carlo power analysis for indirect effects.

[DOCX File, 21 KB - mental_v9i5e27707_app1.docx ]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 373 KB - mental_v9i5e27707_app2.pdf ]

References


21. Probst T, Dehoust M, Br...
Abbreviations

CBT: cognitive behavioral therapy
CSE: coping self-efficacy
ER: emotion regulation
TRIG: Texas Revised Inventory of Grief
Ownership, Use of, and Interest in Digital Mental Health Technologies Among Clinicians and Young People Across a Spectrum of Clinical Care Needs: Cross-sectional Survey

Imogen H Bell¹², BDSc, PhD; Andrew Thompson¹²³, MD; Lee Valentine¹², BA, MSocWk; Sophie Adams⁴, MD, PhD; Mario Alvarez-Jimenez¹², DClinPsy, PhD; Jennifer Nicholas¹², BA, BSc, PhD

¹Orygen, Parkville, Australia
²Centre for Youth Mental Health, University of Melbourne, Parkville, Australia
³Division of Mental Health and Wellbeing, Warwick Medical School, University of Warwick, Coventry, United Kingdom
⁴Austin Health, Melbourne, Australia

Corresponding Author:
Imogen H Bell, BDSc, PhD
Orygen
35 Poplar Rd
Parkville, 3052
Australia
Phone: 61 430156222
Email: imogen.bell@orygen.org.au

Abstract

Background: There is currently an increased interest in and acceptance of technology-enabled mental health care. To adequately harness this opportunity, it is critical that the design and development of digital mental health technologies be informed by the needs and preferences of end users. Despite young people and clinicians being the predominant users of such technologies, few studies have examined their perspectives on different digital mental health technologies.

Objective: This study aims to understand the technologies that young people have access to and use in their everyday lives and what applications of these technologies they are interested in to support their mental health. The study also explores the technologies that youth mental health clinicians currently use within their practice and what applications of these technologies they are interested in to support their clients’ mental health.

Methods: Youth mental health service users (aged 12-25 years) from both primary and specialist services, young people from the general population (aged 16-25 years), and youth mental health clinicians completed a web-based survey exploring technology ownership, use of, and interest levels in using different digital interventions to support their mental health or that of their clients.

Results: A total of 588 young people and 73 youth mental health clinicians completed the survey. Smartphone ownership or private access among young people within mental health services and the general population was universal (611/617, 99%), with high levels of access to computers and social media. Youth technology use was frequent, with 63.3% (387/611) using smartphones several times an hour. Clinicians reported using smartphones (61/76, 80%) and video chat (69/76, 91%) commonly in clinical practice and found them to be helpful. Approximately 50% (296/609) of the young people used mental health apps, which was significantly less than the clinicians (χ²=28.8, n=670; P<.001). Similarly, clinicians were significantly more interested in using technology for mental health support than young people (H³=55.90; P<.001), with 100% (73/73) of clinicians being at least slightly interested in technology to support mental health compared with 88% (520/591) of young people. Follow-up tests revealed no difference in interest between young people from the general population, primary mental health services, and specialist mental health services (all P>.23). Young people were most interested in web-based self-help, mobile self-help, and blended therapy.

Conclusions: Technology access is pervasive among young people within and outside of youth mental health services; clinicians are already using technology to support clinical care, and there is widespread interest in digital mental health technologies among these groups of end users. These findings provide important insights into the perspectives of young people and clinicians regarding the value of digital mental health interventions in supporting youth mental health.

(JMIR Ment Health 2022;9(5):e30716)  doi:10.2196/30716
Introduction

Digital mental health interventions (DMHIs) are interventions that use technologies, such as smartphones, smartwatches, or computer programs, to provide information, support, or treatment for mental health, most commonly using the internet [1]. As either standalone self-help tools or those used in conjunction with standard care in the form of blended therapy [2], DMHIs have the potential to support mental health and well-being [3]. Interest in DMHIs often centers on young people because of the high prevalence of mental health difficulties in this age group coupled with their frequent use of technology [4-6]. Indeed, the potential to access mental health support via technology may be particularly important for young people, given that their accessed and sustained engagement with mental health services are limited [7,8].

The role of technology in supporting mental health was made starkly clear by the global COVID-19 pandemic. During this time, many nations became reliant on technology-enabled service delivery to provide mental health care at a distance [9,10]. For many, telehealth has become the norm, enabling direct client contact via telephone or videoconferencing [11]. This unique moment in history has catalyzed an important shift in the perceived value of technology-supported care. However, although research indicates that telehealth can be as effective as face-to-face treatment and may improve service quality in the eyes of young people with mental health difficulties [12,13], it only represents the tip of the iceberg of how technology can support mental health. In the wake of the pandemic, there exists an opportunity to capitalize on the increased interest and acceptance of technology-enabled care to deliver new digital interventions that not only provide a more convenient way of delivering treatment but also the potential to enhance it [14].

Despite the potential of DMHIs, a lack of long-term engagement has often been reported. This is true both in clinical trials [15] and, particularly, in naturalistic studies in which apps are used in the wild [16], where good initial uptake is commonly followed by a dramatic drop in use over time [15]. Poor engagement with DMHIs has been highlighted as a significant problem in the field and has formed the focus of several reviews [17-20]. A common theme identified in this literature is a lack of fit between evidence-based DMHIs and the needs of end users for whom they are designed to help [20]. A lack of emphasis on understanding end user needs has resulted in an early generation of DMHIs that have generally lacked relevance and interest for users. Critical learnings from this early work have resulted in the greater emphasis placed on involving end users in the design process, as well as a need for research dedicated to understanding their needs and preferences. Young people constitute a particular type of end user who tends to be highly exposed to technologies in daily life, making them particularly critical of digital products [19]. Therefore, understanding the unique perspectives of young people is important to inform the development of DMHIs for youth.

A shift toward practices that prioritize the needs and preferences of young people as end users is required to ensure that DMHIs are engaging and fit for purpose [19,21]. However, there is currently a dearth of research on the technologies that young people are interested in using to support their mental health. Qualitative studies have explored experiences with DMHIs among young people, finding preferences toward their use to support, rather than replace, face-to-face services, as well as a desire to tailor digital interventions to individual preferences [22-24]. Quantitative findings in youth populations are limited, although 2 studies in small samples of young people in early psychosis services found high levels of technology ownership and use in these populations [25,26] and an interest in technology for a variety of purposes to support self-management and functional recovery [25]. Although these findings provide some insight into technology use and preferences, the qualitative findings are limited in generalizability, and quantitative research has involved young people with specific mental health conditions. To fill this gap, this study aims to understand what technologies young people, both within youth mental health services and in the general population, have access to and use in their everyday lives, and which applications of these for supporting their mental health they are most interested in. Furthermore, as DMHIs are most effective when combined with human support [12,27,28], a likely use case is the blending of these tools within youth mental health services. Despite this, there are very few examples of the successful implementation of DMHIs within clinical settings, highlighting the significant gap between research and practice in digital mental health [29].

As such, in addition to young people, this study aims to investigate the use of and interest in different DMHIs among clinicians in youth mental health services to support their clinical work.

Methods

Ethics Approval

The study was approved by the Melbourne University human research ethics committee (approval numbers 2057299 and 2056793) and the Melbourne Health human research ethics committee (reference number QA2020096) and complied with the Declaration of Helsinki.

Study Design and Context

Young people and mental health clinicians completed a web-based survey as part of the BRACE project, which examined the effects of COVID-19 on the mental health and well-being of young people living in Australia, telehealth service quality, and the potential of technology to support youth mental health care. Data collection for the project occurred during and immediately after Australian Federal and State government–mandated lockdown restrictions (stage 3) that included socially distancing from individuals not part of a household and limited ability to leave home [30]. During the
Technology Access and Use

Technology access and use were explored by asking young people if they owned or had private access to various technologies, ranging from smartphones and laptops to social media and gaming consoles. Those who indicated that they had access to the technology were asked how often they used it on a Likert scale ranging from less than once a week to several times an hour. Similarly, clinicians were asked if they had used the same technologies in their clinical practice. For the technologies they had used, they rated how helpful they thought the technology was for their clients. Both young people and clinicians were asked whether they had used a mental health app or recommended a smartphone app for their clients’ mental health. Young people who had used apps to support their mental health were asked which apps they had used and to rate their helpfulness. Clinicians who had recommended apps to their clients were asked to name the apps they had recommended and rate how helpful they were for their clients.

Technology Interest

The level of interest in using 20 different technologies commonly used to support mental health was measured on a 5-point Likert scale ranging from not at all interested to extremely interested. Technologies ranged from established resources such as telehealth, websites, and helplines to emerging digital mental health tools such as virtual reality (VR), serious games, and chatbots. Young people rated their interest in using each technology to support their mental health, whereas clinicians rated their interest in using or recommending each technology to support the mental health and well-being of their clients.

All quantitative items were measured on Likert scales, with anchors varying depending on the question, as specified in the results. A full copy of the survey is provided in Multimedia Appendix 1.

Mental Health Measures

The Patient Health Questionnaire-4 [32] was used to characterize the mental health status of the participants in the sample. The Patient Health Questionnaire-2 (PHQ-2) is a 2-item, brief self-report screening questionnaire for clinical depression. Similarly, the Generalized Anxiety Disorder-2 (GAD-2) is a 2-item brief self-report screening questionnaire for clinical anxiety. Items are on both measures rated on a 4-point Likert type scale from 1 (not at all) to 4 (nearly every day). The total scores range from 0 to 6, with higher scores indicating greater levels of depression or anxiety. A score of ≥3 on the 2-item PHQ-2 indicates probable depressive disorder, and a score of ≥3 on the 2-item GAD-2 indicates probable anxiety disorder for adults and young people in primary care settings and the general population [32,33].
between analyses and is reported where it differed. Chi-square statistics were used to examine differences among participant groups (young people from the general population, young people from primary mental health services, young people from specialist mental health services, and clinicians) and the use of apps for mental health. To gain an indication of participants’ overall interest in technology to support mental health, overall interest in technology was calculated as the mean of an individual’s interest scores across the 20 technology types. Kruskal-Wallis tests with Bonferroni-corrected follow-up contrasts were used to examine differences among participant groups in terms of overall interest in using technology to support mental health. Similar technologies were grouped to examine differences among participant groups concerning interest in technology types. The following seven groups were formed by the research team based on the original 20 technology items:

1. Web-based self-help (web-based therapy, mental health websites, and web-based employment support)
2. Mobile self-help (apps to support mental health, apps to track mental health, and wearables to track mental health such as smartwatches)
3. Telehealth (video chat with clinician, telephone with clinician, texting with clinician, and mental health support lines)
4. Blended therapy (blended therapy and sharing mental health information with clinicians on the web)
5. Social media (secure social media to connect with young people about mental health and social media to connect with clinicians about mental health)
6. Immersive technologies (VR for mental health strategies, augmented reality for mental health strategies, VR with clinicians, and virtual worlds for mental health groups)
7. Interactive technologies (chatbots for mental health support and digital games for mental health support)

Results

Sample Characteristics

Within primary care services, an SMS text message link to the survey was sent to 1868 young people, 308 (16.49%) of whom responded to the survey, and of the 308 respondents, 229 (74.4%) completed it. Within specialist services, the survey was distributed to approximately 650 young people, of whom 59 (9.1%) responded, and of these 59 respondents, 53 (90%) completed it. The survey was also advertised on social media, and of the 693 people who clicked the link, 498 (71.9%) provided consent and were eligible, and of those who were eligible, 306 (61.4%) completed the survey items reported in this study. Finally, of the approximately 370 clinicians who received the survey link, 92 (25%) initiated the survey, and of those 92 clinicians, 73 (79%) completed it. The final sample comprised 73 clinicians across specialist and primary services and 588 young people (age range 12-25 years) from primary care, specialist services, and the general population. Demographic characteristics of the youth sample are shown in Table 1.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>General population (n=306)</th>
<th>Primary services (n=229)</th>
<th>Specialist services (n=53)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>21.20 (2.90)</td>
<td>18.77 (3.48)</td>
<td>21.08 (2.54)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>222 (72.5)</td>
<td>142 (62.0)</td>
<td>26 (49)</td>
</tr>
<tr>
<td>Male</td>
<td>58 (19)</td>
<td>63 (27.5)</td>
<td>26 (49)</td>
</tr>
<tr>
<td>Transgender</td>
<td>1 (0.3)</td>
<td>10 (4.4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>14 (4.6)</td>
<td>7 (3.1)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Unspecified</td>
<td>11 (3.6)</td>
<td>7 (3.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Aboriginal or Torres Strait Islander</td>
<td>6 (2)</td>
<td>4 (1.7)</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Current living situation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living with parents, caregivers, or siblings</td>
<td>201 (65.7)</td>
<td>191 (83.4)</td>
<td>39 (74)</td>
</tr>
<tr>
<td>Living with friends</td>
<td>29 (9.5)</td>
<td>3 (1.3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Living with romantic partner</td>
<td>30 (9.8)</td>
<td>11 (4.8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Living in shared accommodation</td>
<td>23 (7.5)</td>
<td>9 (3.9)</td>
<td>5 (9)</td>
</tr>
<tr>
<td>Living alone</td>
<td>23 (7.5)</td>
<td>14 (6.1)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Homeless or couch surfing</td>
<td>0 (0)</td>
<td>1 (0.4)</td>
<td>3 (6)</td>
</tr>
<tr>
<td><strong>State of residence, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>11 (2.4)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>New South Wales</td>
<td>31 (10.1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>1 (0.2)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Queensland</td>
<td>16 (5.2)</td>
<td>0 (0)</td>
<td>16 (30)</td>
</tr>
<tr>
<td>South Australia</td>
<td>10 (3.3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Tasmania</td>
<td>17 (5.6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Victoria</td>
<td>211 (69.0)</td>
<td>229 (100)</td>
<td>37 (70)</td>
</tr>
<tr>
<td>Western Australia</td>
<td>9 (2.9)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Employment status, b n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time student</td>
<td>182 (59.5)</td>
<td>126 (55.0)</td>
<td>13 (25)</td>
</tr>
<tr>
<td>Part-time student</td>
<td>35 (11.4)</td>
<td>15 (6.6)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Hours of study each week, mean (SD)</td>
<td>24.98 (12.22)</td>
<td>22.14 (17.96)</td>
<td>16.43 (8.77)</td>
</tr>
<tr>
<td>Full-time paid employment, n (%)</td>
<td>54 (17.6)</td>
<td>13 (5.7)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Part-time paid employment, n (%)</td>
<td>103 (33.7)</td>
<td>34 (14.8)</td>
<td>9 (17)</td>
</tr>
<tr>
<td>Hours of work each week, mean (SD)</td>
<td>23.35 (13.33)</td>
<td>19.72 (12.75)</td>
<td>24.02 (11.68)</td>
</tr>
<tr>
<td>Unpaid worker as a parent or carer, n (%)</td>
<td>6 (2.0)</td>
<td>1 (0.4)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Currently unemployed, n (%)</td>
<td>43 (14.1)</td>
<td>72 (31.4)</td>
<td>30 (57)</td>
</tr>
<tr>
<td><strong>Mental health, c,d n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potential clinical depression</td>
<td>133 (43.5)</td>
<td>69 (62.7)</td>
<td>30 (57)</td>
</tr>
<tr>
<td>Potential clinical anxiety</td>
<td>152 (49.7)</td>
<td>65 (59.1)</td>
<td>31 (60)</td>
</tr>
</tbody>
</table>

<sup>a</sup>ACT: Australian Capital Territory.

<sup>b</sup>Categories are not mutually exclusive.

<sup>c</sup>Patient Health Questionnaire-2 and Generalized Anxiety Disorder-2.

<sup>d</sup>A score of ≥3 on the 2-item depression and anxiety screening measures indicates probable depressive or anxiety disorder (n=110).
Technology Access and Use

Access and Use of Technology by Young People

Young people’s access to different technologies is displayed in Table 2. Smartphone access was universal (611/617, 99%), including among young people from primary and specialist services. Across the groups, young people reported high rates of video chat, instant messenger, and social media access and lower levels of access to wearable technologies and VR. Overall, technology use was frequent. Of the young people that had access to the various technologies, use varied according to technology type, as illustrated in Figure 1. Most young people (387/611, 63.3%) reported using their smartphones several times an hour, and a high proportion used social media (540/584, 92.5%), instant messaging (509/574, 88.7%), and computers (397/540, 73.5%) at least once or several times a day, with hourly use being the most common.

Table 2. A comparison of access to different technologies among young people from the general population, primary services, and specialist services and use of technology for clinical care among clinicians (N=693).

<table>
<thead>
<tr>
<th>Technologies</th>
<th>Young people from the general population (n=327), n (%)</th>
<th>Young people from primary services (n=236), n (%)</th>
<th>Young people from specialist services (n=54), n (%)</th>
<th>Clinicians (n=76), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone</td>
<td>321 (98.2)</td>
<td>236 (100)</td>
<td>54 (100)</td>
<td>61 (80)</td>
</tr>
<tr>
<td>iPhone</td>
<td>233 (71.2)</td>
<td>145 (61.4)</td>
<td>29 (54)</td>
<td>—</td>
</tr>
<tr>
<td>Android</td>
<td>88 (26.9)</td>
<td>91 (38.5)</td>
<td>25 (46)</td>
<td>—</td>
</tr>
<tr>
<td>Social media</td>
<td>316 (96.6)</td>
<td>222 (94.1)</td>
<td>46 (85)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Instant messenger</td>
<td>313 (95.7)</td>
<td>215 (91.1)</td>
<td>46 (85)</td>
<td>7 (9)</td>
</tr>
<tr>
<td>Laptop</td>
<td>306 (93.6)</td>
<td>197 (83.5)</td>
<td>37 (69)</td>
<td>55 (72)</td>
</tr>
<tr>
<td>Video chat</td>
<td>286 (87.4)</td>
<td>185 (78.4)</td>
<td>44 (81)</td>
<td>69 (91)</td>
</tr>
<tr>
<td>Gaming console</td>
<td>153 (46.8)</td>
<td>152 (64.4)</td>
<td>40 (74)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Tablet</td>
<td>117 (35.8)</td>
<td>84 (35.6)</td>
<td>15 (28)</td>
<td>27 (36)</td>
</tr>
<tr>
<td>Wearables</td>
<td>92 (28.1)</td>
<td>43 (18.2)</td>
<td>9 (17)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Desktop</td>
<td>80 (24.5)</td>
<td>66 (28)</td>
<td>16 (30)</td>
<td>43 (57)</td>
</tr>
<tr>
<td>Landline</td>
<td>75 (22.9)</td>
<td>57 (24.1)</td>
<td>13 (24)</td>
<td>33 (43)</td>
</tr>
<tr>
<td>Virtual reality</td>
<td>18 (5.5)</td>
<td>9 (3.8)</td>
<td>5 (9)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

aData not available.

Figure 1. Young people’s average frequency of use across technologies that they have access to (as presented in Table 2).
Technology Use by Clinicians

The proportion of clinicians who used different technologies in their clinical work is presented in Table 2. Although most clinicians used video chat (69/76, 91%) and smartphones (61/76, 80%) within their practice, few reported using newer or social technologies such as wearables (3/76, 4%), social media (4/76, 5%), or VR (1/76, 1%). The perceived helpfulness of the technology recommendations for clients is presented in Figure 2. Of the technologies used in clinical practice, most clinicians rated them helpful or very helpful for their clients.

Figure 2. Clinicians’ perceived helpfulness of different technologies that they have used within clinical care (as presented in Table 2).

Mental Health App Use

Approximately half of all participants (347/670, 51.8%) had used a mental health app themselves (young people: 296/609, 48.6%) or recommended one to their clients (clinicians: 51/61, 84%). A chi-square test for independence indicated a significant difference between participant groups and the use of apps for mental health ($\chi^2=28.8, n=670; P<.001; \text{Cramer } V=0.21$), with clinicians significantly more likely to recommend apps to support mental health care than young people were to have used mental health apps. The percentage of young people and clinicians who had used apps for mental health and the most common apps used are presented in Table 3. These apps were similar across groups of young people and clinicians and were used for mindfulness, meditation and relaxation, mood monitoring, and safety planning.

Table 3. Young people’s use of smartphone apps and clinicians’ use or recommendations of smartphone apps for clients (N=670).

<table>
<thead>
<tr>
<th>Participant groups</th>
<th>Used or recommended apps for mental health, n (%)</th>
<th>Most commonly used or recommended apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young people from the general population (n=319)</td>
<td>162 (50.8)</td>
<td>Smiling Mind, Headspace, Calm, and Calm harm</td>
</tr>
<tr>
<td>Young people from primary services (n=236)</td>
<td>111 (47)</td>
<td>Headspace, Smiling Mind, Calm, and Daylio</td>
</tr>
<tr>
<td>Young people from specialist services (n=54)</td>
<td>23 (43)</td>
<td>Calm, Headspace, Daylio, Smiling Mind, and YouTube</td>
</tr>
<tr>
<td>Clinicians (n=61)</td>
<td>51 (84)</td>
<td>Smiling Mind, BeyondNow, Headspace, and Calm</td>
</tr>
</tbody>
</table>

Overall, most young people from services (specialist services and primary care) reported that using apps to support their mental health was helpful or very helpful (82/132, 62.1%). Approximately 20.5% (27/132) neutral and 17.4% (23/132) reported apps to be unhelpful. Similarly, on average, young people from the general population found apps to be somewhat helpful (124/161, 77%). Approximately 13% (21/161) found them unhelpful. The vast majority of clinicians (45/48, 93.8%) felt that the apps were helpful to their clients.

Interest in Technology to Support Mental Health

Young people’s and clinicians’ interest in different technologies to support mental health across the 4 participant groups ($H_1=55.90; P<.001$). Follow-up tests with Bonferroni corrections were used to compare all pairs and indicated that clinicians were significantly more interested in using technology to support mental health than each of the groups of young people (general population: $\chi^2=-171.6, P<.001$; primary services: $\chi^2=-158.7, P<.001$; specialist services: $\chi^2=-218.9, P<.001$). However, there was no significant difference in interest between young people in specialist services, primary services, or the general population (all $P>.23$). Although responses varied among the range of technologies surveyed, most participants (593/664, 89.3%) were at least slightly interested in a use of technology to support mental health and well-being (young people: 520/591, 88%) or that of their clients (clinicians: 73/73, 100%).
The technologies with the most consistently high levels of interest across these populations were telehealth, apps to track mental health, and web-based and blended therapies. The lowest level of interest overall was for chatbots, wearables, and immersive technologies (VR and augmented reality). However, close to half of all respondents across the sample reported at least some interest in all technology types.

**Figure 3.** The average level of interest in different technological approaches to support mental health across the 4 participant groups: young people general population (n=306), young people primary services (n=229), young people specialist services (n=53), and clinicians (n=73). AR: augmented reality; MH: mental health; SM: social media; VR: virtual reality; YP: young people.

Similar technology types were then grouped to observe the patterns of interest more clearly between the participant groups (Figure 4). Young people in the general population were most interested in web-based and mobile self-help, whereas young people from primary services were most interested in web-based self-help and blended therapy. Similar to the general population, those in specialist services were most interested in web-based and mobile self-help, as were clinicians, who also had high levels of interest in blended therapy.

Technology interest for mental health may be influenced by the respondents’ familiarity with the technology. A Mann-Whitney U test of independence found that young people who owned or had access to VR were significantly more interested in using VR to support their mental health than those who did not (U=5473.5; z=-2.811; P=.005).
Discussion

Principal Findings

This study examined the access and use of digital technologies among young people from youth mental health services and the general population, as well as the interest in digital technology use for supporting mental health care among young people and clinicians. The findings indicate that young people had widespread access to technologies, with 99% (611/617) having access to a smartphone and 63.3% (387/611) using it on average every hour. Clinicians reported similarly high rates of technology use to support their clinical care, with 91% (69/76) reporting the use of video chat, 80% (61/76) reporting the use of smartphones, and most finding common technologies such as laptops and the internet helpful or very helpful.

Approximately 50% (296/609) of young people from within services and the general population reported using smartphone apps to support their mental health, and 84% (51/61) of the clinicians reported recommending them to their clients. Apps were reported to be helpful by 62.1% (82/132) of young people within services and 77% (124/161) in the general population. The vast majority of clinicians (45/48, 94%) found apps helpful for their clients. Levels of interest varied across different technologies for supporting youth mental health, although 100% (73/73) of clinicians were at least slightly interested in technology to support their clients’ mental health, and 88% (520/591) of the young people were interested in technology.

Rates of access to technology were high across young people from within and outside of youth mental health services, with 98% to 100% of those surveyed having access to an internet-enabled device such as a smartphone or computer. Furthermore, young people reported very frequent use of these technologies throughout their daily lives, averaging several times an hour for smartphones. This is in line with prior research showing access rates between 95% and 99% in youth populations within high-income countries [4,5], with young people describing they use these almost constantly [4]. Research into young people within youth mental health services has been limited, although some studies have found similar rates of approximately 90% within small clinical samples of young people with early psychosis [25,26]. The current findings add to this literature by demonstrating high rates of access and use of technologies within populations of young people who use youth mental health services, supporting the potential reach of DMHIs in this population.

Overall, 88% (520/591) of young people reported at least some interest in technologies to support their mental health and well-being, and this did not differ depending on whether they were using youth mental health services. However, the patterns of interest appeared to differ across groups. Although all young people showed high levels of interest in self-help technologies, particularly smartphone apps and web-based therapy, those from within the services were most interested in technologies that worked alongside a clinician, including blended therapies and telehealth. This highlights the perceived need among young people for technologies to support care delivery, a finding supported by research indicating that DMHIs are the most effective and engaging when used in conjunction with human support [20]. However, it is also important to note that access to youth mental health care is limited [34]; therefore, young people in the general population who may have an unmet need for care could rely more heavily on digital technologies as self-help tools to support their mental health. Young people within services, who, on the basis of the current findings, are likely to receive care that incorporates digital technology (ie,
blended therapy), may have a greater appreciation for technology to support the care they are receiving. This highlights the differences in the needs of the 2 populations and the important role that both self-help and blended technologies play in meeting the demand for mental health support among young people. Furthermore, although levels of interest did not significantly differ overall, contextual factors such as the level and type of mental health support being sought (eg, low-intensity psychological treatment vs crisis intervention) or stage of care (eg, in remission vs acutely unwell) suggest that the needs and interests in different digital mental health technologies are likely to differ. For example, a relatively well young person who is in remission may be interested in smartphone-based symptom monitoring to prevent relapse, whereas a young person in active treatment may be interested in telehealth services and web-based therapy support.

Clinicians also endorsed high rates of interest in recommending a wide range of digital technologies to support youth mental health, with 100% (73/73) reporting at least some interest. Patterns of interest appeared to map well with young people, primarily for video calls, self-help apps, and web-based therapy. The most consistently endorsed technology across young people and clinicians was websites providing web-based therapy or mental health information and smartphone apps to track and support mental health. Indeed, 40% (29/73) of the clinicians were extremely interested in apps, and 33% (24/73) were extremely interested in web-based therapy, 38% (28/73) when used in a blended way. This aligns with most research and development that has occurred in digital mental health, particularly for smartphone apps [35], supporting the clear consumer demand for these products. Furthermore, research supports young people’s interest in blending technology with standard treatment as a way of increasing accessibility, continuity, and consolidation of treatment, as well as a means of accessing posttherapy support and strengthening the face-to-face relationship between clients and therapists [23]. However, there is a lack of evidence-based web-based therapies and smartphone apps currently available to support youth mental health [36], with some key exceptions [24,37], highlighting a critical area for further research and development.

In contrast, clinicians and young people were relatively less interested in automated therapies, such as chatbots, and technologies that made use of platforms that were infrequently accessed and used, such as VR. Although this may represent genuinely lower levels of interest in these technologies, it is also possible that this reflects a lack of familiarity and experience with their use for mental health treatment. Indeed, people tend to hold less positive attitudes and are less likely to adopt technologies with which they are less familiar [38]. This interpretation is supported by the finding that those who had used VR were significantly more interested in using it for mental health support. VR has a strong emerging evidence base for supporting the delivery of psychological interventions [39,40], particularly for exposure therapy; however, these interventions have not been widely implemented in clinical services. As the technology landscape is changing rapidly, levels of interest may increase as novel technologies such as VR become more common.

Clinicians also reported frequently using technology to support their practice, with 91% (69/76) using video chat, 80% (61/76) using smartphones, and >80% finding these helpful. Furthermore, overall, clinician interest in recommending digital technologies to support youth mental health was significantly higher than young people’s interest in using them (although both groups displayed high levels of interest). This finding is consistent with prior results from the BRACE survey, showing that 98% of youth mental health clinicians endorsed the ongoing use of telehealth beyond the COVID-19 pandemic [13]. However, these findings contrast with prior research findings that clinicians hold tentative views about the role of technology in mental health [41], particularly in regards to these replacing their care. Although a comparison sample is not available, the widespread adoption of technologies to support care delivery during the COVID-19 pandemic may account for the positive attitude change among clinicians. Therefore, the level of clinician interest is a positive finding, as the field seeks to promote the adoption of technologies within care systems, traditionally a challenge partly because of staff resistance [42,43]. The current findings may exemplify the paradigm shift in digital mental health arising from the global pandemic toward more digitally enhanced models of care [14,44]. This contemporary model of care has been heralded as potentially overcoming critical limitations of current mental health care systems; therefore, this shift brings about new hope for reform [45]. However, the degree to which this optimism will continue as the COVID-19 pandemic normalizes and the critical reliance on digital technology reduces is yet to be determined.

Half of the young people reported using smartphone apps for their mental health, and 84% (51/61) of the clinicians had recommended them to their clients, with most finding these helpful. This difference between young people and clinicians was statistically significant, indicating that although apps may be commonly recommended by clinicians, this does not correspond directly with uptake by young people. Given that young people have high levels of exposure to digital technologies within their everyday lives [4], it is likely that their motivation to use these for mental health arises from multiple sources, including social influences [46]. Indeed, research studies have found that both adults [47] and young people [48] with mental ill health most commonly use social media, searches (including Google and app store), and informal recommendations to select mental health apps. These prior studies also show that recommendations from friends and family were a more common source of mental health apps than recommendations from health care providers. Future research would benefit from exploring the best means of engaging and supporting young people in using evidence-based DMHIs for their mental health, particularly using participatory methodologies that involve young people as the ultimate end users of these products.

Notably, the apps most commonly used by clinicians and young people were those with significant market dominance. A recent app store review by Lau et al [35] found that 90% of mental health app downloads are accounted for by only 4 different apps (Headspace, Calm, Youper, and Wysa). Headspace and Calm are widely used in the current sample, as well as others.
supporting mindfulness or relaxation, mood tracking, and safety planning. With estimates that 325,000 health apps are currently available [35], the restricted range of apps being used highlights the driving force of marketing behind consumer choice and demand. However, strong marketing rarely translates to effectiveness, with only 2% of the available apps being supported by any sort of research evidence [35], and many have been found to undergo questionable ethical practices around privacy and security [49,50]. Furthermore, the apps most commonly used or recommended by clinicians and young people to support youth mental health were not specifically designed for this purpose. Given the importance of designing DMHIs for end users and ensuring they are backed by strong evidence, maximizing the benefits of technologies to help young people with mental health difficulties requires more research to develop, evaluate, and disseminate purpose-built solutions designed specifically for, and alongside, young people with lived experience of mental health difficulties [51]. In particular, there is a clear dearth of available smartphone apps designed to be integrated into clinical treatment, despite the clear interest in these products among young people using services and clinicians. As young people and clinicians have reported high levels of interest in blended therapies, it is surprising that there are very few digital technologies designed to support clinical care currently available, with some exceptions [37]. This highlights a critical discrepancy between what young people and clinicians want and what is available, which may reflect the challenges in implementing digital interventions in service settings. Informing efforts to implement evidence-based DMHIs to support clinical care is a critical area for future research [29,52].

Strengths and Limitations

Although this study has a number of strengths, including its large sample of young people across the spectrum of need for care, the inclusion of clinicians as important additional stakeholders and end users of DMHIs, as well as the depth of the survey regarding different DMHIs, the findings should be interpreted with knowledge of study limitations. First, data were collected via technology; thus, respondents likely represent a sample of digitally enabled young people, and only a proportion of young people responded to the survey. A range of demographic factors such as income and education may have influenced young people’s access to, and beliefs about, technology; however, this information was not captured in this study. Importantly, particular populations of young people, such as those from culturally and linguistically diverse or low socioeconomic backgrounds, who may have a greater need for mental health care, may not be well represented in this survey because of lower rates of technology access in these populations. This was highlighted in another report from the BRACE survey as a primary consideration among clinicians regarding the suitability of DMHIs for some young people [13]. Other factors considered by clinicians included client willingness, access, and complexity of clinical presentation, highlighting the need for an individualized approach. This survey provides an overall picture of interest levels in DMHIs; however, there was clear variability within the sample. Understanding who these technologies are suited to, at what time, and in what context remains a critical area of future research to overcome the limitations of a one size fits all approach.

Second, we cannot guarantee that young people from the general population were not users of services or did not experience mental health issues. Indeed, the high rates of depression and anxiety reported in our general population sample indicate a potential need for care. However, we did not ask participants about their help seeking. Notably, these levels of mental health concerns match those of surveys conducted on the general Australian youth population during the pandemic, supporting the representativeness of the sample [53]. Third, there is an important distinction between clinicians’ recommendations for young people to use DMHIs and their use to support clinical care activities (ie, blended therapy). Additional research is required to gain insight into how clinicians use technology within the mental health treatment they provide and what technologies are most appealing to support their clinical work. Fourth, this survey was conducted at a time during which strict lockdown measures were instituted in Australia, limiting daily activities. Although technology use rates in this study were similar to populations of young people before the pandemic [4], it is possible that use rates increased in this sample during this time, as well as increased demand and interest in mental health support because of increased stress.

Finally, Australia is a high-income country with mental health services supported by government funding. Youth mental health services are free for young people, although capacity limitations and geographical barriers limit access to them. These results may not be generalizable to countries with more limited youth mental health services, in which the demand for and interest in DMHIs may be higher [54,55]. However, these findings establish a strong case in which young people across a spectrum of clinical needs are interested in DMHIs, and most have access to the technologies required to receive them.

Conclusions

The global pandemic has brought forth a critical juncture in developing a new system of digitally enabled care that is aligned with the needs of those it intends to support. These findings provide valuable insights into the perspectives of clinicians and young people as end users of digital mental health technologies and provide a compelling case for further development and expansion of technologies to enhance youth mental health care.
young people and clinical staff who gave their time to complete the survey. This work was supported by Future Generation Global and a Victorian Government Innovation Grant.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Full survey on digital mental health use and interest among young people and clinicians.

References


Abbreviations

DMHI: digital mental health intervention
GAD-2: General Anxiety Disorder-2
PHQ-2: Patient Health Questionnaire-2
VR: virtual reality
Therapeutic Alliance in Online and Face-to-face Psychological Treatment: Comparative Study

Josep Mercadal Rotger¹, PhD; Victor Cabré¹, PhD
Institut Universitari de Salut Mental Vidal i Barraquer, Universitat Ramon Llull, Barcelona, Spain

Corresponding Author:
Josep Mercadal Rotger, PhD
Institut Universitari de Salut Mental Vidal i Barraquer
Universitat Ramon Llull
C/ Sant Gervasi de Cassoles 88-90
Barcelona, 08022
Spain
Phone: 34 660221557
Email: jmercadal@fvb.cat

Abstract

Background: Since the COVID-19 pandemic, the number of online mental health treatments have grown exponentially. Additionally, it seems inevitable that this technical resource is here to stay at health centers. However, there is still very little scholarly literature published on this topic, and therefore, the impact of the changes that have had to be dealt with in this regard has not been studied.

Objective: This study aims to evaluate the differences in the establishment of the therapeutic alliance (TA) based on the intervention modality (online or face-to-face), the type of attachment, and diagnosis.

Methods: A total of 291 subjects participated in the study, 149 (51.2%) of whom were men and 142 were (48.8%) women between the ages of 18 and 30 years. The instruments used were sociodemographic data, SOFTA-o (System for Observing Family Therapeutic Alliances—observational), and Relationship Questionnaire.

Results: The results show that the treatments conducted face-to-face obtain significantly better scores in the creation of the TA than those conducted online ($t = -42.045, df = 289, P < .001$). The same holds true with attachment, in that users with secure attachment show a better TA than those with insecure attachment ($t = 6.068, P < .001$), although there were no significant differences with the diagnosis ($F = 4.566, P = .44$), age ($r = 0.02, P = .70$), and sex ($t = 0.217, P = .33$).

Conclusions: We believe that professionals are not yet prepared to conduct remote treatment with a degree of efficacy similar to that of face-to-face. It is essential for professionals to receive training in this new technical resource and to understand and incorporate the variants it entails into their daily practice.

(JMIR Ment Health 2022;9(5):e36775) doi:10.2196/36775

KEYWORDS
online psychological intervention; therapeutic alliance; digital health; mental health; mental health education; mental health treatment; health interventions; health professional; online health; web-based health; intervention modality

Introduction

Background
It seems inevitable that online psychological treatments are here to stay in mental health centers and services. The pandemic caused by COVID-19 has accelerated their advent and normalization among mental health professionals, forcing most of their psychotherapeutic activity to shift to the online methodology. Therefore, in a brief period of time, therapists and patients have had to adapt to conditions that forced them to change certain variables, especially the setting, without prior planning or awareness of what other changes they would have to grapple with besides technological ones [1]. Nonetheless, the future of online and face-to-face treatments, once the health crisis is over, is still unclear.

Some authors [2,3] claim that online modalities have facilitated the availability of mental health services during the pandemic. Acero et al [4] further claim that online treatment has facilitated access to mental health services not only in situations caused by COVID-19 but also for people living in rural environments.
or far from urban nuclei. In this sense, several studies have been published, which conclude that online psychological treatments during the pandemic have led to significant improvements in patients’ concerns with COVID-19 and a significant drop in symptoms such as anxiety, depression, and insomnia [5,6]. However, there is still a lot of room to study the differences between the therapeutic alliance in online and face-to-face psychological treatment in terms of efficacy and quality.

**Online Psychological Interventions**

Different authors [7,8] warn that the use of these digital resources is not without consequences in the patient-therapist relationship and that therapists should use these new communicative devices with a great deal of care and knowledge, especially not knowing the risks they could entail for the patient and the therapeutic relation. In fact, some authors agree that there is limited knowledge about the feasibility and acceptability of eHealth interventions in relation to the clinical characteristics of certain types of patients, such as psychotics. The studies by these authors conclude that the level of acceptance among patients with psychosis is high and offer evidence that both online interventions and the use of artificial intelligence can serve as a profitable, accessible, and effective therapeutic agent [9,10].

In some countries such as Brazil, online psychological treatment may only be carried out if the purpose is to research its efficacy [11], with the argument that this new technical resource may have limitations or legal or ethical problems related to its practice. Other countries such as Italy claim that many professionals are not prepared either methodologically or technologically for the change from traditional therapies to digital or online therapies [12]. In another study also conducted in Italy, only 18.3% of the therapists reported having experience with online treatments, and even though 62.6% of the psychologists were in favor of online treatment, they saw many limitations and had many reservations about ethical and legal issues, in addition to technical and methodological ones [13]. In this sense, De la Torre and Pardo [14] do not recommend holding online sessions at times of crisis or under specific conditions such as a lack of emotional control characteristic of people with psychotic disorders, severe depression, or situations of severe violence and abuse, among others, as they must be addressed in a specific way and in some cases by a multidisciplinary team. In fact, in a study conducted in Germany, therapists claim that treatment conducted face-to-face is much more efficacious than online treatment [15].

Rollman et al [16] conducted a study in which they compared the application of online and face-to-face treatment in a sample of 704 patients who had anxiety and depression; they concluded that online therapy did not provide any additional benefit over face-to-face therapy. However, Rathenau et al [17] affirm that the main predictive factor of the efficacy of online treatments is the therapist’s attitude toward it. Other authors claim that live, face-to-face human treatment is not comparable to online treatment, and that while at times it can be a good resource and even a good complement, under no circumstances can it be better and “more real” than face-to-face treatment [7]. In this sense, Knaevelsrud and Mearcker [18] cautioned that we know little about how the therapeutic relationship evolves over the internet and whether it influences the outcome of the treatment, as it does in traditional face-to-face treatments. However, the meta-analysis carried out by Lin et al [19], in which the findings between teletherapy and in-person therapy were compared, concluded that there were no significant differences between teletherapy and face-to-face therapy in the results at posttreatment (g=0.043), at follow-up (g=0.045), or in attrition rates (rate ratio=1.006). In addition, the within-group findings showed that teletherapy produced a large reduction in symptoms at posttreatment (g=1.026) and at follow-up (g=1.021). Thus, these findings provide empirical support for the practice of teletherapy, and client outcomes in teletherapy do not differ from in-person versions of treatments.

**Therapeutic Alliance**

The TA is one of the most investigated variables related to success in psychological interventions, regardless of the theoretical orientation. Many authors affirm that the TA is the main predictor variable of results in mental health treatments [20-24]. Bordin [25] proposes that the TA has three components: agreement between therapist and patient about the goals of therapy; agreement on the tasks necessary to achieve those goals; and affective bond between therapist and patient, necessary to withstand the difficulties of therapeutic change. For Muran [26], the TA implies that an intersubjective negotiation between patient and therapist about the needs and desires of the other underlies all treatment. Luborsky et al [27] also made interesting contributions by distinguishing two phases in the development of the TA. At the beginning of treatment, the Type I alliance implies that the patient trusts that the treatment will help, and the therapist offers a warm, supportive, and caring relationship. Both aspects create the conditions for the treatment to start and develop. Later, the Type II alliance is based on joint effort to overcome difficulties and bring about change. This implies trust and commitment on the part of the patient and a solid experience of collaboration with the therapist.

In this sense, there is still no certainty as to whether the establishment of the TA in online interventions is as powerful as in face-to-face interventions. However, a study by Anderson et al [28], in which the differences in the establishment of the TA in adolescents with anxiety were studied, the results showed that the adolescents did not report differences between those who had received face-to-face treatment and those who had received it online. Along the same lines, in a systematic review that evaluated the differences in the establishment of the TA between web-based and face-to-face interventions, it was concluded that the quality of the TA established in web-based interventions is, at least, the same as in face-to-face interventions. In addition, it also indicated that there was a relationship between the TA and the results of the interventions [29]. Flückiger et al [30] conducted a meta-analysis in which they collected 295 independent studies that covered more than 30,000 patients in online and face-to-face treatment. The study investigated the relationship between TA and treatment outcome. The results indicated that a good TA was a predictor of better therapeutic results in both treatments (online and face-to-face).
However, the results were significantly better in face-to-face treatments than in web-based treatments. There is also another meta-analysis carried out by Kaiser et al [31], which aimed to summarize the association between TA and outcome in therapist-assisted online interventions. Overall, 51 effect sizes were extracted from 20 included studies. The average weighted effect size is \( r = 0.203 \) (\( P < 0.001 \)). The correlation was larger when alliance was measured near the end of an intervention. There was no impact of therapist contact frequency or mode and availability of self-help content on the effect size. Therefore, it is concluded that TA and outcome are significantly correlated in web-based therapy. That is, it highlights the importance of a stable alliance in web-based interventions and suggests that fostering the alliance could be beneficial for treatment success.

**Therapeutic Alliance and Attachment**

Attachment theory provides a model for understanding development within the context of the child’s primary and formative relationships, on the one hand, and an adult’s orientation toward lifelong intimate connections and social relationships, on the other. Researchers in psychotherapy have linked measures of patient attachment to the therapeutic alliance, therapeutic process, and therapeutic outcomes. The attachment organization and the therapist’s ability to mentalize play an important role in establishing a good therapeutic alliance and, therefore, in therapeutic success [32].

Smith et al [33] conducted a systematic review of research that has examined the relationship between self-reported patterns of attachment and TA. The results suggest that patients who rate themselves as having a more secure attachment pattern are likely to rate the alliance as stronger. The idea is that patients project their internal working models onto the therapist and the therapist-patient relationship, so that the patient’s attachment patterns affect how the two parties interact with each other and thus the formation and maintenance of their TA [34]. Patients who have a secure attachment are better able to engage in self-exploration, engage in self-disclosure, develop collaborative understanding with the therapist, and be able to reflect on and evaluate their past and current relationships [35]. These skills would help securely attached patients to form a good-quality TA and maintain it by repairing any breaks that develop. Conversely, patients with an insecure attachment pattern may avoid interpersonal closeness with the therapist or worry about the therapist’s investment in them. As a result, this can prevent or delay the formation of a good quality TA [33,36,37].

Daniel [38] advances the idea that therapeutic change occurs when insecure clients, contrary to their previous experience, experience a supportive and responsive relationship with their therapist. If this experience deviates significantly from the individuals’ early prototype model, their central attachment pattern may change. Consistent with this idea, studies have reported that decreases in symptom severity during psychotherapy are associated with increases in self-reported secure attachment [39,40].

This is the context within which we set out to conduct this study, whose main objective is to evaluate the differences in the establishment of the TA in online compared to face-to-face treatments.

Likewise, we shall also evaluate the subjects’ type of attachment and what effects this has on the establishment of the TA.

**Methods**

**Participants**

A total of 291 subjects participated in this study anonymously and voluntarily, 149 (51.2%) of whom were men and 142 (48.8%) women. The subjects were between the ages of 18 and 30 years, with a mean age of 23.1 (SD 2.82; Table 1).

The participants came to the psychological guidance and consulting service voluntarily and free of charge and were invited to participate in the study. The main objective of this service is to psychologically assess or explore the users from 2 universities in Barcelona, and if needed, to refer them to the corresponding services in the public health care network. Participants who were involved in fewer than 3 sessions were excluded.

**Instruments**

The participants responded to the following questionnaires: (1) sociodemographic data—sociodemographic data such as sex, age, whether the treatment was online or face-to-face, and the diagnostic was collected ad hoc; (2) therapeutic alliance—SOFTA-o (System for Observing Family Therapeutic Alliances—observational) for patients [41]; this instrument was created simultaneously in English and Spanish as a transtheoretical tool for research and practice on the TA. In this case, the patient version was used. The measure is based on three dimensions: engagement in the process, emotional connection, and safety. It also provides an overall score. The 12 items, both negative and positive, are related to patients’ behaviors, which are grouped within these 3 dimensions; and (3) attachment—Relationship Questionnaire is a brief self-report that was developed by Bartholomew and Horowitz [42] to evaluate adults’ attachment style based on continuous measures and categorical results. First, the person being evaluated is presented with four prototypical descriptions of the types of attachment in Bartholomew’s model (secure, anxious-preoccupied, dismissive-avoidant, and fearful-avoidant) and is asked to decide with which one they identify the most. Secondly, they are asked to rate their degree of agreement with each of the prototypical definitions of attachment on a 7-point Likert scale [43-47].

**Procedure**

All the subjects filled out the SOFTA-o and the Relationship Questionnaire before the exploration began and filled out only the SOFTA-o after it. It is understood that the TA with the therapist will change if the exploration was a positive experience, but the type of attachment will not, as this construct is stable over time.

The explorations lasted between 3 and 5 sessions. The subjects themselves chose whether they wanted to be treated face-to-face or online. The online interventions were carried out through videoconference.
The subjects filled out the questionnaires individually and independently, and they were only assisted by the researcher if they requested help.

**Ethics Approval**

The study was approved by Research Ethics Committee of the Vidal i Barraquer Mental Health University Institute.

**Results**

**Description of Analyses**

The statistical analyses were conducted using SPSS statistical package (version 27.0, SPSS Inc). First, the descriptive results of the sociodemographic data, the TA, attachment, and the diagnosis were presented. Subsequently, the relations between the TA and the intervention modality, attachment, sex, age, and diagnostic were presented. Next, the mixed model analysis was conducted. To do so, an unstructured variance-covariance matrix was calculated via the restricted estimation of maximum likelihood. The TA before and after treatment, treatment modality, attachment scale, and their interactions were considered fixed effects. Finally, gender and age were also included as fixed factors. The random effect was the subjects’ intersection parameter. The degrees of freedom were calculated with the Satterthwaite approximation. The end model was chosen by recalculating the models with and without interaction via maximum likelihood in order to compare the significance of the change on the Akaike information criterion (AIC). The residuals of the prediction and of the random factor were inspected via a quartile-quartile plot to assess the suitability of the model.

**Descriptive Results of the Sociodemographic Data, Therapeutic Alliance, Attachment, Intervention Modality, and Diagnosis**

As shown in Table 1, the percentage of men and women was almost similar, with 142 (48.8%) women and 149 (51.2%) men. The mean age was 23.1 (SD 2.82) years; 43.6% (n=127) chose the web-based option while 56.4% (n=164) chose face-to-face. The differences were not significant (t=0.210, df=289, P=.91).

The most prevalent diagnosis was anxiety (n=91, 31.3%), followed by depression (n=45, 15.5%) and grief (n=29, 10%). We can also see that 63.6% (n=185) of the participants had a secure attachment, while 36.4% (n=106) had an insecure attachment. Finally, regarding the TA, we see that prior to the treatment, the mean SOFTA-o score of the subjects was 8.62 while after treatment, it was 36.78.

Comparison between age, sex, modality, diagnosis, and attachment in relation to the therapeutic alliance before and after treatment.

We conducted t tests for the variables sex, modality, and attachment; we used the Pearson correlation coefficient for age and TA and ANOVA for the diagnosis.

Via the Pearson correlation coefficient, Table 2 shows significant relations in the scores on the TA at the two times when the questionnaire was administered. We see that between the pre- and postadministrations, there is a correlation of r=0.09 and P<.001.

If we examine the relationship between TA and age, we see that prior to the treatment, there is a correlation of r=0.10 and P=.08, while afterward, it was r=0.02 and P=.70. Therefore, there are no significant differences in the establishment of a better TA according to age.

As we can also see in Table 2, the t test for independent samples revealed that there are no significant differences in the TA prior to the treatment, with the treatment modality (web-based and face-to-face) t=0.150, df=289, P=.88; attachment (secure and insecure) t=–0.835, P=.39; and sex (male and female) t=1.430, P=.16. By contrast, after the treatment, we do find significant differences in the treatment modality t=–42.045, P<.001, and the type of attachment t=6.068, P<.001, but not sex, t=0.217, P=.33. Therefore, we can conclude that the face-to-face modality shows significantly better results in terms of establishing a good TA compared to web-based treatments. The same holds true for attachment, where having a secure attachment leads to significant differences in the development of a better TA.

Finally, regarding the diagnosis, we conducted an ANOVA to determine whether there were differences in the establishment of a better TA by diagnosis, and the results both before and after the treatment showed that there are no significant differences (F=1.097, P=.37 and F=4.566, P=.44, respectively; degrees of freedom between groups, within groups, and total were 9, 281, and 290, respectively; Tables 2-4).
### Table 1. Descriptive results.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>23.1 (2.82; 18-29)</td>
</tr>
<tr>
<td>Pre-TA(^a) scores, mean (SD; range)</td>
<td>8.6 (3.03; 3-18)</td>
</tr>
<tr>
<td>Post-TA scores, mean (SD; range)</td>
<td>36.8 (13.88; 11-56)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>149 (51.2)</td>
</tr>
<tr>
<td>Female</td>
<td>142 (48.8)</td>
</tr>
<tr>
<td>Modality, n (%)</td>
<td></td>
</tr>
<tr>
<td>Web-based</td>
<td>127 (43.6)</td>
</tr>
<tr>
<td>Face-to-face</td>
<td>164 (65.4)</td>
</tr>
<tr>
<td>Attachment, n (%)</td>
<td></td>
</tr>
<tr>
<td>Secure</td>
<td>185 (63.6)</td>
</tr>
<tr>
<td>Insecure</td>
<td>106 (36.4)</td>
</tr>
<tr>
<td>Diagnosis, n (%)</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>91 (31.3)</td>
</tr>
<tr>
<td>Depression</td>
<td>45 (15.5)</td>
</tr>
<tr>
<td>Grief</td>
<td>29 (10)</td>
</tr>
<tr>
<td>Mistreatment</td>
<td>25 (8.6)</td>
</tr>
<tr>
<td>Family problems</td>
<td>16 (5.5)</td>
</tr>
<tr>
<td>Couple problems</td>
<td>16 (5.5)</td>
</tr>
<tr>
<td>Concentration problems</td>
<td>15 (5.2)</td>
</tr>
<tr>
<td>Social relation problems</td>
<td>28 (9.6)</td>
</tr>
<tr>
<td>Adaptation problems</td>
<td>23 (7.9)</td>
</tr>
<tr>
<td>Others</td>
<td>3 (1)</td>
</tr>
</tbody>
</table>

\(^a\)TA: therapeutic alliance.

### Table 2. Therapeutic alliance and age correlation before and after intervention.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Value</th>
<th>P value</th>
<th>Age, r (P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA(^a) before treatment</td>
<td>0.092</td>
<td>&lt;.001</td>
<td>−0.102 (.08)</td>
</tr>
<tr>
<td>TA after treatment</td>
<td>0.092</td>
<td>&lt;.001</td>
<td>0.022 (.70)</td>
</tr>
</tbody>
</table>

\(^a\)TA: therapeutic alliance.
Table 3. Therapeutic alliance comparison between the groups before and after intervention.

<table>
<thead>
<tr>
<th></th>
<th>Tests</th>
<th>Values</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>df</td>
<td>P value</td>
</tr>
<tr>
<td>TA* before treatment</td>
<td>Modality</td>
<td>0.15</td>
<td>268.130</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>Attachment</td>
<td>-0.853</td>
<td>203.183</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>1.403</td>
<td>284.221</td>
<td>.16</td>
</tr>
<tr>
<td>TA after treatment</td>
<td>Modality</td>
<td>-42.045</td>
<td>222.357</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Attachment</td>
<td>6.068</td>
<td>217.342</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>0.22</td>
<td>287.029</td>
<td>.33</td>
</tr>
</tbody>
</table>

*TA: therapeutic alliance.

Table 4. Therapeutic alliance comparison by diagnosis.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>ANOVA</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean square</td>
<td>F</td>
<td>P value</td>
</tr>
<tr>
<td>TA* before treatment</td>
<td>10.084</td>
<td>1.097</td>
<td>.37</td>
</tr>
<tr>
<td>TA after treatment</td>
<td>792.356</td>
<td>4.566</td>
<td>.44</td>
</tr>
</tbody>
</table>

*TA: therapeutic alliance.

Analysis of the Mixed Model

In the model without interactions, the pre-post change in the TA was significant ($t_{576.0}=44.020$, $P<.001$), as was the treatment modality ($t_{576.0}=18.804$, $P=.72$). Age, gender, and attachment did not reach the level of significance ($t_{576.0}=0.492$, $P=.62$; $t_{576.0}=0.17$, $P=.87$; and $t_{576.0}=1.048$, $P=.30$, respectively).

The model with interactions (AIC=3305.5, with 12 parameters) was significantly better ($\chi^2_{4}=742.78$, $P<.001$) than the model without interactions (AIC=4040.3, with 8 parameters). The interaction between the time of the evaluation and the therapeutic modality was highly significant ($t_{287.0}=32.296$, $P<.001$). In the web-based treatment, the mean score on the SOFTA rose by 13.5 points (SD 5), while in the face-to-face treatment, it rose 39.6 points (SD 5.1). The interactions between evaluation and attachment and modality and attachment were not significant ($t_{287.0}=1.248$, $P=.21$ and $t_{534.3}=0.363$, $P=.72$, respectively). In the inspection of the residuals, no gross deviations were found compared to a normal distribution.

Discussion

Principal Findings

The results of this study show that the interventions carried out in person, with a sample of subjects aged between 18 and 30 years, obtain significantly better scores in the creation of the TA compared with those carried out with the web-based methodology. The same occurs with attachment, where users with secure attachment establish a better TA compared with those with insecure attachment. In relation to the variables’ diagnosis, age and sex, there were no significant differences.

Complementary Results (Sample, Diagnosis, and Sociodemographic Data)

First, we should highlight that this is a sample of university students, so we can assume a high sociocultural level with a social network (at least in terms of their belonging to the educational community: teachers, classmates, etc) and a certain predisposition to establish relational bonds (at least with referents in education). Likewise, they belong to an age group with knowledge and skills of the new technologies and therefore have a low level of interference and inconvenience associated with the use of this variable.

In terms of the modality chosen, the members of the sample distributed themselves in a balanced fashion (43.6% web-based and 56.4% face-to-face), with a slight preference for face-to-face treatment. We may think that this may be a pattern that is tending to gain ground in this age group, in a socioeconomic milieu that enables them to have sufficient technological resources and in a metropolitan setting that minimizes the difficulties of access to face-to-face encounters (remote residences, precarious environments, etc). It is likely that based on the experience of the pandemic, these patients’ initiative, at least in initial contacts, includes both methodologies. The fact that there was a slight predominance of those who requested face-to-face treatment seems to reflect the caregiving logic, in which the vast majority of conflicts associated with mental health directly imply other people with whom one has interactions in face-to-face settings (family, friends, partner, etc). In fact, Cabré and Mercadal [8] claimed that live treatment in person is not comparable to web-based treatment, even though at times it may be a good resource or be complementary; however, under no circumstances can it be better and “more real” than face-to-face. Nonetheless, the significant percentage
Attachment and Therapeutic Alliance

The results of our study show that participants with secure attachment developed better TA compared with those with insecure attachment. These results are in line with other research projects in which it is concluded that a secure attachment predicts a better TA [33,35]. In fact, there are also studies that provide the same conclusions from the opposite side: people with insecure attachment have worse TA [36,37]. In our opinion, it is logical that the subjects with secure attachment develop a better TA. Therefore, if we see that the TA consolidates over the course of a few sessions (which comprise the exploration), it is likely that in a new (and hypothetical) choice to continue treatment via safety or distance measures (web-based) would prefer to see the therapist face-to-face as they feel safer, less threatened, and more trusting of the other and themselves.

In terms of the diagnosis, it is difficult to establish patterns with such a general and unspecific set of symptoms. Nonetheless, it is likely that in symptoms in which clinically active depressive features predominate, the first choice will tend to be the contact that is the “easiest” and entails the least effort, which is apparently the online connection (even though these same clinical components may respond better to closer human contact). In grief (even though it also contains these components of sadness and anhedonia), we may believe that the need to have close contact with the other and receive affection from them, without filters, may push the demand for face-to-face over web-based treatment. Finally, in symptoms in which anxious contact predominates, especially regarding human or relational contact (eg, social phobias, separation anxieties, and persecutory anxieties), the first choice may be heavily conditioned by this experience, defensive strategies will probably predominate, and thus web-based methodologies may be preferred. In fact, 73.2% (n=23) of the subjects in our study with a diagnosis of depression chose the web-based modality, 67.8% (n=20) of those who were grieving chose the face-to-face modality, and 59.6% (n=54) of the subjects who had anxiety preferred the web-based option.

Therapeutic Alliance and Modality (Web-Based or Face-to-face)

When we compared at the moment before the intervention if there were differences between TA and modality (web-based or face-to-face), the results showed that there were no differences between these two groups. However, when comparing the modality and the TA at the time after the intervention, the results showed that the face-to-face modality presents significantly better results when establishing a good TA, compared to web-based interventions. These results dispute the conclusions reached by investigations such as that of Anderson et al [28], in which it was concluded that there were no differences between web-based and face-to-face interventions when establishing a good TA.

Finally, we see how the interaction at the time of evaluation and the therapeutic modality were significant; indeed, we found that the score on the TA was 3 times higher in the face-to-face modality (39.6), compared with that in the web-based modality (13.5). Therefore, even though the outcomes may be quite positive in the web-based modality (since it is assumed that TA has improved throughout the intervention), these results are contrary to those reported by Sucala et al [29], who conclude that the quality of TA in web-based interventions is, at least, the same as in face-to-face interventions. However, several studies [20-24,30,31] state that TA improves therapeutic results both in face-to-face and web-based interventions. In fact, Eichenberg [15] stated that face-to-face is more effective than the web-based modality and therefore should be used whenever possible.

Furthermore, we believe that it is obvious, as Tullio et al [12] noted, that professionals are not yet prepared to conduct remote treatment with a degree of efficacy similar to that of face-to-face. Furthermore, only 18.3% of therapists reported having experience with web-based interventions, although 62.6% are in favor of them [13].

For all these reasons, we believe that, as Mercadal and Cabré [1] stated, it is essential for professionals to receive training in this new technical resource and to understand and incorporate the variants it entails into their daily practice.
Acknowledgments

The authors wish to acknowledge the contributions and support of Manel Salamero throughout his entire professional career and, specifically, in this article.

Conflicts of Interest

None declared.

References


**Abbreviations**

- **AIC**: Akaike information criterion
- **SOFTA-o**: System for Observing Family Therapeutic Alliances—observational
- **TA**: therapeutic alliance
Engagement, Predictors, and Outcomes of a Trauma Recovery Digital Mental Health Intervention: Longitudinal Study

Carolyn M Yeager¹, MA, MSc, PhD; Charles C Benight¹, PhD

Lyda Hill Institute for Human Resilience, University of Colorado Colorado Springs, Colorado Springs, CO, United States

Abstract

Background: Worldwide, exposure to potentially traumatic events is extremely common, and many individuals develop posttraumatic stress disorder (PTSD) along with other disorders. Unfortunately, considerable barriers to treatment exist. A promising approach to overcoming treatment barriers is a digital mental health intervention (DMHI). However, engagement with DMHIs is a concern, and theoretically based research in this area is sparse and often inconclusive.

Objective: The focus of this study is on the complex issue of DMHI engagement. On the basis of the social cognitive theory framework, the conceptualization of engagement and a theoretically based model of predictors and outcomes were investigated using a DMHI for trauma recovery.

Methods: A 6-week longitudinal study with a national sample of survivors of trauma was conducted to measure engagement, predictors of engagement, and mediational pathways to symptom reduction while using a trauma recovery DMHI (time 1: N=915; time 2: N=350; time 3: N=168; and time 4: N=101).

Results: Confirmatory factor analysis of the engagement latent constructs of duration, frequency, interest, attention, and affect produced an acceptable model fit ($\chi^2 = 8.3; P = .02; \text{comparative fit index } 0.973$; root mean square error of approximation 0.059; 90% CI 0.022-0.103). Using the latent construct, the longitudinal theoretical model demonstrated adequate model fit (comparative fit index 0.929; root mean square error of approximation 0.052; 90% CI 0.040-0.064), indicating that engagement self-efficacy ($\beta = .35; P < .001$) and outcome expectations ($\beta = .37; P < .001$) were significant predictors of engagement ($R^2 = 39\%$). The overall indirect effect between engagement and PTSD symptom reduction was significant ($\beta = -.065; P < .001; 90\% \text{ CI } -0.071 \text{ to } -0.058$). This relationship was serially mediated by both skill activation self-efficacy ($\beta = .80; P < .001$) and trauma coping self-efficacy ($\beta = .40; P < .001$), which predicted a reduction in PTSD symptoms ($\beta = -.20; P = .02$).

Conclusions: The results of this study may provide a solid foundation for formalizing the nascent science of engagement. Engagement conceptualization comprised general measures of attention, interest, affect, and affect produced an acceptable model fit ($\chi^2 = 8.3; P = .02; \text{comparative fit index } 0.973$; root mean square error of approximation 0.059; 90% CI 0.022-0.103). Using the latent construct, the longitudinal theoretical model demonstrated adequate model fit (comparative fit index 0.929; root mean square error of approximation 0.052; 90% CI 0.040-0.064), indicating that engagement self-efficacy ($\beta = .35; P < .001$) and outcome expectations ($\beta = .37; P < .001$) were significant predictors of engagement ($R^2 = 39\%$). The overall indirect effect between engagement and PTSD symptom reduction was significant ($\beta = -.065; P < .001; 90\% \text{ CI } -0.071 \text{ to } -0.058$). This relationship was serially mediated by both skill activation self-efficacy ($\beta = .80; P < .001$) and trauma coping self-efficacy ($\beta = .40; P < .001$), which predicted a reduction in PTSD symptoms ($\beta = -.20; P = .02$).

Conclusions: The results of this study may provide a solid foundation for formalizing the nascent science of engagement. Engagement conceptualization comprised general measures of attention, interest, affect, and affect produced an acceptable model fit ($\chi^2 = 8.3; P = .02; \text{comparative fit index } 0.973$; root mean square error of approximation 0.059; 90% CI 0.022-0.103). Using the latent construct, the longitudinal theoretical model demonstrated adequate model fit (comparative fit index 0.929; root mean square error of approximation 0.052; 90% CI 0.040-0.064), indicating that engagement self-efficacy ($\beta = .35; P < .001$) and outcome expectations ($\beta = .37; P < .001$) were significant predictors of engagement ($R^2 = 39\%$). The overall indirect effect between engagement and PTSD symptom reduction was significant ($\beta = -.065; P < .001; 90\% \text{ CI } -0.071 \text{ to } -0.058$). This relationship was serially mediated by both skill activation self-efficacy ($\beta = .80; P < .001$) and trauma coping self-efficacy ($\beta = .40; P < .001$), which predicted a reduction in PTSD symptoms ($\beta = -.20; P = .02$).

Conclusions: The results of this study may provide a solid foundation for formalizing the nascent science of engagement. Engagement conceptualization comprised general measures of attention, interest, affect, and affect produced an acceptable model fit ($\chi^2 = 8.3; P = .02; \text{comparative fit index } 0.973$; root mean square error of approximation 0.059; 90% CI 0.022-0.103). Using the latent construct, the longitudinal theoretical model demonstrated adequate model fit (comparative fit index 0.929; root mean square error of approximation 0.052; 90% CI 0.040-0.064), indicating that engagement self-efficacy ($\beta = .35; P < .001$) and outcome expectations ($\beta = .37; P < .001$) were significant predictors of engagement ($R^2 = 39\%$). The overall indirect effect between engagement and PTSD symptom reduction was significant ($\beta = -.065; P < .001; 90\% \text{ CI } -0.071 \text{ to } -0.058$). This relationship was serially mediated by both skill activation self-efficacy ($\beta = .80; P < .001$) and trauma coping self-efficacy ($\beta = .40; P < .001$), which predicted a reduction in PTSD symptoms ($\beta = -.20; P = .02$).

Conclusions: The results of this study may provide a solid foundation for formalizing the nascent science of engagement. Engagement conceptualization comprised general measures of attention, interest, affect, and affect produced an acceptable model fit ($\chi^2 = 8.3; P = .02; \text{comparative fit index } 0.973$; root mean square error of approximation 0.059; 90% CI 0.022-0.103). Using the latent construct, the longitudinal theoretical model demonstrated adequate model fit (comparative fit index 0.929; root mean square error of approximation 0.052; 90% CI 0.040-0.064), indicating that engagement self-efficacy ($\beta = .35; P < .001$) and outcome expectations ($\beta = .37; P < .001$) were significant predictors of engagement ($R^2 = 39\%$). The overall indirect effect between engagement and PTSD symptom reduction was significant ($\beta = -.065; P < .001; 90\% \text{ CI } -0.071 \text{ to } -0.058$). This relationship was serially mediated by both skill activation self-efficacy ($\beta = .80; P < .001$) and trauma coping self-efficacy ($\beta = .40; P < .001$), which predicted a reduction in PTSD symptoms ($\beta = -.20; P = .02$).

Conclusions: The results of this study may provide a solid foundation for formalizing the nascent science of engagement. Engagement conceptualization comprised general measures of attention, interest, affect, and affect produced an acceptable model fit ($\chi^2 = 8.3; P = .02; \text{comparative fit index } 0.973$; root mean square error of approximation 0.059; 90% CI 0.022-0.103). Using the latent construct, the longitudinal theoretical model demonstrated adequate model fit (comparative fit index 0.929; root mean square error of approximation 0.052; 90% CI 0.040-0.064), indicating that engagement self-efficacy ($\beta = .35; P < .001$) and outcome expectations ($\beta = .37; P < .001$) were significant predictors of engagement ($R^2 = 39\%$). The overall indirect effect between engagement and PTSD symptom reduction was significant ($\beta = -.065; P < .001; 90\% \text{ CI } -0.071 \text{ to } -0.058$). This relationship was serially mediated by both skill activation self-efficacy ($\beta = .80; P < .001$) and trauma coping self-efficacy ($\beta = .40; P < .001$), which predicted a reduction in PTSD symptoms ($\beta = -.20; P = .02$).

Conclusions: The results of this study may provide a solid foundation for formalizing the nascent science of engagement. Engagement conceptualization comprised general measures of attention, interest, affect, and affect produced an acceptable model fit ($\chi^2 = 8.3; P = .02; \text{comparative fit index } 0.973$; root mean square error of approximation 0.059; 90% CI 0.022-0.103). Using the latent construct, the longitudinal theoretical model demonstrated adequate model fit (comparative fit index 0.929; root mean square error of approximation 0.052; 90% CI 0.040-0.064), indicating that engagement self-efficacy ($\beta = .35; P < .001$) and outcome expectations ($\beta = .37; P < .001$) were significant predictors of engagement ($R^2 = 39\%$). The overall indirect effect between engagement and PTSD symptom reduction was significant ($\beta = -.065; P < .001; 90\% \text{ CI } -0.071 \text{ to } -0.058$). This relationship was serially mediated by both skill activation self-efficacy ($\beta = .80; P < .001$) and trauma coping self-efficacy ($\beta = .40; P < .001$), which predicted a reduction in PTSD symptoms ($\beta = -.20; P = .02$).

(JMIR Ment Health 2022;9(5):e35048) doi:10.2196/35048

KEYWORDS

engagement; digital health; digital mental health intervention; social cognitive theory; SCT; self-efficacy; outcome expectations; trauma; posttraumatic stress disorder; PTSD
Introduction

Background

The World Mental Health Survey Consortium indicated that >70% of adults are exposed to traumatic events [1]. In the United States, approximately 90% of people are estimated to have at least one exposure to a traumatic event during their lifetime [2]. The COVID-19 pandemic has presented additional challenges resulting in an increased global demand for mental health services, along with increases in trauma exposure in its aftermath [3]. As trauma exposure increases, so do the risks of developing posttraumatic stress disorder (PTSD), along with other mental and physical health conditions [4]. Costs associated with mental health disorders are significant, accounting for approximately 7% of the global burden of disease and 19% of all years lived with disability [5]. Despite its high prevalence and societal costs, treatment coverage remains poor, resulting in a global mental health treatment gap [6]. Rural and underserved communities are especially vulnerable to these barriers, which have been exacerbated in the wake of the COVID-19 pandemic [7].

Digital Mental Health Interventions

A promising approach to overcoming these barriers is the use of technology to reach more people at a low cost in a structured and confidential format [8,9]. Technology to promote mental health and behavior change is referred to as a digital mental health intervention (DMHI). The acceptance of DMHI apps continues to increase, and downloads have risen exponentially since the proliferation of the COVID-19 pandemic [10]. Several systematic reviews have concluded that although DMHIs are growing in popularity, evidence of their efficacy is still limited [10,11]. Research suggests that some inconclusive findings on DMHI effectiveness may be related to a lack of engagement [12]. As engagement may influence intervention outcomes, a greater understanding of engagement and the factors that influence DMHI engagement is essential [13].

However, despite its importance, a consistent engagement conceptualization is lacking [14]. The term engagement, broadly defined as attention, interest, and use of a DMHI [15], has been used in many ways, yielding inconsistent findings and making it challenging to synthesize reliable models and measures. The lack of guidelines or specificity makes it difficult to measure, interpret, and compare the engagement metrics across DMHIs. A recent systematic review concluded that the field of DMHIs depends on user engagement, and the lack of clear definitions and standards can be harmful to the field [16].

This study sought to address this gap by extending previous engagement research [17] to examine a theoretical framework for the conceptualization of engagement, predictors of engagement, and the relationship between engagement and outcomes. Before presenting the model, the engagement conceptualization is described.

Engagement

Most studies agree that engagement includes some interaction with a DMHI [18]; however, there is little agreement as to what exactly engagement is, its bounds, and a precise conceptualization of the concept in general (see Yeager and Benight [19] for a full review). Systematic reviews of engagement research concluded that the definition of engagement must go beyond objective measures of use to include subjective measures of attention, interest, and affect [14-16]. This definition aligns with the social cognitive theory (SCT) framework [20], where observed behaviors (ie, DMHI use), cognitive factors (ie, attention and interest), and personal factors (ie, affect) interact to define the more complex process of engagement. As far as we are aware, this is one of the first studies to explore a multidimensional conceptualization of engagement that includes subjective measures of attention, interest, and affect.

Longitudinal Engagement Research Model

Overview

On the basis of this engagement conceptualization, a longitudinal research model was developed and tested (Figure 1). This model was built on several frameworks [15,17,21] and included predictors of engagement (shown in brown), objective and subjective measures of engagement (shown in purple), and direct and indirect relationships to DMHI outcomes (shown in blue).
**Engagement Predictors**

Predictors of engagement can include countless combinations of user characteristics and DMHI design components [22,23]. Using a theoretical framework is essential to begin to whittle down the most important factors [12]. SCT, which offers a parsimonious, self-regulatory framework for motivating DMHI engagement [20], suggests that behaviors are performed if one perceives confidence in one’s ability, there are few external barriers, and the behavior is worth the effort. Self-efficacy [24] and outcome expectations [25] are key constructs in this self-regulatory process. SCT suggests that predictors of engagement include both personal (eg, appraisals) and external (eg, DMHI characteristics) factors.

Engagement self-efficacy reflects appraisals of one’s ability to initiate and maintain engagement with a DMHI despite barriers associated with using a DMHI [26] that focuses on trauma recovery [27]. Engagement self-efficacy incorporates both confidence in using technology and confidence in addressing traumatic stress symptoms. Individuals high in engagement self-efficacy imagine success in using a DMHI and are more likely to initiate a new behavior, invest more effort, and persist longer than those who are less self-efficacious [28]. These individuals may persist despite challenges associated with using technology and avoidant behaviors, a hallmark symptom cluster of PTSD [29]. Therefore, engagement self-efficacy is predicted to have a positive effect on engagement.

Outcome expectations are the estimation that a given behavior, once performed, will lead to desired outcomes [30]. Outcome expectancy includes DMHI characteristics such as beliefs that the DMHI will be effective [31]. For this study, outcome expectations are defined as perceptions that using the DMHI will increase one’s ability to cope with symptoms associated with their trauma. Low outcome expectations are often cited as barriers to in-person evidence-based treatment for those with PTSD [32]. Therefore, higher outcome expectations are predicted to have a positive effect on DMHI engagement.

**Engagement Outcomes**

Understanding the full picture of DMHI effectiveness must also include the anticipated effects of engagement on important postintervention outcomes. The ambiguity demonstrated in the predictors of engagement research was also found in research examining the relationship between engagement and DMHI outcomes [33,34]. Research has shown that there is a dose-response relationship: the greater the use, the greater the positive effects [17,35]. However, not all DMHIs show this relationship [31,36,37], which may be attributable to several factors, including a lack of engagement consensus and lack of consideration of DMHIs’ mechanisms of action [34].

These mechanisms can serve as mediators between DMHI engagement and desired outcomes. Perski et al [15] found that mechanisms of action include beliefs, knowledge, motivation, self-efficacy, and skill practice. To increase our understanding of these potential mechanisms, our model included 2 serial mediators between engagement and outcomes (Figure 1).

On the basis of the SCT, we examined 2 forms of self-efficacy, measured at subsequent periods, as potential mechanisms of action. According to Bandura [20], self-efficacy is a context-specific assessment of competence in performing a specific task. These different types of task-specific self-efficacy, which is an assessment targeting distinct appraisals across different groups of behaviors, include skill activation and coping with trauma symptoms.

Skill activation self-efficacy refers to a person’s confidence in performing the specific skills taught by the DMHI, and coping self-efficacy for trauma (CSE-T) refers to a person’s confidence in managing their internal and external posttraumatic recovery demands [27]. In our theorizing, greater engagement with the
DMHI enhanced skill activation self-efficacy, which led to subsequent increases in CSE-T. In the proposed model, skill activation self-efficacy was measured after using the DMHI for approximately 1 week (time 2 [T2]), and CSE-T was measured after using the DHMI for approximately 2 weeks (time 3 [T3]). Skill activation self-efficacy reflects confidence in one’s ability to use the new skills learned from the DMHI, despite associated barriers. According to Bandura [20], confidence in practicing DMHI skills precedes the actual use of the skill, and this step is often ignored in the literature. The specific trauma recovery skills taught by the DHMI are relaxation, increasing social support, managing triggers, identifying unhelpful thoughts, and using healthy coping strategies. Practicing these skills may prove to be more challenging to adhere to than expected; however, a self-efficacious person will respond more confidently, with better strategies, effort, and persistence in overcoming such hurdles. We hypothesized that as a person engages more with the DMHI, their confidence in using DMHI skills will also increase.


Study hypotheses
- Hypothesis 1: a relationship exists between the observed subjective and objective engagement variables and their underlying engagement latent construct, as demonstrated by the adequate model fit of the latent construct. The engagement latent construct included objective measures of use and subjective measures of attention, interest, and affect.
- Hypothesis 2: participants with higher engagement self-efficacy would experience higher levels of engagement.
- Hypothesis 3: participants with higher outcome expectations would experience higher levels of engagement.
- Hypothesis 4: skill activation self-efficacy and coping self-efficacy for trauma would serially mediate the relationship between engagement and outcome:
  - Hypothesis 4A: participants with higher levels of engagement would experience higher levels of skill activation self-efficacy.
  - Hypothesis 4B: participants with higher levels of skill activation self-efficacy would experience higher levels of CSE-T.
  - Hypothesis 4C: participants with higher CSE-T would experience a greater reduction in trauma symptoms.

Methods

Overview
A 6-week correlational, longitudinal study was performed using a DMHI with a population of survivors of trauma to test the proposed engagement conceptualization and theoretical model. The study was completed entirely on the web without human interaction to examine engagement in the wild with minimal engagement-related confounds [40].

Participants
To improve external validity, recruitment for the web-based study comprised a national sample of survivors of trauma who had experienced a variety of traumatic events. Specifically, participants were recruited from flyers, social media groups, and the university’s Sona web-based study system. The inclusion criteria were as follows: (1) having experienced at least one traumatic event based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth edition Criterion A [29], (2) English speaking, (3) aged ≥18 years, and (4) having a score >0 on a measure of posttraumatic distress. No other inclusion or exclusion criteria were specified to maximize the number and diversity of respondents.

Sample Size Criteria
Although most researchers advise a sample size of 10 participants for each parameter being estimated [41], the ratio of sample size to free parameters may be as low as 5:1 with certain model specifications (ie, large factor loading and multiple indicators for each latent variable [42]). This study recruited more participants (at time 1 [T1]: 915/1367, 66.93% who met the criteria) to achieve this minimum while allowing for nonuse, dropout attrition, and flexibility to handle missing data and other unanticipated procedural or methodological issues.

Materials: DMHI (My Trauma Recovery)
The DMHI used in this study was My Trauma Recovery (MTR), which was designed by developers to improve an individual’s CSE-T [43]. MTR is a self-guided, theoretically based, DMHI with no interactions with a therapist. MTR is mainly based on SCT [43] and focuses on increasing an individual’s ability to cope with trauma via six self-directed modules: (1) unhelpful ways of coping, (2) relaxation, (3) social support, (4) self-talk, (5) ways of coping, (6) managing stress, and (7) social support. Each module contains several steps, including baseline assessment, targeted skill building, and review and practice opportunities. The modules are designed to be completed in a 6-week period, with each module covering a different aspect of trauma recovery.
(5) trauma triggers and memories, and (6) seeking professional help. There is empirical support for the efficacy of MTR, and its mechanism of action (increasing CSE-T) is understood [39].

Study Design
The overall study design is shown in Figure 2. Participants were assigned a random participant identification number to maintain anonymity and track their progress throughout the study. This study used a longitudinal research design comprising 4 periods. T1 was the baseline, T2 was approximately 1 week after T1, and T3 was approximately 2 weeks after T1. Time 4 (T4) was a follow-up questionnaire sent out 30 days after the completion of the study (approximately 6 weeks after T1).

Figure 2. Fully automated, longitudinal research design. The total duration is 6 weeks. DMHI: digital mental health intervention; T1: time 1; T2: time 2; T3: time 3; T4: time 4.

The study was designed to allow participants enough time to practice the skills they learned from the DMHI. Each period indicated the completion of a survey. To move from T1 to T2 and from T2 to T3, the participants had to complete a minimum of 2 modules of the DMHI. Therefore, the individuals moved through the study at different rates. On average, participants moved from one period to the next every 8 days (from T1 to T2: mean 8.88, SD 13.31 days; from T2 to T3: mean 8.49, SD 7.72 days; from T1 to T3: mean M15.67, SD 13.56 days).

After T3, the use of the DMHI was optional. After finishing a module, participants were asked to complete a short, postmodule survey that measured engagement. To increase adherence, automatic reminder emails were sent to participants throughout the study.

Procedures
Overview
Data were collected over 2 years between August 2018 and August 2020. Recruitment followed three primary strategies: (1) Western university’s Sona system posting to recruit undergraduate students, (2) paid advertising on various social media platforms (Facebook, LinkedIn, and Twitter), and (3) free advertising on various email listservs. Sona participants were compensated with extra credit, and web-based participants were compensated US $15 when they completed all study protocols. All participants were entered in a raffle for one of four US $50 gift cards after completing all study protocols (T1, T2, T3, and T4). A list of resources was provided to participants after completing each survey.

Screening
All participants were provided with a brief statement explaining the procedure and a link to the study on Qualtrics. Participants indicated that they voluntarily agreed to participate in the web-based informed consent form. After the participants provided informed consent via Qualtrics (by pressing the I Agree button), they completed the screening questionnaires. Those who did not meet the criteria (ie, did not experience a traumatic event or were not experiencing any symptoms of posttraumatic distress) were not included in the study.

T1 Procedures
Those who met the inclusion criteria were automatically provided the baseline T1 questionnaire, which included all measures assessing demographics, engagement self-efficacy, outcome expectations, and CSE-T. To increase the validity of the responses, 7 questions were embedded throughout the T1 survey as an attention assessment. These questions asked participants to select a specific response (eg, “For this item, please select ‘None at all’”). A valid case was identified as a participant who answered most (4 of 7) validity questions correctly [44].

All participants who validly completed the T1 survey and met the inclusion criteria were provided access to MTR via email. This email provided participants with a link to MTR and instructions on how to create a user account, log into the site, and begin using the site. Participants were asked to watch an introductory video and use MTR as much as needed but were asked to use a minimum of 2 of the 6 modules to receive compensation. To increase adherence, a second reminder email was sent to the participants 3 days after qualification. This email reminded the participants to create an MTR account as soon as possible and contained instructions to do so.

Engagement
Objective engagement levels were tracked and recorded by MTR throughout the study. Subjective engagement was assessed after the completion of a module.

T2 and T3 Procedures
T2 and T3 used identical procedures and occurred approximately 1 week apart. After completing ≥2 MTR modules, participants were eligible to participate in the next period and were asked to fill out a questionnaire that assessed engagement self-efficacy, activation self-efficacy, CSE-T, and posttraumatic distress. After T3, participants could continue to use MTR as often as needed; however, this was not required.

T4 Procedures
A month after completing the T3 survey (approximately 6 weeks from the start of the study), participants received a request via
email to fill out a brief T4 questionnaire that assessed skill activation self-efficacy, CSE-T, and posttraumatic distress.

Measures

The measures used in this study are described in the following sections and are included in Multimedia Appendix 1. The internal consistencies of the measures are included in the Results section.

Demographics (T1)

Demographic information such as participants’ ethnicity, age, gender, relationship status, income, mental health treatment history, and education was measured.

Traumatic Event (T1)

The presence of a Criterion A traumatic event was assessed using the Life Events Checklist-5 (LEC-5), which is a 17-item self-report measure that assesses exposure to potentially traumatic events across the life span [45]. The LEC-5 was used to determine whether participants had experienced a qualifying traumatic event over the course of their lives. If participants endorsed Happened to me or Witnessed it on the LEC-5, they were eligible for the study. The LEC-5 demonstrated good test-retest reliability (r=0.82) [46].

Engagement Self-efficacy (T1, T2, and T3)

Engagement self-efficacy was measured using 8 questions at the beginning of T1, T2, and T3. During T1, the questions began with the sentence stem “I am confident that I can begin to use My Trauma Recovery...” and during T2 and T3, the same questions began with the sentence stem “I am confident that I can continue to use My Trauma Recovery...” Questions comprised items representing technological and coping-related barriers. The answers were provided on a 5-point Likert scale ranging from 1=not at all confident to 5=very confident.

Outcome Expectancies (T1, T2, and T3)

Outcome expectations were measured at the beginning of each period with 9 questions that started with the sentence stem “If I use My Trauma Recovery regularly I expect that...” Responses were scored on a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree. Cons were reversed scored, and the total score was computed by summing the answers for all items of the positive and negative outcome expectancies scales.

PTSD Symptoms (T1, T2, T3, and T4)

The PTSD checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (PCL-5) [47] was used to measure PTSD symptoms and was anchored to the most relevant trauma on the LEC-5. The PCL-5 was assessed at T1, T2, T3, and the 1-month follow-up. Items assessed symptoms across 4 symptom clusters of PTSD (re-experiencing, avoidance, negative mood, and hyperarousal) on a 0- to 4-point Likert scale. Responses ranged from 0=not at all to 4=extremely. The PCL-5 was scored using the total symptom severity score (range 0-80) by summing the scores for each of the 20 items.

Engagement (T2, T3, and T4)

Overview

Engagement was measured both subjectively and objectively. Subjective self-report engagement measures comprised perceived use and attention, interest, and affect. Objective data were automatically measured from the system logs throughout each period. These measures are described in the following sections.

Engagement Subjective Use (Postmodule)

The subjective perception of use was measured at the completion of each module and included depth (how much of the module did you use [0%-100%]?) and duration (minutes). In addition, frequency of use was assessed at T2, T3, and T4 with the question “How often did you use My Trauma Recovery during the past week?” Answers included 0=never, 1=less than once a week, 2=once a week, 3=twice to three times per week, 4=daily, and 5=more than once a day.

Engagement Subjective Interest and Attention (Postmodule)

The subjective experiences of interest and attention were measured at the completion of each module. Participants were asked to rank how true several statements were while using the DMHI surrounding interest and attention (eg, “I was absorbed”). The answers were provided on Likert scales, ranging from 0=not at all true to 4=extremely true.

Engagement Subjective Affect (T1 and Postmodule)

The subjective dimension of affect was measured at baseline and after the completion of each module using the Positive and Negative Affect Schedule short form. The Positive and Negative Affect Schedule is a 20-item measure that assesses both positive and negative affect [48]. Each subscale comprised 10 items. The answers were provided on Likert scales, ranging from 1=not at all to 5=extremely.

Engagement Objective Use (Continuous)

Objective engagement measures included continuously recorded data that quantified the frequency (number of logins), breadth (number of pages), depth (number of modules completed), and duration (total number of minutes logged in) of the DMHI use. The data were stored in a secure web-based database.

Skill Activation Self-efficacy (T2, T3, and T4)

Skill activation self-efficacy was measured at T2, T3, and T4 using 8 questions that began with the sentence stem “I am confident that I can practice the skills I learned from My Trauma Recovery...” followed by items representing coping-related barriers. The answers were provided on a 5-point Likert scale ranging from 1=not at all confident to 5=very confident.

CSE-T (T1, T2, T3, and T4)

CSE-T was assessed at baseline, T2, T3, and T4 using the CSE-T scale [27]. The CSE-T is a 9-item scale that assesses coping self-efficacy for challenges and demands in the trauma recovery process. Questions such as “I feel capable that I can manage distressing dreams or images about the traumatic experience” were measured on a 7-point scale ranging from 1=not at all capable to 7=totally capable.
Statistical Analyses

Preliminary Analysis

SPSS (version 28; IBM Corp) was used for the demographic and initial analysis. Data were inspected for invalid surveys, outliers, missingness, and other characteristics influencing fit before the analyses. The data were then assessed for outliers and normality.

Missing Data and Attrition

Missing data were estimated with the full information maximum likelihood procedure using AMOS (version 28; IBM Corp). Full information maximum likelihood assumes that data are either missing completely at random (MCAR) or missing at random (MAR) but is also robust when data are missing not at random [49]. The first to be analyzed was the item- or scale-level missing data within each period. Next, tests were performed to analyze missing data because of attrition, which is common in web-based DMHI studies, more lenient criteria were chosen as guidelines for acceptable model fit to the data [57]. Given to the sample size, a CFI >0.90 and RMSEA <0.10 were used as guidelines for acceptable model fit to the data [57]. Given the high attrition rates typically associated with self-directed web-based DMHI studies, more lenient criteria were chosen a priori and were used to evaluate model fit in this study [58].

Reliability

SPSS (version 28) was used to calculate reliability. Cronbach α [53] was used to calculate the internal consistency of the measures, where each item was measured on the same scale (ie, all items were measured on a 0-5-point Likert scale). Cronbach α is based on the essentially tau-equivalent measurement model, which assumes that each item measures the same latent variable on the same scale (variance) but with possibly different degrees of precision [54]. All measures met this assumption apart from the engagement latent construct, which comprised heterogeneous items measured on different scales (eg, pages viewed, interest, affect, and attention) with different SDs. Therefore, the reliability of the engagement latent construct was calculated using the composite reliability (CR) coefficient [55].

MTR Effectiveness

Although not the primary focus of this study, outcomes were analyzed with SPSS (version 28) using repeated-measures ANOVAs. Participants used MTR for approximately 2 weeks, during which their PTSD symptoms were measured at baseline (T1), 1-week (T2), 2-week (T3), and 4-week follow-ups (T4; approximately 6 weeks from baseline).

Fit Indices

This study evaluated and interpreted model fit on two indices in addition to the chi-square value: (1) comparative fit index (CFI) [42] and (2) root mean square error of approximation (RMSEA) [56]. The guidelines for acceptable fit included a nonsignificant chi-square value. However, it should be noted that the chi-square goodness-of-fit test statistic uses traditional statistical significance testing procedures and is highly subject to the sample size. A CFI >0.90 and RMSEA <0.10 were used as guidelines for acceptable model fit to the data [57]. Given the high attrition rates typically associated with self-directed web-based DMHI studies, more lenient criteria were chosen a priori and were used to evaluate model fit in this study [58].

Engagement Measurement Model

To verify the construct validity of the engagement measurement model (hypothesis 1), a confirmatory factor analysis using AMOS (version 28) was used to confirm the factor structure of the set of observed subjective and objective engagement variables. Initially, all observed objective and subjective measures were included, such as objective items measuring the extent of use (minutes, logins, pages, and modules completed) and subjective measures of use, attention, affect, and interest. Items that had a poor factor loading were deleted to improve overall fit unless the deletion compromised the validity of the construct such that it no longer supported engagement conceptualization.

Longitudinal Research Model

The proposed longitudinal structural equation model was tested using AMOS (version 28) to confirm hypotheses 2 to 4 (A, B, and C) [59]. Our model specified 2 exogenous predictors of engagement, a multidimensional engagement latent construct, and skill activation self-efficacy and CSE-T as positive serial mediators between engagement and symptom reduction (Figure 1). To analyze the indirect mediational effects between engagement and outcomes, AMOS (version 28) bootstrapping (2000 samples) analysis was performed with bias-corrected CIs (90% CI).

Ethics Approval

All study materials and procedures were approved by the University of Colorado, Colorado Springs, Institutional Review Board (19-011) before participant contact and data collection.

Results

Preliminary Analysis

Demographics

Table 1 depicts the descriptive statistics of the demographic variables. Of the 1367 participants who signed up for the study, 915 (66.93%) qualified and completed the T1 survey. The participants who met the criteria were 76.9% (704/915) White, 17.8% (163/915) Hispanic or Latino, 7.4% (68/915) Black, 6.4% (59/915) Asian or Pacific Islander, 4.7% (43/915) Native American or Alaskan Native, and 3.3% (30/915) other (some participants specified multiple races). Of the 915 who qualified, 404 (44.1%) created an account on the DMHI, 350 (38.2%) participated in the T2 survey, and 168 (18.4%) participated in T3. Of the 168 participants who completed the T3 survey, 101 (60.12%) completed the T4 1-month follow-up survey.

All participants who met the criteria were directly exposed to ≥1 traumatic event either through experiencing or witnessing the event, including an accident (741/915, 81%), physical assault (564/915, 61.6%), sexual assault (547/915, 59.8%), life-threatening illness or injury (472/915, 51.6%), natural disasters (370/915, 40.4%), sudden unexpected death of someone close (350/915, 38.3%), military combat (72/915, 7.9%), captivity (51/915, 5.6%), and other distress (588/915, 64.3%).
Data were assessed for outliers and normality. A series of comparisons using ANOVA (Cronbach $\alpha$=.05) were conducted to determine whether any relevant differences existed in the variables of interest (eg, engagement variables, activation self-efficacy, and CSE-T) by demographic characteristics (eg, age and education). No significant differences were found. Regarding multicollinearity, there was no correlation between variables above $>0.80$. Therefore, there was no indication of collinearity [60].

Table 1. Descriptive statistics for demographics for time 1, time 2, time 3, and time 4 (N=915)$^a$.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time 1 (N=915)</th>
<th>Time 2 (n=350)</th>
<th>Time 3 (n=168)</th>
<th>Time 4 (n=101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>24.11 (8.53; 18-62)</td>
<td>26.13 (9.39; 18-60)</td>
<td>28.12 (9.04; 18-60)</td>
<td>30.32 (9.08; 18-54)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>698 (76.3)</td>
<td>270 (77.1)</td>
<td>117 (69.7)</td>
<td>67 (66.3)</td>
</tr>
<tr>
<td>Male</td>
<td>205 (22.4)</td>
<td>77 (22.0)</td>
<td>49 (28.9)</td>
<td>34 (33.7)</td>
</tr>
<tr>
<td>Other</td>
<td>10 (1.1)</td>
<td>3 (1)</td>
<td>2 (1.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Relationship status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>646 (70.6)</td>
<td>219 (62.6)</td>
<td>96 (57)</td>
<td>53 (52.4)</td>
</tr>
<tr>
<td>Married</td>
<td>143 (15.6)</td>
<td>73 (20.9)</td>
<td>46 (27.5)</td>
<td>34 (33.7)</td>
</tr>
<tr>
<td>Divorced</td>
<td>48 (5.2)</td>
<td>27 (7.7)</td>
<td>14 (8.4)</td>
<td>10 (9.9)</td>
</tr>
<tr>
<td>Widowed</td>
<td>5 (0.5)</td>
<td>2 (0.6)</td>
<td>1 (0.6)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Other</td>
<td>73 (8)</td>
<td>29 (8.3)</td>
<td>11 (6.5)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Highest education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td>173 (18.9)</td>
<td>57 (16.3)</td>
<td>26 (15.5)</td>
<td>14 (14.1)</td>
</tr>
<tr>
<td>Some college</td>
<td>617 (67.4)</td>
<td>225 (64.3)</td>
<td>98 (58.4)</td>
<td>52 (51.5)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>81 (8.9)</td>
<td>45 (12.9)</td>
<td>31 (18.3)</td>
<td>24 (23.9)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>35 (3.8)</td>
<td>20 (5.7)</td>
<td>13 (7.7)</td>
<td>11 (10.8)</td>
</tr>
<tr>
<td>Sona (vs web-based), n (%)</td>
<td>713 (77.9)</td>
<td>238 (68)</td>
<td>83 (49.3)</td>
<td>29 (28.3)</td>
</tr>
<tr>
<td>Mental health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment (current), n (%)</td>
<td>248 (27.1)</td>
<td>117 (33.5)</td>
<td>62 (36.7)</td>
<td>33 (32.6)</td>
</tr>
<tr>
<td>Treatment (lifetime), n (%)</td>
<td>548 (59.9)</td>
<td>246 (70.3)</td>
<td>114 (67.8)</td>
<td>72 (71.7)</td>
</tr>
<tr>
<td>Baseline PCL-5$^b$, mean (SD)</td>
<td>35.83 (19.10)</td>
<td>40.14 (18.27)</td>
<td>41.98 (18.32)</td>
<td>44.99 (16.97)</td>
</tr>
<tr>
<td>Traumatic event, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency (lifetime)</td>
<td>9.79 (19.01)</td>
<td>10.49 (15.63)</td>
<td>11.36 (16.32)</td>
<td>10.92 (16.46)</td>
</tr>
<tr>
<td>Intensity (0-5)</td>
<td>3.08 (0.93)</td>
<td>3.27 (0.79)</td>
<td>3.95 (0.97)</td>
<td>4.12 (0.91)</td>
</tr>
</tbody>
</table>

$^a$Some percentages did not add up to 100% because of missing data.

$^b$PCL-5: PTSD checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

**Missing Data and Attrition**

The missing data for all relevant items within each period were 0.012% for T1, 0.006% for T2, and 0.005% for T3. As all item-level missing data per period were <1%, they were deemed negligible.

Of the 915 participants who met the criteria for the study, 511 (55.8%) did not create an account (ie, nonuse attrition). Of the 404 participants who created an account, 236 (58.4%) did not complete the T3 protocol. Hence, the nonuse and dropout rates were 55.8% and 58.4%, respectively. To analyze the missing data patterns, the Little MCAR test, a stricter criterion than MAR, was performed with all variables used in the full model (T1, T2, and T3) simultaneously, with age as a reference variable. The results of this test showed that the missing data were not MCAR ($\chi^2_{34}$= 278.2; $P<.001$). This suggests that missing data from T1 to T3 are either MAR or missing not at random; however, there are no definitive tests for these conditions [61].

Independent-sample 2-tailed $t$ tests (equal variances assumed) showed significant baseline differences between noncompleters and completers on the predictors of engagement. Completers reported higher engagement self-efficacy (noncompleters: mean 16.85, SD 7.40; completers: mean 20.71, SD 6.70; $t_{915}$=-6.21; $P<.001$) and higher outcome expectations (noncompleters: mean 21.91, SD 5.21; completers: mean 24.23, SD 4.61; $t_{915}$=-5.33; and $P<.001$).
Multinomial logistic regression was used to identify covariates and interactions that were simultaneously predictive of missingness for the different groups (ie, nonuse, dropouts, and completers). This allows researchers to reasonably assume that the data are MAR. All baseline measures were included (eg, age, education, baseline symptoms, engagement self-efficacy, and outcome expectations), and the results indicated that engagement self-efficacy ($B=-0.07$; odds ratio=0.93; $P=0.003$; 95% CI 0.91-0.97) and outcome expectations ($B=-0.11$; odds ratio=0.90; $P=0.19$; 95% CI 0.88-0.95) were the only significant predictors of dropout or nonuse group membership. As these significant covariates were included in the model, bias because of missingness may be reduced, and the assumptions of maximum likelihood were assumed to be met.

### Reliability

Cronbach $\alpha$ [53] or the CR coefficient [54] was calculated for each measure in the model. The results are shown in Table 2.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Number of items</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Engagement predictors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement self-efficacy</td>
<td>8</td>
<td>.94</td>
<td>.95</td>
<td>.95</td>
<td>—^e</td>
</tr>
<tr>
<td>Outcome expectations</td>
<td>9</td>
<td>.84</td>
<td>.86</td>
<td>.78</td>
<td>—</td>
</tr>
<tr>
<td><strong>Engagement variables and latent construct</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention or interest</td>
<td>7</td>
<td>—</td>
<td>.86</td>
<td>.84</td>
<td>.81</td>
</tr>
<tr>
<td>PANAS$^f$ (positive affect)</td>
<td>10</td>
<td>.89</td>
<td>.94</td>
<td>.93</td>
<td>.94</td>
</tr>
<tr>
<td>Subjective frequency</td>
<td>6</td>
<td>—</td>
<td>.89</td>
<td>.91</td>
<td>.96</td>
</tr>
<tr>
<td>Engagement latent construct (composite reliability)</td>
<td>4</td>
<td>—</td>
<td>.70</td>
<td>.66</td>
<td>.74</td>
</tr>
<tr>
<td><strong>Outcome predictors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activation self-efficacy</td>
<td>8</td>
<td>—</td>
<td>.97</td>
<td>.96</td>
<td>.97</td>
</tr>
<tr>
<td>Trauma coping self-efficacy</td>
<td>9</td>
<td>.91</td>
<td>.92</td>
<td>.93</td>
<td>.93</td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCL-5$^g$</td>
<td>20</td>
<td>.95</td>
<td>.96</td>
<td>.97</td>
<td>.98</td>
</tr>
</tbody>
</table>

^aT1: time 1. ^bT2: time 2. ^cT3: time 3. ^dT1: time 4. ^eIndicates scale not measured during the period. ^fPANAS: Positive and Negative Affect Schedule. ^gPCL-5: PTSD checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

### MTR Effectiveness

The average baseline PTSD symptom severity, as indicated by the PCL-5, was above the 33.00 cutoff value suggested by the National Center for PTSD (T1: mean=43.77, SD 17.43). A repeated-measures ANOVA showed that PTSD symptoms differed significantly between time points (Wilks $\lambda=0.507$; $F_{3,97}=31.42$; $P<.001$), with a large effect size ($\eta^2=0.49$). Post hoc tests using the Bonferroni correction revealed that PTSD was reduced by an average of 11.10 points on the PCL-5 after 1 week ($P<.001$), 13.25 after 2 weeks ($P<.001$), and 18.15 ($P<.001$) after 6 weeks.
Figure 3. Engagement latent confirmatory factor analysis that includes attention, interest, affect, subjective frequency of use, and objective pages viewed. Results of the confirmatory factor analysis provided an adequate model fit and reliability. All regression weights were significant, \( P < .05 \). CFI: comparative fit index; RMSEA: root mean square error of approximation; T2: time 2; TLI: Tucker-Lewis index.

Longitudinal Research Model

Overview

The engagement latent construct was used in the longitudinal research model (Figure 1). Table 3 shows the bivariate correlation coefficients for the study variables included in the tested model. Overall, the model demonstrated good fit (\( \chi^2_{25} = 85.9; P < .001; \) CFI 0.929; RMSEA 0.052; 90% CI 0.040-0.064) and supported hypotheses 2 to 4 (Figure 4). The details of this process are described below.

Table 3. Correlations of variables used in the full structural equation model: T1 \(^a\) (N=915), T2 \(^b\) (n=350), and T3 \(^c\) (n=168)\(^d,e\).

<table>
<thead>
<tr>
<th>Categories</th>
<th>Predictors</th>
<th>Engagement</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>T1 engagement self-efficacy</td>
<td>1</td>
<td>0.489(^f)</td>
<td>0.328(^f)</td>
</tr>
<tr>
<td>T1 outcome expectations</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>T2 subjective affect positive</td>
<td>—</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>T2 subjective interest or attention</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>T2 subjective engagement frequency</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Objective engagement pages viewed</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>T2 skill activation self-efficacy</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>T3 CSE-T(^i)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>T3 to T1 PCL-5(^j)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)T1: time 1.
\(^b\)T2: time 2.
\(^c\)T3: time 3.
\(^d\)Objective measures were continuously measured.
\(^e\)Predictor 1: mean 17.56 (SD 7.43); predictor 2: mean 22.33 (SD 5.18); engagement 3: mean 10.79 (SD 7.14); engagement 4: mean 14.83 (SD 4.45); engagement 5: mean 2.36 (SD 1.07); engagement 6: mean 84.69 (SD 63.42); outcomes 7: mean 26.65 (SD 6.98); outcomes 8: mean 42.52 (SD 10.89); outcomes 9: mean −10.78 (SD 15.86).

\(^f\)\(P < .01\).
\(^g\)Not applicable.
\(^h\)\(P < .05\).
\(^i\)CSE-T: coping self-efficacy for trauma.
\(^j\)PCL-5: PTSD checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.
**Engagement Predictors**

Analysis of the hypothesized predictors of engagement indicated that engagement self-efficacy ($\beta=.35; P<.001$) and outcome expectations ($\beta=.37; P<.001$) were significant positive predictors of engagement ($R^2=39\%$). Adequate fit and significant pathways supported hypotheses 2 to 3.

**Engagement Outcomes**

Regarding the relationship between engagement and outcomes, the direct effect of engagement on changes in PTSD symptoms was nonsignificant ($\beta=.04; P=.58$). However, the indirect serial mediational pathway between engagement and T3 PTSD symptom reduction was found to be statistically significant ($\beta=-.065; P<.001; 90\% CI ~0.071 \text{ to } ~0.058$). Examination of the mediators between engagement and symptom change revealed that engagement was a strong and significant predictor of T2 skill activation self-efficacy ($\beta=.80; P<.001; R^2=64\%$), and skill activation self-efficacy was a significant predictor of T3 CSE-T ($\beta=.40; P<.001; R^2=16\%$). Importantly, the CSE-T significantly predicted PTSD symptom reduction between T1 and T3 ($\beta=-.20; P=.02; R^2=4\%$). Thus, these results provided support for hypotheses 4A, 4B, and 4C (Figure 4).

**Exploratory Model**

In addition to the above model, the second similar model included those who completed the 4-week follow-up (101/915, 11\%). The T1 to T4 model provided an excellent fit ($\chi^2_{25}=49.8; P=.002; \text{CFI} 0.950; \text{RMSEA} 0.033; 90\% CI 0.019-0.046$), suggesting that at the 4-week follow-up, participants continued to show significant improvement through the indirect effects of skill activation self-efficacy and CSE-T. Combined, the above results support hypotheses 1 to 4.
Discussion

Principal Findings
The present research aimed to address the gap in the engagement literature surrounding the definition, measurement, and modeling of engagement with the ultimate goal of understanding ways of effectively increasing engagement. Using a multidimensional definition of engagement that included both subjective and objective components, the proposed conceptualization was tested with a trauma recovery DMHI. As far as we are aware, this is the first study that offers empirical support for a multidimensional definition of engagement. On the basis of the confirmed measurement model of engagement, a theoretically based model of DMHI engagement was tested with a national sample of survivors of trauma. The results confirmed the validity and reliability of the comprehensive engagement measurement model and the relationships between the DMHI engagement, predictors of engagement, and clinical outcomes.

A strength of this study is the variety of trauma-exposed individuals recruited. Trauma experiences included accidents, physical and sexual assault, natural disasters, and military combat. On average, survivors of trauma reported baseline PTSD scores that may be interpreted as above the diagnostic threshold (mean 35.83, SD 19.10), and those who completed the study experienced, on average, a clinically significant reduction in PTSD symptoms (mean 13.24 point reduction on the PCL-5, SE 1.83; P<.001; n=168)) that persisted at the 1-month follow-up (mean 18.15-point reduction on the PCL-5, SE 1.86; P<.001; n=101).

Engagement Measurement Model
The final model demonstrated adequate reliability in this sample (CR=0.70) and included all required components of the proposed definition. A strength of this model is that it is not DMHI specific; rather, it contains general measures of attention, interest, affect, and use that could be applied to other applications, although this has yet to be determined. Another advantage is its parsimony. Measuring subjective experiences while using a DMHI can be burdensome for users [62]. Therefore, short, valid, and reliable measures of engagement may increase compliance. Importantly, the final model did not confound the predictors of engagement (eg, aesthetics and satisfaction) with engagement.

Longitudinal Research Model
Engagement Predictors
The results revealed that 2 exogenous variables, engagement self-efficacy (β=.35; P<.001) and outcome expectations (β=.37; P<.001), were significant predictors of engagement. This confirms previous research, where engagement self-efficacy and outcome expectancies were major determinants of DMHI use [17,63,64]. In these studies, highly motivated participants felt capable of using the DMHI and perceived it as useful and worth the effort.

Engagement Outcomes
The relationship between engagement and outcomes is not well understood [14]. Our model tested 2 different task-specific self-efficacies as serial mediators between engagement and outcomes. Specifically, engagement was found to influence skill activation self-efficacy, where those with higher levels of engagement experienced greater levels of skill activation self-efficacy (ie, belief in the ability to enact skills learned through the DMHI). In turn, higher levels of skill activation self-efficacy predicted higher levels of CSE-T, which mediated an improvement in PTSD symptoms.

Skill activation self-efficacy has been shown to increase health management behaviors in nondigital health care settings, such as heart failure [65], diabetes management [66], and HIV [67], but, as far as we are aware, has never been tested as a DMHI mechanism of action. As predicted by SCT, our results suggest that augmenting beliefs about personal efficacy in DMHI skills practice may be an antecedent to improved confidence in managing posttrauma recovery demands.

Similar to previous research, our study found that CSE-T was the most proximal predictor of symptom improvement. This confirms other studies in which CSE-T mediated posttrauma recovery from several traumatic experiences [68], including accidents [69], sexual abuse [70], life-threatening illnesses [71], and natural disasters [72]. Combined, skill activation and CSE-T mediated the relationship between DMHI engagement and outcomes. This finding is consistent with an extensive literature base that identifies cognitive changes as mediators of mental health symptom improvement (refer to Ehlers et al [73] and Kleim et al [74]). However, the relatively small amount of explained outcome variance (R²=4%) suggests that there may be additional mechanisms of action not included in the model.

Interestingly, the direct pathway between engagement and symptom reduction was not significant after 2 weeks. This supports previous short-term research that failed to find a relationship between engagement and outcomes [75-77] and underscores the importance of understanding and targeting the mechanisms of action between engagement and outcomes to improve DMHI efficacy. Simply increasing engagement to improve outcomes without considering these mediating factors may not suffice.

Limitations
Overview
A major limitation of this model is that it views predictors, engagement, and outcomes as unidirectional processes in which predictors influence engagement and engagement influences the outcomes. According to SCT, behaviors, personal factors, and the environment interact with each other over time (ie, triadic reciprocal causation). It is highly probable that these components operate in a nonrecursive fashion of reciprocal determinism. Further examination of this dynamic framework can reveal how engagement changes and influences predictors and recovery as it unfolds across time. Modeling these dynamic reciprocating processes is beyond the scope of this study and will be investigated in future studies.
Attrition
This study had a high attrition rate, which is consistent with other longitudinal DMHI studies [78,79]. Attrition can cause potential biases and threats to generalizability [51].

Engagement Latent Construct
Several limitations surround the engagement latent construct. As seen in much of the literature, different objective measures have been used to define engagement with disparate results [14,79]. For this study, only 1 objective measure was included in the final model. Researchers have suggested that multiple objective and subjective measures may more accurately represent engagement [80]. However, the equivocal findings in the DMHI engagement literature suggest that the most appropriate measure of use may vary for each DMHI. Further improvements to these measures may be warranted. Although the final model had an excellent fit and supported hypothesis 1, the relatively weak factor loading and explained variance leave some questions for discussion and future research regarding which components are most relevant.

In addition, SCT suggests that predictors and outcomes of engagement influence engagement throughout the DMHI experience [20]. Future research may want to examine the differential and recursive effects of outcomes such as CSE-T and symptom reduction on engagement. An examination of engagement over time may reveal that other measures of engagement may become more influential as users move deeper into the intervention and experience greater (or lesser) changes because of their efforts [81].

Study Design
This study used a longitudinal correlational design, suggesting that cause and effect are only interpreted based on theory and time lag and not on experimental manipulation. The fitted models do not necessarily represent the only true models, and there may be others that also fit the data [41]. Several engagement predictors were not investigated in this study, such as social support [64]. This will be an area for future exploration.

Regarding outcomes, skill activation self-efficacy is assumed to increase the practice of DMHI skills. However, a measure of skill practice was not included in our model. Ideally, an accurate measure of skill practice should be captured through daily ecological momentary assessments. Future studies should incorporate daily ecological momentary assessments of skill use as a mediator in symptom reduction.

DMHI-Related Limitations
In this study, only 1 DMHI was tested that targeted a mental health disorder (PTSD). MTR is a web-based web intervention that does not use many recent advances in digital technology such as social networking, virtual reality, machine learning, sensor technology, and mobile computing. Examination of the engagement measurement model and theoretical models with flexible and novel DMHIs for a variety of mental health issues may help confirm the generalizability of these findings.

Due to the design of MTR, participants were led through each module by way of several predetermined steps (ie, tunneling). The participants generally moved through the intervention at the same rate, which provided limited variability in engagement use patterns. These types of tunnelled interventions have been found to generate more page views than self-paced interventions [79]. However, this may be an artifact of making users click through a prespecified number of pages to progress through the DMHI and may not be at all related to engagement.

Sample
Although a national sample of survivors of trauma was recruited from throughout the United States, most of the participants were White female psychology students enrolled in a Western university.

Implications
Engagement
The findings of this study established the validity and reliability of a multidimensional engagement measurement model, although questions for future research remain. In principle, empirically supported behavioral and experiential dimensions of engagement can be measured in every DMHI. With a valid and reliable measurement of engagement, the therapeutic dose of DMHIs can be established, and the relationships between individual characteristics, engagement, and intervention effectiveness can be better understood. Ultimately, an adequate measure of engagement may provide the opportunity to automatically detect disengagement and help identify the factors that improve engagement.

Theoretical
This study provides a theoretical foundation for understanding numerous predictors of engagement. Although several models could potentially fit the data, the present findings tend to replicate earlier findings in the context of engagement predictors and are in line with the SCT.

Clinical
Importantly, these findings have implications for mental health interventions, whether in person or on the web. Treatment dropout and its causes remain top research priorities in both settings [82]. Improving engagement can potentially lead to improved therapeutic outcomes. By understanding the impact of engagement self-efficacy and outcome expectations, interventions can be designed to enhance these perceptions before treatment, which could, in turn, lead to improved engagement. Skill activation self-efficacy and the CSE-T were shown to mediate the path from engagement to symptom reduction. Although skill training is an essential component of most DMHIs, ensuring that users feel confident in practicing those skills appears to be an important component of DMHI effectiveness.

Directions for Future Research
This research provides a strong foundation for several different explorations surrounding DMHI engagement. Although subjective measures demonstrated strong factor loadings in the engagement measurement model, low response rates to embedded DMHI engagement surveys were common. Combining nonintrusive sensor data with machine learning may
be an important area of research to help alleviate the participant burden [83].

This study also provides support for future research on engagement predictors. We offered a theoretically based predictor model. Future experiments that manipulate predictors of engagement, such as outcome expectations, are encouraged. Logically, engagement alone does not make an intervention effective. Our model revealed the significant indirect effects of engagement on symptom reduction. A further understanding of the mechanisms of action may contribute to overall intervention effectiveness [84]. In theory, these components can be a part of every DMHI.

Conclusions
The empirically supported engagement latent construct and structural equation model provide steps toward formalizing the science of engagement. In turn, this may help improve the design of engaging and effective digital interventions. Unique individual difference variables related to engagement may then emerge, offering a more refined approach to intervention customization.

The therapeutic dose of DMHIs can be established with a valid and reliable measurement of engagement, and DMHI efficacy can be evaluated in a more standardized way. Comparisons among similar DMHIs can then be accomplished through clinical trials to establish the safety and effectiveness of the DMHI. Once established, DMHIs can be designed to increase engagement in early interventions, meet the specific needs of populations, and be used at the exact moment they are needed. Taken together, the future is bright for the role of DMHIs in overcoming significant barriers to care and improving outcomes for a variety of mental health disorders.

Acknowledgments
The authors would like to thank Dr Terrance Boult of the University of Colorado Colorado Springs Computer Science department, who provided important insights into the complexity of engagement with technology and provided the invaluable computer resources necessary to support this research. This research was funded in part by the 2018 Dissertation Research Award from the American Psychological Association.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Measures used in this research study.

References


https://mental.jmir.org/2022/5/e35048 JMIR Ment Health 2022 | vol. 9 | iss. 5 | e35048 | p.143 (page number not for citation purposes)


Abbreviations

CFI: comparative fit index  
CR: composite reliability  
CSE-T: coping self-efficacy for trauma  
DMHI: digital mental health intervention  
LEC-5: Life Events Checklist-5  
MAR: missing at random  
MCAR: missing completely at random  
MTR: My Trauma Recovery  
PCL-5: PTSD checklist for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition  
PTSD: posttraumatic stress disorder  
RMSEA: root mean square error of approximation  
SCT: social cognitive theory  
T1: time 1  
T2: time 2  
T3: time 3  
T4: time 4

Edited by J Torous; submitted 18.11.21; peer-reviewed by W Cao, K Uludag; comments to author 20.12.21; revised version received 28.02.22; accepted 05.03.22; published 02.05.22.

Please cite as:
Yeager CM, Benight CC  
Engagement, Predictors, and Outcomes of a Trauma Recovery Digital Mental Health Intervention: Longitudinal Study  
JMIR Ment Health 2022;9(5):e35048  
URL: https://mental.jmir.org/2022/5/e35048  
doi:10.2196/35048  
PMID:35499857

©Carolyn M Yeager, Charles C Benight. Originally published in JMIR Mental Health (https://mental.jmir.org), 02.05.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on https://mental.jmir.org/, as well as this copyright and license information must be included.
Abstract

Background: Major depressive disorder (MDD) is a global crisis with increasing incidence and prevalence. There are many established evidence-based psychotherapies (EBPs) for depression, but numerous barriers still exist; most notably, access and dissemination. Virtual reality (VR) may offer some solutions to existing constraints of EBPs for MDD.

Objective: We aimed to examine the feasibility, acceptability, and tolerability of using VR as a method of delivering behavioral activation (BA) for adults diagnosed with MDD during a global pandemic and to explore for signs of clinical efficacy by comparing VR-enhanced BA (VR BA) to a standard BA treatment and a treatment as usual control group for individuals diagnosed with MDD.

Methods: A feasibility trial using a 3-armed, unblinded, randomized controlled pilot design was conducted. The study took place remotely via Zoom telehealth visits between April 8, 2020, and January 15, 2021. This study used a 3-week, 4-session protocol in which VR BA participants used a VR headset to complete their BA homework. Feasibility was measured using dropout rates, serious adverse events, completion of homework, an adapted telepresence scale, the Simulator Sickness Questionnaire, the Brief Agitation Measure, and an adapted Technology Acceptance Model. Efficacy was assessed using the Patient Health Questionnaire–9.

Results: Of the 35 participants assessed for eligibility, 13 (37%) were randomized into VR BA (n=5, 38%), traditional BA (n=4, 31%), or a treatment as usual control (n=4, 31%). The mean age of the 13 participants (5/13, 38% male; 7/13, 54% female; and 1/13, 8% nonbinary or third gender) was 35.4 (SD 12.3) years. This study demonstrated VR BA feasibility in participants with MDD through documented high levels of acceptability and tolerability while engaging in VR-induced pleasurable activities in conjunction with a brief BA protocol. No adverse events were reported. This study also illustrated that VR BA may have potential clinical utility for treating MDD, as the average VR BA participant’s clinical severity decreased by 5.67 points, signifying a clinically meaningful change in severity from a moderate to a mild level of depression as per the Patient Health Questionnaire–9 score.

Conclusions: The findings of this study demonstrate that VR BA is safe and feasible to explore for the treatment of MDD. This study documented evidence that VR BA may be efficacious and justifies further examination in an adequately powered randomized controlled trial. This pilot study highlights the potential utility that VR technology may offer patients with MDD, especially those who have difficulty accessing real-world pleasant activities. In addition, for those having difficulty accessing care, VR BA could be adapted as a first step to help people improve their mood and increase their motivation while waiting to connect with a health care professional for other EBPs.

Trial Registration: ClinicalTrials.gov NCT04268316; https://clinicaltrials.gov/ct2/show/NCT04268316

International Registered Report Identifier (IRRID): RR2-10.2196/24331
Introduction

Background

Major depressive disorder (MDD) is a global crisis with increasing incidence and prevalence [1]. Depressive disorders are among the leading drivers of years lived with disability, and those who meet the criteria for MDD experience significant distress or impairment in areas of functioning [1,2].

Many evidence-based treatments have been identified for MDD [3]. Behavioral activation (BA) is considered one of the first-line treatments for MDD as the behavioral theory of depression states that a dearth of response-contingent positive reinforcement catalyzes symptoms of depression owing to less frequent engagement in pleasant activities or behavioral avoidance [3-5]. BA helps those who experience depression become less avoidant and more behaviorally activated by engaging in activities that are pleasurable or lead to a sense of accomplishment, which restores lost positive reinforcement and improves mood.

However, even in pre–COVID-19 times, only 56.8% of people diagnosed with MDD received some type of care to address their symptoms of depression over the course of 12 months [6]. For those who reach out for help, it is estimated that only 37.5% receive minimally adequate or evidence-based treatment [6]. Systemic barriers such as a lack of access to care and long wait times for appointments prevent individuals from engaging in mental health care [7]. Furthermore, there may be external obstacles that prevent those who experience MDD from engaging in BA, such as a lack of resources, financial constraints, physical limitations, and pandemic restrictions. The COVID-19 outbreak and subsequent widespread confinement to one’s home with shelter-in-place and community shutdown orders prevented individuals from partaking in enjoyable activities.

The use of technology as an adjunct to or a method of delivering mental health treatments is becoming increasingly popular, as technology can solve multiple barriers to care and grant increased access to evidence-based care when providers are unavailable [8]. One technology medium, virtual reality (VR), has been successfully used to help treat a variety of mental health conditions, with a study illustrating that VR video 360 was able to elicit similar emotional intensity and feelings of presence to real-life exposures [8,9]. Given the plethora of VR options readily available on the web for free and the cheaper headset selections, VR is now more publicly accessible than in previous years [10], and thus could help eliminate many of the aforementioned barriers to care.

Although using a VR headset presents minimal risk, studies have indicated that users may experience cybersickness, which may include symptoms such as headaches, nausea, dizziness, eye strain, reduced limb control, and reduced postural control [11-13]. However, there are ways to mitigate cybersickness, such as limiting prolonged continuous exposure to the virtual world [11].

Objectives

VR-enhanced psychotherapy may enable increased access to BA by creating solutions to various barriers to engaging in pleasant activities, including pandemic restrictions and social isolation. The primary aim of this study was to examine whether using VR to engage in pleasurable activities within a BA protocol was a feasible, acceptable, and tolerable treatment. In addition, the study explored evidence of clinical efficacy in VR-enhanced BA for MDD compared with traditional BA and a treatment as usual (TAU) control. Finally, this study explored how mood was affected after partaking in a VR activity compared with engaging in an activity in real life.

Methods

Study Design

This was primarily a feasibility study conducted as a preliminary step in deciphering whether VR can be used as a method of delivering pleasurable or mastery activities during BA in a clinical sample of patients with MDD. This study was a 3-arm, nonblinded, between-participant, pilot randomized controlled trial (RCT) created to explore the initial feasibility and efficacy. This study aimed to recruit and enroll 30 participants and took place remotely via Zoom telehealth between April 8, 2020, and January 15, 2021.

Participants

After gaining human-participant consideration and clearance from the Stanford Institutional Review Board (IRB-53483), participants were recruited from a study flyer posted at the Stanford School of Medicine, Department of Psychiatry and Behavioral Sciences, located in Palo Alto, California. The description of the study was also listed on the department’s currently recruiting studies website and on ClinicalTrials.gov. In addition, without solicitation, Curify, a health technology start-up based in San Francisco, placed study advertisements on Facebook without any formal agreement or payment from our research group. The inclusion criteria were as follows: aged ≥18 years; speaking English; and meeting the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, criteria for MDD. The exclusion criteria were as follows: a substance use disorder in the past year, diagnosis of any psychotic or bipolar I disorder, seizure in the last 6 months or untreated epilepsy, current suicidal urges or intent, or current nonsuicidal self-injury or parasuicidal behavior.

Procedures

Overview

The initial screening procedure consisted of 2 steps: an initial phone screen and a face-to-face Zoom intake session. During the phone screen, callers were briefly assessed for initial
eligibility and provided with the opportunity to ask questions about the study. Initial eligibility was determined by a Patient Health Questionnaire–8 score of ≥10 [14] as well as a brief questionnaire. If eligible and still interested in participation, a formal initial intake session was scheduled via Zoom, and the participants were securely emailed the consent form for review before the meeting. After asking any questions and securely emailing the signed consent form back to the protocol director, the intake session occurred. During the Zoom intake session, the participants were asked to verbally complete a demographic questionnaire while the protocol director shared her screen via Zoom. The participants were subsequently administered the Mini-International Neuropsychiatric Interview by the protocol director. The participants were then informed of their eligibility and, if eligible, scheduled for their first session via Zoom. See the previously published case report [15] for further details.

**Randomization**

The participants were randomly assigned to receive one of the 3 study arms in a single-blind fashion by using permuted blocks of 6 in sealed envelopes. A target sample size of 30 patients was selected in keeping with the higher end of the range of sample sizes used for such feasibility studies.

**Intervention**

At the beginning of each session, all participants were verbally administered the Patient Health Questionnaire–9 (PHQ-9). The protocol director shared her screen over Zoom with the participants in the VR BA and traditional BA arms while collecting this measure. The participants in the TAU arm were only read the questions over the phone. If item 9 was endorsed, a risk assessment was conducted in real time, and proper measures were taken in accordance with risk.

**Intervention: Treatment as Usual Arm**

After the 4 meetings were completed, these participants were given the option to meet once with the protocol director via Zoom for 50 minutes so that the protocol director could explain the theory behind BA, provide psychoeducation around pleasurable and mastery activities, and explain how the participants could incorporate BA into their lives. The participants also had the option to receive a Google Cardboard as an incentive to remain in the TAU control group. The protocol director explained how to use the Google Cardboard as a potential method of engaging in pleasurable activities. Only the data accrued during the 4 meetings were used in the study.

**Intervention: VR BA and Traditional BA Arms**

These participants met with the protocol director 4 times, once per week for 3 weeks, over Zoom for 50 minutes to receive BA therapy. The VR participants were shipped a VR headset before the first session. The headset was supplied by Limbix, now partnered with BehaVR. This headset had a 5.5-inch screen size with a resolution of 2560 x 1440 pixels, a screen aspect ratio of 16:9, a field of view of 92 degrees, 3 df, and a refresh rate of 70 Hz. See the previously published case report for further description of the VR device [15]. Both arms followed the protocol for brief BA based on the guidance of published literature [16,17]. The treatment incorporated 4 components: establishing the therapeutic relationship, developing goals for treatment, conducting a functional analysis, and treatment review with relapse prevention [17].

The first session focused on establishing rapport, identifying activities that the participants valued or had felt a sense of mastery or pleasure from in the past, introducing the mood-activity log, and setting activity goals [16]. The traditional BA participants were provided with an in-person activity list and required to schedule *real life* activities, whereas the participants in the VR BA arm were provided with VR activity options and required to choose VR activities for the week. These VR activities consisted of 360-degree videos that did not entail the participants’ active involvement but were simulations of activities that were passively watched, other than allowing the users to change their visual perspectives with head movements. The VR BA participants were also asked to complete a post-VR questionnaire assessing spatial presence, simulator sickness, agitation, and acceptability after each VR activity.

During sessions 2 and 3, the protocol director reviewed the mood-activity log (session 2) and activity schedule (session 3) with the participants and checked in regarding goal attainment to reinforce homework completion [18]. The participants in the VR BA and traditional BA arms were asked to rate their mood on a scale of 1 to 10 (1=worst they ever felt and 10=best they ever felt) before and after their chosen activity. Barriers to completion of activities and problem-solving strategies were again discussed, and new activity goals were introduced and scheduled. During session 4, the treatment and skills were reviewed, and feedback was provided by participants. For further details, see the previously published case report [15].

**Outcomes**

**Feasibility**

Feasibility was assessed using dropout rates, serious adverse events reported, completion of homework, and level of presence experienced in the headset. Sense of presence is a psychological construct and is used as a measure of the ecological validity of VR devices. Sense of presence is defined as a “sense of being there” or a “feeling of being in a world that exists outside of the self” [19]. This presence questionnaire is a validated measure that is correlated with procedural learning enhancement. Dropout rates were assessed using the number of individuals who did not complete the full 4-session protocol after randomization. Serious adverse events were gathered from qualitative interviews and notes. Completion of homework in the VR BA arm was determined by the number of times the headset was used and the number of times the post-VR questionnaire was completed. The number of times the headset was used was obtained from the data collected from the headset after participant termination or completion of the study. The number of post-VR questionnaires completed was calculated from the number of post-VR questionnaires that each participant emailed to the protocol director. The participants were asked to complete ≥4 VR activities per week and a post-VR questionnaire for each VR activity completed. Completion of homework in the traditional BA arm was defined as completing the mood-activity log after session 1 and completing ≥4 activities in real life each week after sessions 2 and 3. These data were collected via participant reports and the completed mood-activity log and
Mood

See the previously published case report [15] for information about the PHQ-9. An exploratory measure of mood was obtained before and after participating in the BA activity of choice by answering the following question—How would you rate your current mood—ranging from 1 (worst ever felt) to 10 (best ever felt). This was adapted from the single-item self-rating scale of happiness, which has good reliability (0.86) and construct validity (Cronbach α=0.55–0.94) [20].

See the previously published case report [15] for information on the following outcome measures: demographics, the Mini-International Neuropsychiatric Interview, acceptability, and tolerability.

Statistical Methods

A power analysis was deemed unnecessary given that the primary purpose of the study was to assess the feasibility of using VR to engage in pleasurable or mastery activities as an adjunct to a brief BA protocol. The feasibility, or the degree to which VR could successfully be integrated into the brief BA protocol, was measured by commenting on qualitative barriers to use observed. Barriers were assessed by rates of dropout, adverse events, and the number of times the headset was used.

The level of presence was obtained via participant reports from a Likert scale of 0 (not at all) to 4 (very strongly) for each question; with 3 questions, there was a possibility of yielding a score between 0 and 12. The average total presence for each participant, intention-to-treat (ITT) participant, and protocol completer was then calculated. The average presence experienced was also calculated as a percentage by dividing the average score by 12 (the maximum score).

Acceptability of the VR BA treatment was measured via participant reports using the Technology Acceptance Model, with the agreement on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree). The number of questions in each category determined the outcome range (either 0-12 for 3 questions or 0-16 for 4 questions). Each participant’s scores were then averaged along with the average ITT participants’ and protocol completers’ scores. The average percentage of acceptance was also calculated by dividing the average score by the maximum score in the outcome range. To determine the degree of acceptance as labeled on the scale, the average score was then scaled back depending on the number of questions. For example, the Perceived Usefulness category included 3 questions, yielding a potential range of 0 to 12, so an average score of 10 would be divided by 3 to assess the degree of acceptance (in this case, it would yield a score of 3.33, which would correlate to agree on the Likert scale).

Physical tolerability of the VR headset was assessed via participant reports using the Simulator Sickness Questionnaire, and the emotional tolerability of the VR headset was assessed via participant reports using the Brief Agitation Measure. Physical tolerability was broken into each item and ranged from 0 (no more than usual) to 3 (severely more than usual) for each item. Each participant’s scores were averaged along with the average ITT participants’ and protocol completers’ scores. The total percentage tolerability rating for a given activity was calculated by dividing a participant’s score by 48, as there were 16 items, yielding a potential range of 0 to 48. The percentage of intolerability for each symptom category was similarly calculated by dividing the average score by the maximum potential score of 3. The average scores for physical tolerability were summed for each participant along with the average emotional tolerability scores of each participant. Emotional tolerability was scored from 1 (strongly disagree) to 7 (strongly agree) per question; with 3 questions, there was a possibility of yielding a score between 3 and 21. These scores were rescaled to a range of 0 to 18 by subtracting 3 from all scores. The percentage of physical and emotional intolerability was calculated by dividing the average scores by the highest potential score (48 for physical tolerability and 18 for emotional tolerability).

To assess the clinical efficacy of the VR BA treatment compared with the traditional BA and TAU control groups, the participants’ depression scores were measured using the PHQ-9 at 4 time points. Owing to the small sample size, statistical analyses were not used; rather, each group’s mean score was graphically represented across time.

To explore whether engaging in an activity in VR increased mood more than engaging in an activity in real life, the participants were asked to rate their mood on a scale of 1 to 10 (1=worst they ever felt and 10=best they ever felt) before and after their chosen activity. The differences in mood before and after each VR activity were cumulatively added across each participant and then divided by the number of activities completed to find the mean. The same was done for the activities completed after sessions 2 and 3 (when the participants were asked to track their pre- and postactivity moods) for the traditional BA group. In addition, the reported pre- to postactivity mood changes of the participants in the VR BA and traditional BA groups were tallied and graphically represented.

Ethics Approval

This study was approved by Stanford University’s IRB (protocol #53483) and registered on ClinicalTrials.gov (ID #NCT04268316). A CONSORT (Consolidated Standards of Reporting Trials) checklist is also included in Multimedia Appendix 1.

Results

Participant Demographics

The sample consisted of 13 adults (mean age 35.4, SD 12.3 years; 5/13, 38% male; 7/13, 54% female; and 1/13, 8% nonbinary or third gender), with 10 (77%) adults (mean age 34.6, SD 11.50 years; 5/10, 50% male; 4/10, 40% female; and 1/10, 10% nonbinary or third gender) completing the full protocol. See Figure 1 for the CONSORT diagram and Table 1 for more participant demographic information.
Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram. BA: behavioral activation; MDD: major depressive disorder; MINI: Mini-International Neuropsychiatric Interview; OCD: obsessive-compulsive disorder; PHQ-8: Patient Health Questionnaire–8; SUD: substance use disorder; VR: virtual reality.
Table 1. Participant demographics (N=13).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>VR(^a) BA(^b) (n=5), n (%)</th>
<th>Traditional BA (n=4), n (%)</th>
<th>TAU(^c) control (n=4), n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1 (20)</td>
<td>3 (75)</td>
<td>1 (25)</td>
<td>5 (38)</td>
</tr>
<tr>
<td>Female</td>
<td>4 (80)</td>
<td>1 (25)</td>
<td>2 (50)</td>
<td>7 (54)</td>
</tr>
<tr>
<td>Nonbinary or third gender</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 to 25</td>
<td>1 (20)</td>
<td>1 (25)</td>
<td>2 (50)</td>
<td>4 (31)</td>
</tr>
<tr>
<td>26 to 30</td>
<td>2 (40)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>31 to 40</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>41 to 45</td>
<td>2 (40)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>3 (23)</td>
</tr>
<tr>
<td>51 to 55</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>56 to 60</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Race or ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>3 (60)</td>
<td>3 (75)</td>
<td>1 (25)</td>
<td>7 (54)</td>
</tr>
<tr>
<td>Chinese</td>
<td>1 (20)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Indian</td>
<td>1 (20)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>African American</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Mexican</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Past mental health treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5 (100)</td>
<td>4 (100)</td>
<td>4 (100)</td>
<td>13 (100)</td>
</tr>
<tr>
<td>Psychotherapy only</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Psychotropic medications only</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Psychotherapy and medications</td>
<td>5 (100)</td>
<td>4 (100)</td>
<td>3 (75)</td>
<td>12 (92)</td>
</tr>
<tr>
<td>No</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Current mental health treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4 (80)</td>
<td>4 (100)</td>
<td>3 (75)</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Psychotherapy only</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>1 (25)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Psychotropic medications only</td>
<td>2 (40)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Psychotherapy and medications</td>
<td>2 (40)</td>
<td>3 (75)</td>
<td>2 (50)</td>
<td>7 (54)</td>
</tr>
<tr>
<td>No</td>
<td>1 (20)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Previous experience using VR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 times</td>
<td>2 (40)</td>
<td>1 (25)</td>
<td>2 (50)</td>
<td>5 (38)</td>
</tr>
<tr>
<td>1 to 4 times</td>
<td>3 (60)</td>
<td>2 (50)</td>
<td>2 (50)</td>
<td>7 (54)</td>
</tr>
<tr>
<td>5 to 9 times</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>≥10 times</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Purpose of past VR use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaming</td>
<td>2 (40)</td>
<td>2 (50)</td>
<td>2 (50)</td>
<td>6 (46)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Research</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Other (conferences)</td>
<td>1 (20)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
</tbody>
</table>

\(^a\)VR: virtual reality.

\(^b\)BA: behavioral activation.
TAU: treatment as usual.

**VR BA Feasibility**

The completion rate was 60% (3/5) in the VR BA arm, 75% (3/4) in the traditional BA arm, and 100% (4/4) in the TAU control arm. No participants reported any serious adverse events. The participants in the VR BA arm used the headset, on average, more than required (Table 2). Of the 5 participants, 2 (40%)—participant 4 and participant 28—-reported that they kept the VR headset nearby so that they could more readily access it and remember to use it. However, only 20% (1/5) of the participants completed a post-VR questionnaire after each VR activity, with the other participants completing less than required. Participant 24 specifically expressed difficulty disentangling headset use with completing the post-VR questionnaires, which she found stressful and tedious to complete.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Dropout (yes or no)</th>
<th>Adverse events, N</th>
<th>Completed mood activity log (yes or no)</th>
<th>Times headset was used between session 1 and session 4, N</th>
<th>Completed homework worksheets, N</th>
<th>Level of presence experienced in headsetb (0-12; 3 items), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 4</td>
<td>No</td>
<td>0</td>
<td>Yes</td>
<td>21</td>
<td>15</td>
<td>9.53 (1.96)</td>
</tr>
<tr>
<td>Participant 12</td>
<td>No</td>
<td>0</td>
<td>Yes</td>
<td>11</td>
<td>11</td>
<td>2.82 (2.99)</td>
</tr>
<tr>
<td>Participant 24</td>
<td>Yes</td>
<td>0</td>
<td>Yes</td>
<td>11</td>
<td>5</td>
<td>6.40 (1.82)</td>
</tr>
<tr>
<td>Participant 28</td>
<td>No</td>
<td>0</td>
<td>Yes</td>
<td>33</td>
<td>9</td>
<td>9.56 (3.88)</td>
</tr>
<tr>
<td>Participant 30</td>
<td>Yes</td>
<td>0</td>
<td>No</td>
<td>5</td>
<td>1</td>
<td>7.00 (N/A)</td>
</tr>
<tr>
<td>Completer average</td>
<td>N/A</td>
<td>0</td>
<td>N/A</td>
<td>21.67</td>
<td>11.67</td>
<td>7.30 (3.88)</td>
</tr>
<tr>
<td>ITTd average</td>
<td>N/A</td>
<td>0</td>
<td>N/A</td>
<td>16.20</td>
<td>8.20</td>
<td>7.06 (2.77)</td>
</tr>
</tbody>
</table>

*aMinimum required headset use and completed homework worksheets was 12 each.

*bLevel of presence contained 3 items with a range of 0 (not at all) to 4 (very strongly) for each item. Higher numbers indicate greater presence.

*cN/A: not applicable.

*dITT: intention-to-treat.

The average total presence rating of the ITT VR BA participants was 59% (7.06/12), whereas the average rating of all the VR BA protocol completers was 61% (7.30/12; Table 2). Participant 24, who reported an average presence rating of 53% (6.40/12), noted that she had difficulty using the head-mounted display (HMD) with her glasses as it led to smudging. Participant 12, who reported a comparatively lower average presence rating of 24% (2.82/12), stated that she “wanted more control of when to stop in the video and look around” and wanted the ability to interact in the virtual environment. She also remarked that the image quality of the videos was not as good as that of real-life imagery. Participant 12 further noted that there was a problem in the lower left visual field of her VR headset, greatly impairing her sense of presence.

**VR BA Acceptability**

Overall, the participants who completed the protocol “agreed” that the VR treatment was acceptable, with an average rating of 87% (45.32/52) acceptability, and all VR BA participants (5/5, 100%) verbally provided positive endorsements for using the headset (Table 3).
Table 3. Virtual reality behavioral activation acceptability.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Perceived usefulness(^a) (0-12; 3 items), mean (SD)</th>
<th>Perceived ease of use(^a) (0-12; 3 items), mean (SD)</th>
<th>Attitudes toward use(^b) (0-16; 4 items), mean (SD)</th>
<th>Intention to use technology(^a) (0-12; 3 items), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>11.00 (0)</td>
<td>12.00 (0)</td>
<td>16.00 (0)</td>
<td>12.00 (0)</td>
</tr>
<tr>
<td>12</td>
<td>7.00 (1.41)</td>
<td>8.90 (0.32)</td>
<td>8.10 (3.63)</td>
<td>9.30 (0.95)</td>
</tr>
<tr>
<td>24</td>
<td>8.80 (1.48)</td>
<td>10.60 (0.89)</td>
<td>11.20 (2.95)</td>
<td>6.40 (1.52)</td>
</tr>
<tr>
<td>28</td>
<td>11.67 (1.00)</td>
<td>12.00 (0)</td>
<td>16.00 (0)</td>
<td>12.00 (0)</td>
</tr>
<tr>
<td>30</td>
<td>10.00 (N/A(^c))</td>
<td>10.00 (N/A)</td>
<td>11.00 (N/A)</td>
<td>8.00 (N/A)</td>
</tr>
<tr>
<td>Completer</td>
<td>9.89 (2.52)</td>
<td>10.97 (1.79)</td>
<td>13.37 (4.56)</td>
<td>11.10 (1.56)</td>
</tr>
<tr>
<td>ITT(^d) average</td>
<td>9.69 (1.85)</td>
<td>10.70 (1.33)</td>
<td>12.46 (3.46)</td>
<td>9.54 (2.47)</td>
</tr>
</tbody>
</table>

\(^a\)Domains comprising the technology acceptance model (higher numbers indicate greater acceptability). Perceived usefulness, perceived ease of use, and intention to use technology contained 3 items with a range of 0 (strongly disagree) to 4 (strongly agree) for each item.
\(^b\)Attitudes toward use contained 4 items with a range of 0 (strongly disagree) to 4 (strongly agree) for each item.
\(^c\)N/A: not applicable
\(^d\)ITT: intention-to-treat.

**VR BA Tolerability**

The average overall physical tolerability of those who completed the full protocol and the ITT participants was 92% (44.23/48) and 94% (45.06/48), respectively (Table 4). *Nausea* was the most endorsed symptom of physical intolerability (Table 5). *Burping* was the least endorsed symptom of physical intolerability, with no participants endorsing it after any activity. Participant 30 stated that she becomes seasick/carsick easily and found some of the VR activities nauseating. Participant 12 informed that she also becomes carsick easily and not being in control of the image’s movement made her feel sick until the headset was removed, with the longest lingering symptom dissipating 30 minutes after headset removal. The average overall emotional tolerability of those who completed the full protocol and the ITT participants was 90% (16.21/18) and 94% (16.93/18), respectively (Table 4).

Table 4. Overall tolerability.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Physical tolerability(^a) (0-48; 16 items), total mean(^b) (SD)</th>
<th>Emotional tolerability(^c) (0-18; 3 items), total mean(^d) (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.73 (0.14)</td>
<td>0.00 (0)</td>
</tr>
<tr>
<td>12</td>
<td>8.73 (0.23)</td>
<td>5.36 (0.14)</td>
</tr>
<tr>
<td>24</td>
<td>0.40 (0.07)</td>
<td>0.00 (0)</td>
</tr>
<tr>
<td>28</td>
<td>0.78 (0.10)</td>
<td>0.00 (0)</td>
</tr>
<tr>
<td>30</td>
<td>3.00 (N/A(^e))</td>
<td>0.00 (0)</td>
</tr>
<tr>
<td>Completer</td>
<td>3.75 (4.34)</td>
<td>1.79 (3.10)</td>
</tr>
<tr>
<td>ITT(^f) average</td>
<td>2.93 (3.39)</td>
<td>1.07 (2.40)</td>
</tr>
</tbody>
</table>

\(^a\)Physical tolerability determined using the Simulator Sickness Questionnaire. Possible responses for the 16 items ranged from 0 (no more than usual) to 3 (severely more than usual). Lower numbers indicate greater tolerability.
\(^b\)The mean scores for physical tolerability were summed for each participant.
\(^c\)Emotional tolerability determined using the Brief Agitation Measure. Possible responses for the 3 items ranged from 0 (strongly disagree) to 6 (strongly agree). Lower numbers indicate greater tolerability.
\(^d\)The mean scores for emotional tolerability were summed for each participant.
\(^e\)N/A: not applicable.
\(^f\)ITT: intention-to-treat.
Table 5. Physical tolerability.

<table>
<thead>
<tr>
<th></th>
<th>Participant 4, mean (SD)</th>
<th>Participant 12, mean (SD)</th>
<th>Participant 24, mean (SD)</th>
<th>Participant 28, mean (SD)</th>
<th>Participant 30, mean (SD)</th>
<th>Completer average, mean (SD)</th>
<th>ITTa average, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nausea(b 0-3)</td>
<td>0.33 (0.62)</td>
<td>0.91 (1.22)</td>
<td>0.20 (0.40)</td>
<td>0.22 (0.44)</td>
<td>1 (N/A)c</td>
<td>0.49 (0.37)</td>
<td>0.53 (0.39)</td>
</tr>
<tr>
<td>General discomfort(b 0-3)</td>
<td>0.20 (0.56)</td>
<td>0.91 (1.22)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.37 (0.48)</td>
<td>0.22 (0.39)</td>
</tr>
<tr>
<td>Stomach awareness(b 0-3)</td>
<td>0.27 (0.70)</td>
<td>0.73 (1.27)</td>
<td>0 (0)</td>
<td>0.33 (0.71)</td>
<td>0 (N/A)</td>
<td>0.44 (0.25)</td>
<td>0.27 (0.30)</td>
</tr>
<tr>
<td>Sweating(b 0-3)</td>
<td>0.27 (0.70)</td>
<td>0.55 (1.21)</td>
<td>0 (0)</td>
<td>0.11 (0.33)</td>
<td>0 (N/A)</td>
<td>0.31 (0.22)</td>
<td>0.19 (0.23)</td>
</tr>
<tr>
<td>Increased salivation(b 0-3)</td>
<td>0.13 (0.35)</td>
<td>0.55 (1.21)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.23 (0.28)</td>
<td>0.14 (0.24)</td>
</tr>
<tr>
<td>Vertigo(b 0-3)</td>
<td>0.33 (0.90)</td>
<td>0.55 (1.21)</td>
<td>0 (0)</td>
<td>0.11 (0.33)</td>
<td>0 (N/A)</td>
<td>0.33 (0.22)</td>
<td>0.20 (0.24)</td>
</tr>
<tr>
<td>Burping(b 0-3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Difficulty concentraing(b 0-3)</td>
<td>0 (0)</td>
<td>0.45 (0.82)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (N/A)</td>
<td>0.15 (0.26)</td>
<td>0.29 (0.44)</td>
</tr>
<tr>
<td>Difficulty focusing(b 0-3)</td>
<td>0 (0)</td>
<td>0.45 (0.82)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.15 (0.26)</td>
<td>0.09 (0.20)</td>
</tr>
<tr>
<td>Eye strain(b 0-3)</td>
<td>0 (0)</td>
<td>0.55 (1.21)</td>
<td>0.20 (0.40)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.18 (0.31)</td>
<td>0.15 (0.24)</td>
</tr>
<tr>
<td>Fatigue(b 0-3)</td>
<td>0 (0)</td>
<td>0.18 (0.40)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.06 (0.10)</td>
<td>0.04 (0.08)</td>
</tr>
<tr>
<td>Headache(b 0-3)</td>
<td>0 (0)</td>
<td>0.64 (1.21)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (N/A)</td>
<td>0.21 (0.37)</td>
<td>0.33 (0.47)</td>
</tr>
<tr>
<td>Blurred vision(b 0-3)</td>
<td>0 (0)</td>
<td>0.36 (0.81)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.12 (0.21)</td>
<td>0.07 (0.16)</td>
</tr>
<tr>
<td>Dizziness (eyes open(b 0-3)</td>
<td>0.20 (0.56)</td>
<td>0.64 (1.21)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.28 (0.33)</td>
<td>0.17 (0.28)</td>
</tr>
<tr>
<td>Dizziness (eyes closed(b 0-3)</td>
<td>0 (0)</td>
<td>0.64 (1.21)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.21 (0.37)</td>
<td>0.13 (0.28)</td>
</tr>
<tr>
<td>Fullness of head(b 0-3)</td>
<td>0 (0)</td>
<td>0.64 (1.21)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (N/A)</td>
<td>0.21 (0.37)</td>
<td>0.13 (0.28)</td>
</tr>
</tbody>
</table>

aITT: intention-to-treat.

bSymptoms included in the Simulator Sickness Questionnaire. Each symptom had a range of 0 (no more than usual) to 3 (severely more than usual). Lower numbers indicate greater tolerability.

cN/A: not applicable.

Clinical Efficacy

Owing to a lower than anticipated sample size, there was not enough power to conduct statistical analyses, and a graphical representation was used. Figure 2 shows the PHQ-9 scores of the participants who completed the full 4-session protocol. Overall, the mean PHQ-9 scores of the VR BA group decreased by 5.67, changing the average diagnostic severity category rating from moderate depression (14.33) to mild depression (8.67; Figure 2), a clinically significant change (>5) [14]. The mean PHQ-9 scores of the traditional BA group decreased by 3, changing the average severity from moderately severe depression (15.33) to moderate depression (12.33). The mean PHQ-9 scores of the TAU control group decreased by 0.25, which did not change the average diagnosis severity level (moderate depression).
Pre- to Postactivity Mood Scores

Descriptive statistics of the pre- to postactivity mood scores between the VR BA and traditional BA groups are presented in Table 6. The mean change in mood reported by the participants who completed the VR BA protocol was 0.18, whereas the mean change in mood reported by the participants who completed the traditional BA protocol was 1.48 (Table 6). The mode-reported mood change was 1 among both the VR BA and traditional BA participants (Figure 3). The lowest reported mood change among both the VR BA and traditional BA participants was −2, whereas the highest reported mood change was 2 among the VR BA participants and 6 among the traditional BA participants.

Table 6. Average change in mood scores pre- to postactivity completion.

<table>
<thead>
<tr>
<th>VR^a BA^b participant</th>
<th>Change in mood after VR activity^c, mean (SD)</th>
<th>Traditional BA participant</th>
<th>Change in mood after real-life activity^d, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 4</td>
<td>0.71 (0.85)</td>
<td>Participant 14</td>
<td>1.58 (1.89)</td>
</tr>
<tr>
<td>Participant 12</td>
<td>−0.36 (1.21)</td>
<td>Participant 21</td>
<td>0.65 (1.46)</td>
</tr>
<tr>
<td>Participant 24</td>
<td>0.36 (1.12)</td>
<td>Participant 22</td>
<td>2.20 (1.06)</td>
</tr>
<tr>
<td>Participant 28</td>
<td>0.18 (0.86)</td>
<td>Participant 23</td>
<td>N/A^e</td>
</tr>
<tr>
<td>Participant 30</td>
<td>0.40 (1.52)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Completer average</td>
<td>0.18 (0.54)</td>
<td>Completer average</td>
<td>1.48 (0.78)</td>
</tr>
<tr>
<td>ITT^f average</td>
<td>0.26 (0.40)</td>
<td>Total average</td>
<td>N/A</td>
</tr>
</tbody>
</table>

^aVR: virtual reality.
^bBA: behavioral activation.
^cThe observed minimum change in the VR BA group was −2, and the observed maximum change was 2.
^dThe observed minimum change in the traditional BA group was −1, and the observed maximum change was 6.
^eN/A: not applicable.
^fITT: intention-to-treat.
Discussion

Principal Findings

The results of this study illustrate that VR is a feasible, acceptable, and tolerable method of engaging in pleasurable activities in conjunction with a BA intervention for MDD. The attrition rate of 23% (3/13) of the participants in this study is comparable with other VR studies [21,22], lower than that of many RCTs of internet-based interventions for depression [23], and lower than that of a small-sample pilot RCT exploring exercise as a treatment for depression [24]. None of the participants in the VR BA treatment arm dropped out of the study because of adverse events, and no adverse events were reported throughout the duration of the study.

Despite the COVID-19 pandemic, on average, the participants in the VR BA and traditional BA arms complied with the homework assignment of completing ≥4 activities each week. However, only 20% (1/5) of the participants in the VR BA arm completed a post-VR questionnaire for every VR activity completed. This lack of full questionnaire completion could be due to the repetitive nature of the questions, as several participants noted survey fatigue. This may also be the reason why the participants in the traditional BA arm reportedly engaged in more activities than the participants in the VR BA arm as those in the traditional BA arm did not need to complete a post-VR questionnaire equivalent after each activity; rather, they were simply required to document their pre- and postactivity mood.

Another potential reason for the shortage of questionnaire completion may have been that the study was conducted via telehealth. It is possible that providing hard copies of the post-VR questionnaire in person at the end of each session and collecting these copies at the beginning of the following session may have yielded an increase in post-VR questionnaire completion [18].

Although the VR presence ratings were lower than expected, they were comparable with the presence ratings in other VR studies [21]. In general, presence may not have been higher for a few reasons. First, the Limbix headset created a subtle effect that one is looking at the image through a screen owing to the simple device technology. Second, in using a 360-degree video, to give the illusion of movement, the image moves while the participant remains still rather than the participant being able to walk around the virtual environment. Participant 12 even noted that she wished she could interact more with the environment and wanted the autonomy to decide when to stop and look around. She stated that she would have preferred a digitally rendered environment that was interactive over a more realistic environment that did not have interactive capabilities, which aligns with research illustrating that interactivity is more important than realism for yielding a greater sense of presence [25]. Furthermore, some of the activities involved sounds that were not natural to or consistent with the environment, such as a voice-over description of the scene or gentle music playing in the background. Although this HMD was chosen for its simplicity of use and portability, it is possible that, with a more advanced device or one with greater interactivity, the presence ratings would be higher [25]. However, the presence ratings were not correlated with the participants’ pre- to post-VR mood ratings in that a lower presence rating could yield a greater increase in mood than a higher presence rating and vice versa, a finding consistent with the literature given the nature of the emotion [26].

Despite the presence ratings potentially being affected by the device simplicity, on average, protocol completers strongly agreed that the VR device was easy to use and agreed that the VR BA protocol was useful. These findings are consistent with the literature stating that the simpler and easier-to-use the VR device is, the more useful it will be [27].

The participants rated the protocol as largely physically tolerable, with an average tolerability rating of 92% (44.23/48) among the protocol completers and 94% (45.06/48) among the ITT participants, and no participants dropped out because of adverse effects. Consequently, using VR to decrease symptoms of depression may be more tolerable than taking antidepressant
medications, with participants in antidepressant trials dropping out because of side effects [28]. However, larger-scale VR trials must be completed to better assert this claim.

Participant 12, who endorsed the lowest physical tolerability, specifically attributed her cybersickness to her not being in control of the image’s movement. The fact that most of the cybersickness symptoms and the strongest reported intensity of symptoms occurred during the adrenaline activities may indicate that it was due to the mismatch between the participants’ vestibular and visual cues as the movement of the image during adrenaline activities happens more quickly than during the other activities, such as watching a sunset or observing nature [11,13]. Despite being the only participant to report symptoms of agitation, participant 12 did not drop out of the study, and the symptoms did not correlate with her reported mood changes pre-to post-VR activity. Participant 12’s report of this emotional intolerability while using the headset may be due to frustration around the aforementioned problem with the visual field of the headset, and her subsequent endorsement of sadness may be due to wishing she was in the physical space of the activity.

Although the sample size was not large enough to statistically comment on whether the fidelity and efficacy of BA withstands the modification of BA to a VR format, the initial signal supports the possibility that it is not inferior. In this sample, VR BA participants experienced a greater decrease in PHQ-9 scores than those who completed the traditional BA or the TAU controls. The overall clinical severity (>5) [14] decrease in depressive symptoms for those in the VR BA arm illustrated that, despite the restrictions in place because of the COVID-19 pandemic, the participants were able to meaningfully clinically improve using VR BA.

The mean scores in the traditional BA group also decreased, with the average severity changing from moderately severe (15.33) to moderate depression (12.33). This aligns with the literature illustrating that a brief BA protocol can decrease symptoms of MDD. Although the change was not considered clinically significant as it did not meet the threshold of at least a 5-point decrease [14,16,29] per PHQ-9 criteria, this decrease in symptoms, which shifted the diagnostic categories, is a good indicator of the fidelity of the traditional BA group protocol.

This discrepancy in PHQ-9 scores between traditional BA and VR BA may have occurred because of the small sample size, and thus may not be significant. These results may also be due to the fact that the VR BA participants could have been more excited than the traditional BA participants when completing their activities—the VR BA participants noted that the novelty of using the HMD was “exciting,” whereas no such equivalent was noted among the traditional BA participants. Furthermore, the BA participants did not have the opportunity to engage in VR activities, whereas the VR BA participants were not discouraged from partaking in real-life activities. Notably, 40% (2/5) of the participants informed the protocol director that they were more motivated to partake in real-life activities after using the headset. Therefore, it is possible that the VR BA participants increased their activities in real life in addition to using the VR headset. This could explain the fact that, although the VR BA group endorsed less of an average change in mood pre-to postactivity measurement compared with the traditional BA group, they still experienced a numerically greater decrease in depression symptoms.

If using VR can improve mood or at least provide enough of a boost in mood to increase one’s motivation to engage in other pleasurable or mastery activities, it could greatly decrease the burden that depression has on individuals and society. This use could also provide some symptom relief for individuals waiting to see a mental health care provider. Furthermore, once an individual is in therapy, the use of VR could provide a sense of novelty, which may encourage individuals struggling with symptoms of depression to engage in the intervention [30]. Thus, providers could consider incorporating VR as a first step in a hierarchy of activity scheduling to incrementally increase their clients’ behavioral motivation. Scheduling activities was not an easy feat during the COVID-19 global pandemic, and using VR as a means of engaging in activities that otherwise could not be explored provided excitement and “escape” for the participants and could continue to do so if preventative barriers occur in the future. Finally, although previous studies have illustrated that BA has higher rates of retention than antidepressant medications among patients who were more severely depressed, this study further illustrated that VR may be more tolerable than antidepressant medications [4,28]. This finding suggests that partaking in a VR BA protocol could be a potential treatment alternative for those who have failed psychiatric medications owing to the side effects.

Going forward, it is necessary to replicate this study with a larger sample size both to confirm the findings and to statistically assess the efficacy and effectiveness compared with traditional BA. In addition, although this study used a Limbix HMD with videos already preloaded onto the headset for ease of use and controllability, it would be interesting to conduct a similar study with some of the more easily accessible, interactive, and immersive content with less expensive headset options. Headsets such as Google Cardboard could provide greater accessibility and content variety for the general population to engage with VR and potentially experience these positive changes.

Furthermore, given the feedback from some participants that they would have preferred more interaction within the VR landscape rather than passively watching the environment around them, research comparing the use of different headsets on feasibility, acceptability, tolerability, and efficacy is needed. This would provide additional data for individuals who have the option of obtaining different headsets and allow them to choose the option to best fulfill their wants and needs. Moreover, future research could incorporate HMDs with options to interact with other users and assess whether the social engagement component is correlated with an increase in mood. This methodology would provide a more realistic opportunity for pleasurable activities for some individuals whose values align with being social, as well as greater social accessibility for people who encounter barriers to engaging in social interaction, such as pandemic restrictions. In addition, a more advanced headset could potentially provide more activity choices, enabling individuals to engage in activities that align with their values and potentially increasing their mood.
Finally, although this study was open to adults aged >18 years, the age range of the VR BA participants was 20 to 41 years. Given that older adults experience an increase in prevalence of MDD after the age of 85 years, especially when residing in a hospital or long-term care facility setting [31], and older adults in these settings often have barriers that prevent them from becoming behaviorally activated in real life, it is important to conduct a VR BA study similar to this one with older adults. If older adults were able to experience an increase in positive mood after using VR in a similar vein to the initial results of this study, then perhaps long-term care facilities could implement the use of VR for their older patrons.

Limitations
Although some of the enumerated findings are promising, this study had several limitations. First, many of the quantitative and qualitative measures were subjective and completed by the participants. Given that the participants completed fewer post-VR questionnaires than corresponding activities, the complete data set could not be analyzed after every activity. In addition, although the VR BA participants’ aforementioned feasibility data were collected from the headset, the participants in the traditional BA arm self-reported their real-life activities, which always yields a potential for inaccuracy. Similarly, although the PHQ-9 is a self-report measure, because of the remote nature of the study, the protocol director shared her screen with the VR BA and traditional BA participants and read the questions aloud to the TAU control participants over the phone while all participants verbally answered the 9 questions. This method may have resulted in less accurate reporting if the participants felt inclined to respond in a certain way. In addition, as there were no follow-ups, it is unknown whether the mood gains that the participants reported were lasting.

Second, one of the largest obstacles to the study design was recruitment. Although the goal was to randomize 30 participants into one of the 3 study arms, only 13 were randomized because other potential participants were excluded based on ineligibility, declining to participate, or being lost to follow-up. This difficulty in recruitment may be due in part to the COVID-19 pandemic and subsequent telehealth design, with people not wanting to participate in an unpaid study during this transition. It may also be due to lack of funding and an inability to broadly advertise but could be an inherent problem with depression studies where comorbidities and misdiagnoses are common and cause exclusion from controlled studies. Given the difficulties in recruiting enough participants to conduct a powered RCT and the subsequent small sample size, the results may not be generalizable and do not indicate causality. The results may also not be applicable to all populations struggling with symptoms of depression owing to the heterogeneity of the disorder.

Conclusions
This was the first study of its kind, a historical first step in applying VR to a clinical population with MDD. Although technology is becoming increasingly popular and many studies have been conducted to analyze the feasibility and efficacy of using VR as an adjunct to or method of delivering mental health interventions for a variety of mental health disorders, this is the first study to analyze the feasibility and initial clinical efficacy of using VR as a method of engaging in pleasurable or mastery activities in conjunction with a brief BA protocol for individuals diagnosed with MDD.

This study illustrated that using VR as a method of administering BA in conjunction with a brief BA protocol for individuals diagnosed with MDD was feasible and that this intervention was able to integrate seamlessly into a telehealth design during a global pandemic. This study also illustrated that using VR as a method of administering BA in conjunction with a brief BA protocol was acceptable and tolerable for participants diagnosed with MDD.

The findings of this study demonstrate that clinicians can offer VR BA as a way for patients to experience pleasurable activities in conjunction with BA treatment to eliminate barriers that some patients may face when attempting traditional BA. VR may also be a viable alternative to psychiatric medications for some individuals given its high tolerability. In addition, given that many people do not receive adequate mental health care, VR could be a first step to help people improve their mood and increase activation while waiting to connect with a health care professional. VR BA may also be a way to operationalize and standardize BA and make it more acceptable for providers to deliver and improve the efficiency of practice. Implementation science examining VR BA is recommended.

Acknowledgments
The authors wish to acknowledge Dr Christine Blasey for her statistical consultation, Talia Lyric Weiss for her early consultation in developing the virtual reality protocol design, and Dr Nancy Haug for providing content feedback.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1242 KB - mental_v9i5e35526_app1.pdf ]

References

https://mental.jmir.org/2022/5/e35526


Abbreviations

BA: behavioral activation
CONSORT: Consolidated Standards of Reporting Trials
EBP: evidence-based psychotherapy
HMD: head-mounted display
ITT: intention-to-treat
MDD: major depressive disorder
PHQ-9: Patient Health Questionnaire–9
RCT: randomized controlled trial
TAU: treatment as usual
VR: virtual reality
Original Paper

A Group-Facilitated, Internet-Based Intervention to Promote Mental Health and Well-Being in a Vulnerable Population of University Students: Randomized Controlled Trial of the Be Well Plan Program

Daniel B Fassnacht1,2,3, PhD; Kathina Ali1,2,3, PhD; Joep van Agteren1,2,3, MSc; Matthew Iasiello3,4, MA; Teri Mavrangelos1, BSc; Gareth Furber5, PhD; Michael Kyrios1,2, PhD

1College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia
2Órama Institute for Mental Health and Wellbeing, Flinders University, Adelaide, Australia
3Wellbeing and Resilience Centre, South Australian Health and Medical Research Institute, Adelaide, Australia
4College of Nursing and Health Science, Flinders University, Adelaide, Australia
5Health, Counselling & Disability Services, Flinders University, Adelaide, Australia

Corresponding Author:
Daniel B Fassnacht, PhD
College of Education, Psychology and Social Work
Flinders University
Sturt Road
Bedford Park
Adelaide, 5042
Australia
Phone: 61 8 8201 2621
Email: dan.fassnacht@flinders.edu.au

Abstract

Background: A growing literature supports the use of internet-based interventions to improve mental health outcomes. However, most programs target specific symptoms or participant groups and are not tailored to facilitate improvements in mental health and well-being or do not allow for needs and preferences of individual participants. The Be Well Plan, a 5-week group-facilitated, internet-based mental health and well-being group intervention addresses these gaps, allowing participants to select a range of activities that they can tailor to their specific characteristics, needs, and preferences.

Objective: This study aims to test whether the Be Well Plan program was effective in improving primary outcomes of mental well-being, resilience, anxiety, and depression compared to a waitlist control group during the COVID-19 pandemic; secondary outcomes included self-efficacy, a sense of control, and cognitive flexibility. The study further seeks to examine participants’ engagement and satisfaction with the program.

Methods: A randomized controlled trial (RCT) was conducted with 2 parallel arms, an intervention and a waitlist control group. The intervention involved 5 weekly 2-hour sessions, which were facilitated in group format using Zoom videoconferencing software. University students were recruited via social media posts, lectures, emails, flyers, and posters.

Results: Using an intentional randomization 2:1 allocation strategy, we recruited 215 participants to the trial (n=126, 58.6%, intervention group; n=89, 41.4%, waitlist control group). Of the 126 participants assigned to the intervention group, 75 (59.5%) commenced the program and were included in modified intention-to-treat (mITT) analyses. mITT intervention participants attended, on average, 3.41 sessions (SD 1.56, median 4); 55 (73.3%) attended at least 4 sessions, and 25 (33.3%) attended all 5 sessions. Of the 49 intervention group participants who completed the postintervention assessment, 47 (95.9%) were either very satisfied (n=31, 66%) or satisfied (n=16, 34%). The mITT analysis for well-being ($F_{1,162}=9.65$, $P=0.002$, Cohen $d=0.48$) and resilience ($F_{1,162}=7.85$, $P=0.006$, Cohen $d=0.44$) showed significant time x group interaction effects, suggesting that both groups improved over time, but the Be Well Plan (intervention) group showed significantly greater improvement compared to the waitlist control group. A similar pattern of results was observed for depression and anxiety (Cohen $d=0.32$ and 0.37, respectively), as well as the secondary outcomes (self-efficacy, Cohen $d=0.50$; sense of control, Cohen $d=0.42$; cognitive flexibility, Cohen $d=0.65$).
Larger effect sizes were observed in the completer analyses. Reliable change analysis showed that the majority of mITT participants (58/75, 77.3%) demonstrated a significant reliable improvement in at least 1 of the primary outcomes.

**Conclusions:** The Be Well Plan program was effective in improving mental health and well-being, including mental well-being, resilience, depression, and anxiety. Participant satisfaction scores and attendance indicated a high degree of engagement and satisfaction with the program.

**Trial Registration:** Australian New Zealand Clinical Trial Registry ACTRN12621000180819; https://tinyurl.com/2p8da5sk

**(JMIR Ment Health 2022;9(5):e37292)** doi:10.2196/37292

**KEYWORDS**
COVID-19; mental health; well-being; depression; anxiety; online; digital; intervention; Be Well Plan; health outcome; online health; digital health; health intervention; primary outcome; cognition; randomized control trial; resilience; participant satisfaction; student

**Introduction**

**Background**

Despite increased investment and a growing awareness and acceptance of the need to address mental illness in an evidence-based way, to date, the prevalence of mental illness worldwide has not reduced [1,2]. On the contrary, the burden of mental illness on society is expected to grow in the next decade, both in economic and in health terms [3,4]. In addition to the significant proportion of individuals experiencing a diagnosable mental illness [5], many individuals experience poor mental health without a diagnosis (ie, psychological distress or low mental well-being), often referred to as the languishing group [6]. Despite suboptimal mental health (ie, the total number of individuals experiencing a mental illness, psychological distress, or low mental well-being) affecting a large proportion of the population, our health care systems and associated expenditures are devoted to servicing a small proportion of individuals with the most severe mental illnesses in tertiary or community care [7]. With respect to the use of evidence-based psychological therapies, as is seen in many mental health systems globally (eg, the National Health System in the United Kingdom and the Better Access initiative through the Medical Benefits Schedule in Australia), these are predominantly focused on providing care for clients with comparatively complex problems, such as moderate-to-severe common mental illnesses. Access to evidence-based, widely accessible help for less severe mental health needs is limited, ultimately leaving a large group at risk to develop more serious problems, particularly during times of community stress, such as global pandemics, which in turn is likely to increase the incidence of mental disorders. Therefore, a focus on prevention and early intervention should be a priority if we realistically wish to reduce the growing burden of suboptimal mental health [8].

**Stepped Care Models in Mental Health Promotion**

Stepped care models have been proposed and implemented as a solution to resourcing challenges and subsequent access issues across the spectrum of mental health care [9,10]. These models aim to improve the match of service needs to symptom severity and complexity, while ensuring similar or improved effectiveness compared to care as usual [11,12]. The aim is to ensure that highly specialized care will mainly focus on more severe cases, while ensuring that appropriate and effective evidence-based help remains available for those with less severe needs. A well-known example of a system using stepped care principles is the United Kingdom’s Improving Access to Psychological Therapies (IAPT) system [13], where individuals have access to psychological interventions according to their needs (eg, offering low-cost and low-intensity guided self-help or group-based services as an initial step). Other examples of stepped care models can be readily found in the Western world, including in European countries and Australia [14,15].

Although stepped models of care theoretically include a focus on building or promoting good mental health, the models are conceptually designed to deal with the impact of illness. As such, the solutions implemented across the continuum focus on preventing or treating symptoms of illness, not necessarily promoting good mental health [16,17]. Although this differentiation seems semantic at first glance, when considering a parallel with physical health, clear differences can and should be noted. The treatment of symptoms of physical illness and activities to promote general physical health (or fitness) do not necessarily equate to one another, with some interventions being meritorious for both, while others only work for either domain. For example, chemotherapeutic intervention to treat cancer but generally not considered to be helpful for improving overall health status [18,19]. In contrast, good nutrition and physical activity do improve overall health and may aid in the recovery process from cancer but are generally not sufficient to stop established cancer from advancing on its own. There is a clearly overlooked opportunity for mental health care to mirror this parallel and to systematically adopt solutions with a broader focus than simply targeting or aiming to prevent and treat symptoms of mental illness. In other words, instead of an abundant reliance on reactive solutions, there may be a place for proactive solutions that promote mental health and well-being more broadly [16].

**Promotion of Mental Health vs Treatment of Mental Illness**

Within mental health intervention research, there have been different streams seeking to promote mental health and well-being (eg, focusing on positive functioning and feeling well, high life satisfaction, more positive than negative emotions, a sense of purpose, and self-acceptance) [20]. These streams include proponents of dual-factor models [21], research on personal recovery [22], well-being therapy [23], positive...
facilitators from a variety of different backgrounds, including peers of participants. This supports a high level of tailoring for different target groups as well as for the sustainable implementation and scalability of the program.

Given the tailored nature of the Be Well Plan that considers participants’ characteristics, needs, and preferences, there is the potential that the intervention also improves outcomes beyond mental health and well-being, such as self-efficacy, a sense of control, and cognitive flexibility. Although preliminary evidence for the Be Well Plan program’s impact has been established [36], there is a clear need now to establish its efficacy using a more robust methodological design.

### Scalable, Group-Facilitated, Internet-Based Mental Health Solutions

Internet-based solutions are an avenue to deliver scalable and effective mental health interventions without draining clinical resources from existing models of care [37]. Notably, the effectiveness of internet-based interventions for mental health problems has been widely established [38-40]. Although less is known about the long-term effects (eg, follow-ups of 2 years or longer) of these interventions, a recent meta-analysis summarizing studies that have examined the long-term effects of internet-supported CBT showed diminished but large effects sizes over an average follow-up period of 3 years [41]. Internet-based or web-based solutions often utilize pre-recorded content, are self-guided, or involve smartphone apps [42]. The past 2 decades have pointed to the utility of these “self-directed” interventions in mental health care at all levels, demonstrating improvements in outcomes of mental illness, as well as outcomes of mental health and mental well-being [43,44].

Although self-guided modalities can be effective, high dropout rates are commonly reported, and research has demonstrated that these interventions often require highly self-motivated participants and do not appeal to everyone [45,46]. For example, a meta-analysis based on individual patient data of 10 randomized controlled trials (RCTs) of self-guided web-based interventions for depression suggested that almost 60% of participants dropped out before completing half of the treatment modules, while less than 20% completed all treatment modules [47]. One solution to overcome these drawbacks is the utilization of a hybrid approach that uses technology to facilitate “active” in-person or group-based care. The most well known of these approaches is telehealth, where psychological therapy is delivered using teleconferencing software [48]. Although telehealth has been used for a long time [49], it has been predominantly used within rural and remote clinical populations. Similar to the delivery of clinical care, teleconferencing software can be used to deliver interventions focusing on the promotion of mental health and well-being. Videoconferencing software, such as Skype (Microsoft Corp.), Zoom (Zoom Video Communications), or Microsoft Teams, has experienced a huge uptake in recent years, proliferating during the COVID-19 pandemic [50], where the majority of the global population was forced to shift to remote working as a result of health restrictions. As such, many group-based programs were successfully delivered via the internet, whereas prior to the
pandemic, the delivery of group programs was often met with considerable skepticism.

Internet-based interventions have the advantage of being accessible independent of location, which is particularly relevant at the moment where access to face-to-face interventions is limited due to lockdowns or quarantining as a result of the current pandemic [51]. Internet-based interventions can also counter some existing system inequities, as they facilitate access beyond metropolitan areas in rural and remote areas (with good internet access). In countries such as Australia, where internet penetration is over 90% [52], using a program such as the Be Well Plan via teleconferencing software is particularly valuable, as the program can be delivered by trained facilitators and does not rely on clinicians, which are already limited, particularly outside of metropolitan areas [53,54].

In addition, the Be Well Plan can successfully reach vulnerable populations, who may be isolated or struggling but are not ready or able to access contemporary services due to various barriers, such as cost or time. One example of such a vulnerable population is university students, who have been found to experience significantly higher levels of psychological distress compared to their peers. There is a sizeable body of research [55] investigating the mental health of university students, demonstrating high rates of mental health problems [56-58]. University students are often going through a phase of transition, are financially vulnerable, or are removed from their support systems at home [59], which increases their risk of experiencing psychological distress. Among others, these factors can account for why students experience such difficulties and why they are considered a key priority group to be targeted using innovative mental health and well-being interventions [60].

Study Aims

This study aims to advance the literature in 2 ways. First, we aim to test the efficacy of the Be Well Plan with a vulnerable population (university students) in improving primary outcomes of mental well-being, resilience, anxiety, and depression and secondary outcomes of self-efficacy, a sense of control, and cognitive flexibility. Second, we aim to examine participants’ engagement and satisfaction with the Be Well Plan facilitated in group format using teleconferencing software.

Methods

Trial Design

A 2-arm RCT was conducted comparing an active intervention (Be Well Plan) with a waitlist control condition.

Ethics Approval

The trial was approved by the local Human Research Ethics Committee (#2163) and registered with the Australian New Zealand Clinical Trial Registry (ACTRN12621000180819).

Recruitment and Procedure

University students were recruited between August 2020 and April 2021 through emails, lectures, social media posts, posters, and flyers at a medium-size public university (~24,500 enrolled students in 2020) in Adelaide, Australia. Recruitment messaging focused on inviting students to participate in a new program that aimed to build their mental health and well-being. All enrolled students across the university were eligible to participate; no other eligibility criteria applied. Participants were not paid, and the study was not part of the university’s credit system to perform research. English language proficiency was assumed, as all students had passed the university language requirement for English before their university enrolment. Similarly, given the requirements of tertiary study, computer and internet literacy was assumed.

Participants registered their interest via an online survey and indicated their preferred day and time for the 5 intervention sessions. They could then attend a general information session about the content and structure of the program, hosted online via the teleconferencing software Zoom, or watch a pre-recorded version. After providing their informed consent electronically to participate in the trial, individuals completed an online baseline assessment, including general demographic questions (ie, age, gender, ethnicity, student and employment status) and their overall health status (ie, diet, activity, sleep), as well as primary (ie, well-being, resilience, depression, anxiety) and secondary outcome measures (self-efficacy, perceived sense of control, cognitive flexibility); see a detailed description for outcome measures later. The baseline survey included 170 questions, and the median completion time was 25 minutes.

After completing the baseline assessment, participants were randomized into either the intervention or the waitlist control group. As it was expected that some participants in the intervention group would not be able to commence the program due to unavailability, we chose a 2:1 allocation ratio for the intervention group. Randomization was stratified by gender, performed by a researcher who was not involved in the delivery of the intervention using a random number generator [61]. Participants in the intervention group took part in the weekly 5-session, group-based program, which was delivered online via Zoom and was accessible for students regardless of their study location (ie, students who were not physically located in Adelaide). The group sizes ranged from 18 to 26 participants, and in total, 10 individual groups were facilitated from August 2020 to June 2021. Participants from the waitlist control group gained access to the program after the intervention group; facilitators were not aware whether they were delivering the program to the intervention or the waitlist control group.

Next, 6 weeks after the baseline assessment, participants in both groups were asked to complete another online assessment including the primary and secondary outcome measures. This survey included 175 questions, and the median completion time was 23 minutes.

This meant participants from the intervention group completed the postintervention assessment 1 week after the final session of the program. Participants from the intervention group were also asked questions about their satisfaction with the program. Participants from the waitlist control group were given access to the intervention following their second assessment. Up to 3 email reminders were sent to participants to complete the assessment.
Study Conditions
The study involved 2 conditions: the intervention group, which underwent the 5-week Be Well Plan program facilitated via Zoom, and the waitlist control group, which gained access to the Be Well Plan after the 5-week intervention period.

Be Well Plan Intervention
Participants allocated to the intervention group received detailed information, including the Zoom link to the first session of the Be Well Plan, prior to the start of the program. They further received a separate email inviting them to complete a brief 10-to-15-minute survey to assess their levels of mental health and well-being via a platform called the Be Well Tracker. The Be Well Tracker uses validated mental health and well-being scales: well-being and life satisfaction were measured using the Mental Health Continuum-Short Form [62] and the Satisfaction with Life Scale [63], respectively; resilience was measured with the Brief Resilience Scale [64]; and psychological distress was assessed using the Depression Anxiety Stress Scales [65]. The Be Well Tracker includes 50 items, and the median completion time for participants was 8 minutes. After completing the Be Well Tracker, participants received a detailed report about their levels of well-being, resilience, and distress, which provided them with relevant information that would be used throughout the Be Well Plan intervention. Thus, outcomes from the Be Well Tracker were solely used within the intervention and not for any analyses examining the efficacy of the program.

The Be Well Plan intervention has been previously described in a detailed paper by van Agteren et al [35], outlining the individual components of the program and providing insights into program materials, including screenshots of the intervention. In summary, the Be Well Plan is a weekly, 5-session internet-based, group-facilitated intervention that aims to improve mental health and well-being. The program assists participants in developing their own well-being plan tailored to their individual circumstances and needs. Participants learn and experiment with a range of evidence-based activities and skills targeted at improving mental health and well-being. Each session provides evidence-based information, self-reflection activities, and sharing of experiences between participants. The Be Well Plan introduces participants to an activity bank consisting of 30 evidence-based activities, which are selected from a large meta-analysis. Participants use various decision-making tools and visual aids (eg, flowcharts to find relevant activities based on self-reflection exercises). They are further supported by technology to find activities for their own unique needs. For example, they use their own results from Be Well Tracker measurements to find activities matched to their needs. Thus, participants can tailor the program according to particular needs and circumstances [30]; for a more detailed description of the individualization of the intervention, see the paper by van Agteren et al [35].

Each session was conducted via Zoom by 2 trained facilitators to ensure that the program adhered to the intervention protocol [35] and was delivered in an engaging and safe way. In total, 5 facilitators (authors KA, JvA, MI, and TM, and GF, AH, and KS) with a variety of professional backgrounds, including well-being research, counselling, and clinical psychology, delivered the program. A detailed description of the weekly content of the 5-week program can be found in Table S1 in Multimedia Appendix 1.

Waitlist Control Group
Participants in the waitlist control group were asked to complete the pre- and postintervention assessments, after which they were provided with access to the 5-week group-facilitated Be Well Plan sessions.

Demographic Questions
At baseline, participants were asked about their age in years, gender (ie, male, female, nonbinary), ethnicity (ie, Caucasian, Asian/Indian, others/prefer not to say), student status (ie, domestic, international), and employment status (ie, part-/full-time, no employment, unemployed/lost job due to COVID-19, other). Furthermore, general health levels were assessed with the following questions: “In general, how would you say your health is?” ranked from 1 (poor) to 5 (excellent); “What best describes your activity level?” ranked from 1 (seldom active, sedentary activities) to 3 (vigorously active for at least 30 minutes, 3 times a week); and “Please report the quality of your sleep over the past 24 hours,” ranked from 1 (worst-possible sleep) to 12 (best-possible sleep).

Outcome Measures
Primary outcome measures assessed participants’ well-being and resilience as well as their levels of depressive and anxiety symptoms. Secondary outcomes included self-efficacy, a sense of control, and cognitive flexibility.

Well-Being
The Warwick Edinburgh Mental Well-Being Scale (WEMWBS) was used to assess mental well-being, including eudaimonic and hedonic aspects of well-being [66]. The 14-item scale asks participants to indicate how often, over the past 2 weeks, from 0 (none of the time) to 5 (all of the time) they have experienced different thoughts and feelings (eg, “I’ve been feeling useful.”). Total scores range from 0 to 70, with higher scores indicating greater levels of mental well-being. Tennant et al [66] found that the WEMWBS demonstrates good content and construct validity, adequate test-retest reliability, and good internal consistency (Cronbach α=.83). The internal consistency of the WEMWBS in this study was excellent (Cronbach α=.91).

Resilience
The 10-item Connor-Davidson Resilience Scale (CD-RISC-10) [67] was used to assess resilience. Participants respond on a 5-point Likert scale from 0 (not true at all) to 4 (true nearly all the time) on how well they cope with adversity (eg, “I am able to adapt when changes occur.”). Total scores range from 0 to 40, with higher scores indicating greater levels of resilience. The CD-RISC-10 has demonstrated good construct validity and internal consistency (Cronbach α=.85) [68]. The internal consistency of the CD-RISC-10 in this study was good (Cronbach α=.88).

Depression
The 9-item Patient Health Questionnaire (PHQ-9) [69] was used to assess symptoms of depression. Participants respond on a
4-point Likert scale from 0 (not at all) to 3 (nearly every day) on how often they have experienced depressive symptoms (eg, “feeling tired or having little energy” or “feeling down, depressed, or hopeless”) over the past 2 weeks. Total scores range from 0 to 27, with higher scores indicating greater levels of depressive symptoms. The PHQ-9 has demonstrated good construct validity and internal consistency (Cronbach α=.86-.89) [69]. The internal consistency of the PHQ-9 in this sample was good (Cronbach α=.85). The following cut-offs were used in this study: 0-9=no-to-mild depressive symptoms and ≥10=moderate-to-severe depressive symptoms [69].

Anxiety
The 7-item General Anxiety Disorder (GAD-7) [70] was used to assess symptoms of anxiety. Participants respond on a 4-point Likert scale from 0 (not at all) to 3 (nearly every day) on how often they have experienced symptoms of anxiety (eg, “feeling nervous, anxious, or on edge” or “trouble relaxing”) over the past 2 weeks. Total scores range from 0 to 21, with higher scores indicating greater levels of anxiety. The GAD-7 has demonstrated good construct validity and internal consistency (Cronbach α=.91) [70]. The internal consistency of the GAD-7 in this sample was good (Cronbach α=.85). The following cut-offs were used in this study: 0-9=minimal-to-mild anxiety and ≥10=moderate-to-severe anxiety [70].

Self-efficacy
The 8-item New General Self-Efficacy Scale (NGSES) [71] was used to assess levels of general self-efficacy. Participants respond on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) on how often they believe they can achieve their goals, despite difficulties (eg, “When facing difficult tasks, I am certain that I will accomplish them.”). Total scores range from 1 to 5, with higher scores indicating greater levels of self-efficacy. The NGSES has demonstrated good predictive validity and internal consistency (Cronbach α=.85-.91) [71]. The internal consistency of the NGSES in this sample was good (Cronbach α=.87).

Sense of Control
The 12-item Sense of Control Scale (SCS) [72] was used to assess participants’ perceived sense of control over their lives. Participants respond on a 5-point Likert scale from 1 (strongly agree) to 7 (strongly disagree) on how much they feel they can control (eg, “Whether or not I am able to get what I want is in my own hands.”) or not control (eg, “Other people determine most of what I can and cannot do.”) their personal lives. Total scores range from 1 to 7, with higher scores indicating greater control. The internal consistency of the SCS in this sample was good (Cronbach α=.84).

Cognitive Flexibility
The 12-item Cognitive Flexibility Scale (CFS) [73] was used to assess participants’ mental and cognitive flexibility. Participants respond on 6-point Likert scale from 1 (strongly disagree) to 6 (strongly agree) on how much they are aware of alternatives available to them (“I have many possible ways of behaving in any given situation.”) and their willingness and ability to be flexible and adapt to situations (“I am willing to work at creative solutions to problems.”). Total scores range from 12 to 72, with higher scores indicating greater cognitive flexibility. The CFS has demonstrated good construct validity and internal consistency (Cronbach α=.76-.77). The internal consistency of the CFS in this sample was acceptable (Cronbach α=.77).

Engagement
Engagement with the program was assessed by recording whether participants attended the individual Be Well Plan sessions. Perceived session engagement was assessed after each session with a single item (“I felt engaged in the session.”). Participants responded on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Statistical Analyses
Data were collected online using Qualtrics [74]. Data analyses were performed using IBM SPSS Statistics v27 [75]. For all analyses, a significance level of Cronbach α=.05 was applied. Cohen d was calculated for both between- and within-subject effect sizes using the following formulae:

\[
d = (M_1 - M_2)/SD_{pooled} \quad \text{(between subjects)},
\]

\[
SD_{pooled} = \sqrt{(S_1^2 + S_2^2)/2},
\]

\[
d = (M_{diff})/SD_{diff} \quad \text{(within subjects)},
\]

\[
SD_{diff} = \sqrt{[(S_1^2 + S_2^2) - 2 \times r_{12} \times S_1 \times S_2]},
\]

where \(M_1\) and \(M_2\) are the means of the intervention and the waitlist control group, respectively; SD is the standard deviation; and \(r_{12}\) is the correlation between the intervention and the waitlist control group.

Pre- and postintervention differences were analyzed with mixed ANOVAs in both modified intention-to-treat (mITT; n=75, 59.5%) and completer (n=49, 38.9%) samples. All participants from the intervention group who participated in at least 1 Be Well Plan session were included in the mITT [76]. The Little missing completely at random (MCAR) test was performed to test whether data were missing completely at random ($\chi^2_{11}=14.23, P=.22$), suggesting missing data in the 7 outcomes measures were missing completely at random. Thus, for the mITT analyses, we imputed missing data on the outcome measures for participants who did not complete the postintervention assessment (n=61, 37.2% of the analyzed sample, N=164) using a Markov chain Monte Carlo method and information from the following variables: gender, age, ethnicity, working and student status, and pre- and postintervention scores in well-being, resilience, depression, anxiety, self-efficacy, a sense of control, and cognitive

https://mental.jmir.org/2022/5/e37292
flexibility. In total, we simulated 10 new data sets using a maximum of 100 iterations from which mean scores for postintervention outcomes were computed and used for the mITT analyses. As there were small but significant age differences between the intervention and waitlist control groups, we additionally performed separate mixed ANOVAs for all outcome measures while controlling for age in years. As age was a nonsignificant contributor in any of the analyses, the following results are presented without age as a covariate.

Reliable change analysis was conducted by calculating a reliable change index (RCI) using the method suggested by Jacobson and Truax [77]. Separate RCIs for the mITT and completer samples were calculated by subtracting participants’ postintervention scores from their baseline and subsequently dividing this difference score by the SE of the difference for the measurements used. The SE of the difference was estimated by

$$SE_{diff} = SD_x \times \sqrt{1 - r_{xx}},$$

where SDX refers to the SD of the difference scores and rXX refers to the internal consistency of the measure (ie, Cronbach α). Any change larger than 1.96 was considered reliable.

**Results**

**Participants**

The participant flow is shown in Figure 1. Based on an a priori power analysis [78], we estimated a sample size of 202 participants: statistical power=.80, Cronbach α=.05, Cohen $d=0.5$, and 40% attrition with a 2:1 allocation ratio for the intervention group. A total of 215 participants were randomized to the intervention (n=126, 58.6%) or the waitlist control (n=89, 41.4%) group. Of the 126 participants who were allocated to the Be Well Plan condition, 51 (40.5%) participants did not commence the program. Participants who did not commence the program reported other time commitments or unavailability for the scheduled session time (n=44, 86.3%). There were no significant differences between participants from the intervention group who attended at least 1 Be Well Plan session (n=75, 59.5%) and those who did not commence the program on any of the baseline outcome measures or demographic variables except for age; participants who did not commence with the program were, on average, 5.88 years younger ($F_{1,125}=9.81, P=.002$). For the waitlist control group, there were no significant differences between those participants who completed (n=54, 61%) versus those who did not complete the postassessment (n=35, 39%); all $P>.24$.

**Figure 1.** Consolidated Standards of Reporting Trials (CONSORT) flow diagram of study. mITT: modified intention to treat.

Next, we report demographic information about the 75 participants from the intervention group who attended at least 1 Be Well Plan session and the 89 participants who were allocated to the waitlist control group; see Tables 1 and 2. On
average, participants were 30.65 years old (SD 10.10), with the majority being female (133/164, 81.1%), Caucasian (94/164, 57.9%; 48/164, 29.3%, were Asian/Indian), and domestic (122/164, 74.4%) students. The majority of participants were employed (part- or full-time, 87/164, 53.0%) or were unemployed/had lost their job due to COVID-19 (25/164, 15.2%). On average, participants rated their overall health (mean 2.81, SD 1.03) and their diet as fair/good (mean 2.78, SD 1.06), their activity level as moderate (mean 1.76, SD 0.72), and their sleep quality as fair/good (mean 7.35, SD 2.78). Almost half of the sample (n=81, 49.4%) reported moderate-to-severe levels of depression, while over one-third (n=59, 36%) reported moderate-to-severe levels of anxiety, suggesting the vulnerability of this student cohort.

There were no significant differences between intervention and waitlist control groups on gender and ethnicity, nor were there differences on student or employment status or health status. However, there was a small but significant difference in age between the 2 groups; participants allocated to the Be Well Plan intervention were, on average, 3.11 years older compared to participants in the waitlist control group.

Table 1. Participants’ demographics and preintervention characteristics (F test).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Be Well Plan (N=75)</th>
<th>Waitlist control (N=89)</th>
<th>Significance statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%) Mean (SD)</td>
<td>n (%) Mean (SD)</td>
<td>F_df P value</td>
</tr>
<tr>
<td>Age</td>
<td>75 (100) 32.33 (10.59)</td>
<td>89 (100) 29.22 (9.50)</td>
<td>F_1,162=3.93 .05</td>
</tr>
<tr>
<td>Overall health</td>
<td>74 (99) 2.77 (1.05)</td>
<td>87 (98) 2.85 (1.01)</td>
<td>F_1,159=0.24 .62</td>
</tr>
<tr>
<td>Diet</td>
<td>74 (99) 2.70 (0.99)</td>
<td>87 (98) 2.85 (1.12)</td>
<td>F_1,159=0.78 .38</td>
</tr>
<tr>
<td>Activity level</td>
<td>74 (99) 1.78 (0.73)</td>
<td>87 (98) 1.75 (0.72)</td>
<td>F_1,159=0.10 .75</td>
</tr>
<tr>
<td>Sleep</td>
<td>74 (99) 7.19 (2.64)</td>
<td>87 (98) 7.49 (2.91)</td>
<td>F_1,159=0.48 .49</td>
</tr>
</tbody>
</table>

Table 2. Participants’ demographics and preintervention characteristics (chi-square test).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Be Well Plan (N=75)</th>
<th>Waitlist control (N=89)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Gender* (χ²=0.10, P=.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13 (17.3)</td>
<td>16 (18.0)</td>
</tr>
<tr>
<td>Female</td>
<td>61 (81.3)</td>
<td>72 (80.9)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>1 (1.3)</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td>Ethnicity (χ²=3.11, P=.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>49 (65.3)</td>
<td>46 (51.7)</td>
</tr>
<tr>
<td>Asian/Indian</td>
<td>18 (24.0)</td>
<td>30 (33.7)</td>
</tr>
<tr>
<td>Others/prefer not to say</td>
<td>8 (10.7)</td>
<td>13 (14.6)</td>
</tr>
<tr>
<td>Student status (χ²=1.72, P=.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domestic</td>
<td>59 (78.7)</td>
<td>63 (70.8)</td>
</tr>
<tr>
<td>International</td>
<td>16 (21.3)</td>
<td>26 (29.2)</td>
</tr>
<tr>
<td>Employment status (χ²=1.34, P=.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-/full-time</td>
<td>40 (53.3)</td>
<td>47 (52.8)</td>
</tr>
<tr>
<td>No</td>
<td>17 (22.7)</td>
<td>25 (28.1)</td>
</tr>
<tr>
<td>Unemployed/lost job due to COVID-19</td>
<td>12 (16.0)</td>
<td>13 (14.6)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (8.0)</td>
<td>4 (4.5)</td>
</tr>
</tbody>
</table>

*Comparison conducted only for male vs female.

Participant Engagement and Satisfaction With the Program

Participants in the intervention group (ie, n=75, 59.5%, who attended at least 1 session) attended, on average, 3.41 sessions (SD 1.56, median 4); 55 (73.3%) attended at least 4 sessions, and 25 (33.3%) attended all 5 sessions. Those included in the completer analyses (n=49, 38.9%, who attended at least 1 session and completed pre- and postintervention assessments) attended, on average, 4.20 sessions (SD 1.00, median 4); 41 (83.7%) of the participants attended at least 4 sessions, and 23 (46.9%) attended all 5 sessions.
Of the 49 intervention group participants who completed the postintervention assessment, 47 (95.9%) were either very satisfied (n=31, 66%) or satisfied (n=16, 34%). Session feedback was available from 28 (37.3%, session 5) to 57 (76%, session 1; overall median response rate=39) of 75 participants across the 5 Be Well Plan sessions. Overall, 68-75 (91.2%-100%) of participants felt engaged during the sessions, while 63-75 (84.2%-100%) participants were either satisfied or very satisfied with the quality of the sessions.

**Primary Outcomes**
A detailed outline of the scores for the Be Well Plan and waitlist control groups for the primary and secondary outcomes, including effect sizes, can be found in Tables 3 and 4. Reliable change analysis showed that the majority (58/75, 77.3%) of the mITT participants demonstrated a significant reliable improvement in at least 1 of the primary outcomes. When looking at data for the completer sample, we found that a vast majority (40/49, 81.6%) of participants showed a reliable change in at least 1 outcome.

| Table 3. Primary outcomes (well-being, resilience, depression, and anxiety): mITT analysis. |
|-----------------------------------------------|-----------------------------------------------|
| Time                                         | Be Well Plan (N=75)                          | Waitlist control (N=89)                        | Effects size, Cohen $d$ (95% CI) |
|                                              | Mean (SD) SE                                 | Mean (SD) SE                                   |Within groups | Between groups |
|                                              | Preintervention 40.44 (8.61) 0.99 43.43 (9.89) 1.05 | N/A b N/A                                      | 0.65 (0.40-0.90) 0.49 (0.17-0.80) |
|                                              | Postintervention 46.12 (8.41) 0.97 44.74 (6.99) 0.74 | N/A N/A                                       | N/A           |
| Resilience                                   | Preintervention 22.93 (5.72) 0.66 24.27 (6.30) 0.67 | N/A N/A                                       | 0.46 (0.22-0.69) 0.44 (0.13-0.75) |
|                                              | Postintervention 25.68 (5.52) 0.64 24.52 (5.54) 0.59 | N/A N/A                                       | N/A           |
| Depression                                   | Preintervention 10.95 (5.81) 0.67 9.67 (5.30) 0.56 | N/A N/A                                       | 0.66 (0.41-0.90) 0.32 (0.01-0.63) |
|                                              | Postintervention 7.70 (4.36) 0.50 8.02 (3.60) 0.38 | N/A N/A                                       | N/A           |
| Anxiety                                      | Preintervention 9.27 (5.06) 0.58 8.15 (4.57) 0.48 | N/A N/A                                       | 0.58 (0.33-0.82) 0.37 (0.06-0.68) |
|                                              | Postintervention 6.46 (4.08) 0.47 7.09 (4.11) 0.44 | N/A N/A                                       | N/A           |

a mITT: modified intention to treat.
b N/A: not applicable.

| Table 4. Primary outcomes (well-being, resilience, depression, and anxiety): completer analysis. |
|-----------------------------------------------|-----------------------------------------------|
| Time                                         | Be Well Plan (N=49)                          | Waitlist control (N=54)                        | Effects size, Cohen $d$ (95% CI) |
|                                              | Mean (SD) SE                                 | Mean (SD) SE                                   |Within groups | Between groups |
|                                              | Preintervention 40.35 (8.55) 1.22 43.74 (10.24) 1.39 | N/A a N/A                                     | 0.77 (0.44-1.08) 0.66 (0.26-1.06) |
|                                              | Postintervention 46.73 (10.24) 1.46 44.37 (8.94) 1.22 | N/A N/A                                       | N/A           |
| Resilience                                   | Preintervention 22.63 (5.20) 0.74 24.83 (6.37) 0.87 | N/A N/A                                       | 0.58 (0.27-0.88) 0.76 (0.36-1.16) |
|                                              | Postintervention 25.92 (6.81) 0.97 24.20 (7.08) 0.96 | N/A N/A                                       | N/A           |
| Depression                                   | Preintervention 11.08 (5.86) 0.84 9.74 (5.31) 0.72 | N/A N/A                                       | 0.79 (0.46-1.11) 0.39 (–0.01 to 0.78) |
|                                              | Postintervention 7.61 (5.37) 0.77 8.09 (4.61) 0.63 | N/A N/A                                       | N/A           |
| Anxiety                                      | Preintervention 8.69 (4.83) 0.69 8.33 (4.71) 0.64 | N/A N/A                                       | 0.56 (0.25-0.86) 0.28 (–0.11 to 0.67) |
|                                              | Postintervention 6.20 (5.02) 0.72 7.15 (5.26) 0.72 | N/A N/A                                       | N/A           |

aN/A: not applicable.
**Well-Being**

The mITT analysis for well-being showed a significant time × group interaction effect ($F_{1,162}=9.65, P=.002$) and a significant main effect of time ($F_{1,162}=24.77, P<.001$); however, there was no significant main effect of group ($F_{1,162}=0.50, P=.48$). The results suggest that both groups improved over time, but the Be Well Plan group showed significantly greater improvement compared to the waitlist control group. These results were replicated with the completer analysis: time × group interaction effect ($F_{1,101}=11.19, P=.001$); main effect of time ($F_{1,101}=16.62, P<.001$); and main effect of group ($F_{1,101}=0.10, P=.76$).

**Resilience**

The mITT analysis for resilience showed a significant time × group interaction effect ($F_{1,162}=7.85, P=.01$) and a significant main effect of time ($F_{1,162}=11.35, P<.001$); however, there was no significant main effect of group ($F_{1,162}=0.01, P=.91$). The results, similar to the pattern found for well-being, suggest that the Be Well Plan group showed significantly greater improvement in resilience compared to the waitlist control group. Results were again replicated in the completer sample: time × group interaction effect ($F_{1,101}=14.91, P<.001$); main effect of time ($F_{1,101}=6.86, P=.01$); and main effect of group ($F_{1,101}=0.04, P=.84$).

**Depression**

The mITT analysis for depression showed again a significant time × group interaction effect ($F_{1,162}=4.14, P=.04$) and a significant main effect of time ($F_{1,162}=39.64, P<.001$); however, there was no significant main effect of group ($F_{1,162}=0.55, P=.46$). Therefore, significantly greater improvements in depression were noted for the Be Well Plan group compared to the waitlist control group. Results were not replicated in the completer sample, as the time × group interaction effect ($F_{1,101}=3.88, P=.05$) did not meet our significance threshold; similar to the mITT analysis, there was a significant main effect of time ($F_{1,101}=30.60, P<.001$) and no significant main effect of group ($F_{1,101}=0.21, P=.65$).

**Anxiety**

Similarly to depression, the mITT analysis for anxiety showed a significant time × group interaction effect ($F_{1,162}=5.64, P=.02$) and a significant main effect of time ($F_{1,162}=27.41, P<.001$); however, there was no main effect of group ($F_{1,162}=0.17, P=.68$). Results thus indicate that the Be Well Plan group improved more in anxiety symptoms compared to the waitlist control group. Results from the completer analysis differed as the time × group interaction effect ($F_{1,101}=2.01, P=.16$) was not statistically significant; similar to the mITT analysis, there was a main effect of time ($F_{1,101}=15.97, P<.001$) and no significant main effect of group ($F_{1,101}=0.11, P=.74$).

**Differential Change in Primary Outcomes**

Of the total mITT participants who demonstrated a reliable change in depression or mental well-being, most (27/48, 56.3%) only showed a change in well-being, with 8 (16.7%) only demonstrating a change in depression and 13 (27.1%) demonstrating a change in both outcomes. Of the 48 participants who demonstrated a change in well-being and anxiety, most (25/48, 52.1%) showed a change in both outcomes, with 8 (16.7%) only demonstrating a change in anxiety and 15 (31.3%) only showing a change in mental well-being.

Results were similar for the completer analysis. For participants who had a change in mental well-being and depression, the majority (17/33, 51.5%) improved in both outcomes, with 7 (21.2%) only improving in depression and 9 (27.3%) only improving in mental well-being. Of the 31 participants who demonstrated a reliable change in mental well-being and anxiety, the majority (16/31, 51.6%) showed a reliable change in both outcomes, with 5 (16.1%) only showing a change in anxiety and 10 (32.3%) only showing a change in mental well-being.

Only 5 participants (mITT: 5/75, 6.7%; completer: 5/49, 10.2%) reported a reliable deterioration in well-being, depression, or anxiety: 1 participant showed a reliable decrease in well-being (mITT: 1/75, 1.3%), whereas 2 participants showed a reliable increase in depression or anxiety symptoms (mITT: 4/75, 5.3%). Importantly, no participant showed reliable deterioration in more than 1 of the mentioned outcomes. Furthermore, no participant reported any harmful effects from participation in the program.

**Secondary Outcomes**

**Self-efficacy**

The mITT analysis for self-efficacy showed a significant time × group interaction effect ($F_{1,162}=4.04, P=.047$) and a significant main effect of time ($F_{1,162}=10.75, P=.001$); however, there was no main effect of group ($F_{1,162}=0.42, P=.52$); see Tables 5 and 6. Results therefore suggest that participants in the Be Well Plan group increased more in self-efficacy compared to the waitlist control group. Results differed in the completer analysis: although the time × group interaction effect was still significant ($F_{1,95}=6.75, P=.01$), the main effect of time ($F_{1,95}=3.83, P=.05$) and group ($F_{1,95}=0.72, P=.40$) was not.
### Table 5. Secondary outcomes (self-efficacy, sense of control, and cognitive flexibility): mITT analysis.

<table>
<thead>
<tr>
<th>Time</th>
<th>Be Well Plan (N=75)</th>
<th>Waitlist control (N=89)</th>
<th>Effects size, Cohen $d$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>SE</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td><strong>Self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>3.55 (0.60)</td>
<td>0.07</td>
<td>3.69 (0.62)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>3.80 (0.51)</td>
<td>0.07</td>
<td>3.75 (0.55)</td>
</tr>
<tr>
<td><strong>Sense of control</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>4.65 (0.90)</td>
<td>0.10</td>
<td>4.89 (0.94)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>5.03 (0.88)</td>
<td>0.12</td>
<td>4.96 (0.81)</td>
</tr>
<tr>
<td><strong>Cognitive flexibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>51.54 (6.16)</td>
<td>0.71</td>
<td>53.51 (7.20)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>54.69 (5.42)</td>
<td>0.65</td>
<td>54.28 (5.94)</td>
</tr>
</tbody>
</table>

*a*mITT: modified intention to treat.

*b*N/A: not applicable.

### Table 6. Secondary outcomes (self-efficacy, sense of control, and cognitive flexibility): completer analysis.

<table>
<thead>
<tr>
<th>Time</th>
<th>Be Well Plan (N=49)</th>
<th>Waitlist control (N=54)</th>
<th>Effects size, Cohen $d$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>SE</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td><strong>Self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>3.54 (0.55)</td>
<td>0.08</td>
<td>3.79 (0.63)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>3.81 (0.61)</td>
<td>0.09</td>
<td>3.75 (0.65)</td>
</tr>
<tr>
<td><strong>Sense of control</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>4.66 (0.81)</td>
<td>0.12</td>
<td>4.98 (0.98)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>5.02 (1.07)</td>
<td>0.15</td>
<td>4.98 (1.03)</td>
</tr>
<tr>
<td><strong>Cognitive flexibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>51.69 (5.91)</td>
<td>0.84</td>
<td>54.89 (7.22)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>54.79 (6.78)</td>
<td>0.98</td>
<td>53.98 (7.69)</td>
</tr>
</tbody>
</table>

*a*N/A: not applicable.

### Sense of Control

The mITT analysis for sense of control showed a significant time × group interaction effect ($F_{1,162}=5.15$, $P=.03$) and a significant main effect of time ($F_{1,162}=10.76$, $P=.001$); however, there was no main effect of group ($F_{1,162}=0.59$, $P=.45$). These results indicate that participants in the Be Well Plan group increased more in their sense of control compared to the waitlist control group. Results were replicated in the completer analysis: although the time × group interaction effect was still significant ($F_{1,99}=10.75$, $P=.001$), the main effect of time ($F_{1,99}=3.23$, $P=.08$) and group ($F_{1,99}=0.92$, $P=.40$) was not.

### Discussion

**Principal Findings**

This study examined the efficacy of a group-facilitated, internet-based program to promote mental health and well-being in a vulnerable population of university students. Compared to waitlist controls, participants in the intervention group significantly improved in all primary outcomes, including mental well-being, resilience, depression, and anxiety, as well as secondary outcomes, including self-efficacy, a sense of control, and cognitive flexibility. Furthermore, participants’ engagement and satisfaction with the Be Well Plan were examined, showing increased more in cognitive flexibility compared to the waitlist control group. Results differed in the completer analysis: although the time × group interaction effect was still significant ($F_{1,99}=10.75$, $P=.001$), the main effect of time ($F_{1,99}=3.23$, $P=.08$) and group ($F_{1,99}=0.92$, $P=.40$) was not.
that students were highly engaged and satisfied with the program.

**Improvements in Well-Being and Resilience**

The study clearly demonstrated the anticipated significant improvements in mental well-being and resilience, confirming preliminary positive effects identified in a previous uncontrolled intervention study [36]. We found medium effect sizes for mental well-being, which is above the average typically reported in the literature [29,79]. For example, previous meta-analyses of psychological interventions in the general population reported small effect sizes for well-being interventions of similar intensity: programs longer than 4 weeks tend to produce small, positive effects according to van Ageten et al [30] (Hedges g=0.32), Sin and Lyubomirsky [80] (r=.36 for 5-7-week interventions). Importantly, effect sizes tend to be much lower for internet-based interventions (Hedges g=0.22) [eg. 30], attesting to the positive impact of the *Be Well Plan* program. For resilience, we found small-to-medium effect sizes, which is in line with what is typically reported in the research literature. For example, a meta-analysis of 11 RCTs examining resilience interventions by Joyce et al [81] reported a standardized mean difference of 0.44 between resilience interventions and waitlist control groups.

There are various potential sources for the observed positive effects. First, the program was rigorously designed based on a best-practice intervention development methodology [35]. The intervention-mapping approach [82] and comparable development methodologies, such as the behavior change wheel [83], are frequently used in health promotion research but not readily in mental health or psychology research. Using the intervention-mapping process meant that the program was (1) designed based on a comprehensive needs analysis, (2) grounded in a well-defined theory of behavior change, (3) co-designed using knowledge and experience from a range of different stakeholders (ie, psychologists, counsellors, mental health researchers, end users), and (4) composed of evidence-based behavior change techniques. The included activities were based on our team’s research into effective well-being interventions [30], resulting in activities from a wide variety of therapeutic approaches, including CBT, acceptance- and commitment-based therapy (ACT), mindfulness, and positive psychology. This “theory-agnostic” approach provides further explanation for the observed positive effects across the outcomes of well-being, resilience, depression, and anxiety, with these approaches having solid evidence for being able to change these outcomes [28,31,33,34,81].

Second, the intervention’s focus on tailoring and individualizing a well-being strategy to a participant’s unique context, needs, and preferences likely aided in achieving positive effects across all outcomes. In contrast to “generic” interventions, which are typically similar for all participants (ie, everyone receives the same content), the *Be Well Plan* was designed to allow participants to experiment with different techniques they wanted to include in their own well-being program (ie, *their Be Well Plan*) based on their perceived characteristics and needs. Previous research has argued for the importance of person-intervention fit through tailoring and individualization of programs or their components as a potential strategy to improve the efficacy of and engagement with psychological interventions, not just for clinical mental health programs, but also for well-being and mental health promotion programs [84-86]. Personalization plays a crucial role in face-to-face therapy but is similarly touted as an important advantage for internet-based interventions (eg, to increase personal relevance and engage users) [87-89]. Although the importance of tailoring interventions to individual characteristics, needs, and preferences has been highlighted in previous research [90], tailored interventions such as the *Be Well Plan*, which center around individual agency, are rare. The program allows individuals to choose their own activities and tailor these to their specific needs and preferences, fostering individual agency and autonomy, which is an important factor in improving health behaviors, including mental health and well-being, and an important component of contemporary well-being theories [91,92]. It is important to note that although this study’s purpose was not to investigate the superiority or noninferiority of our tailored approach over generic programs, research should look further into the impact of higher degrees of personalization on both efficacy and engagement in group-based programs.

Third, the facilitated group-based format of the *Be Well Plan* offered several advantages over a self-guided, individual approach, which likely improved both outcomes and engagement with the program. For example, sharing personal experiences in a safe and supportive environment may have led to vicarious learning and a feeling of being supported by others [93], while the trained facilitators guiding participants through the program and supporting them possibly increased engagement [94,95]. Importantly, the aspect of social connectedness has also been highlighted as a facilitator for user engagement in a recent systematic review [89]. Furthermore, allowing participants in internet-based group interventions to experiment with different evidence-based techniques in an effective manner has become much more within reach with the rise of technology [96,97]. For instance, technology can help guide activity recommendations based on an individual’s response to scientific questionnaires for mental health and well-being, aiding in personalization, as is the case for the *Be Well Plan*.

**Improvements in Depression and Anxiety**

The positive effects on outcomes of depression and anxiety are encouraging, particularly as they build on similar outcomes found in a previous uncontrolled study of the intervention [36]. Psychological distress is an independent outcome to clinical symptoms [98]; therefore, finding improvements in both markers across the 2 studies points to the potential utility of the program for clinical settings. Although within-subject effect sizes were medium, between-subject effects were small. However, it is important to note that the recruited cohort was not a clinical sample, implying that effects could potentially be greater in individuals with clinical depression or anxiety. Having said that, university students are known to be an at-risk population reporting poor mental health, including depression and anxiety [56-58], which was shown by the high proportion of participants reporting moderate-to-severe baseline levels of depression and anxiety.
Our findings also need to be interpreted in the context of the worldwide COVID-19 pandemic. Recruitment and program participation took place between August 2020 and June 2021. Although the effects of COVID-19 in South Australia where the sample was recruited from were modest compared to other jurisdictions in Australia or worldwide, restrictions due to COVID-19 were still in place throughout the study period. For example, in March 2020, the South Australian government declared a public health emergency, which included measures such as closures of state borders and physical distancing requirements (eg, a 3-day lockdown in November 2020). Unsurprisingly, previous studies have found detrimental effects of COVID-19 on mental health in large, representative Australian cohorts [99,100], individuals who have been impacted by the adverse border closure effects of COVID-19 [101], and university students [102,103].

Improvements in Self-efficacy, Sense of Control, and Cognitive Flexibility

After participating in the Be Well Plan, improvements in self-efficacy, sense of control, and cognitive flexibility were observed. It might be that the tailored nature of the intervention, which encourages individuals to initially understand their own mental health and well-being and subsequently identify effective strategies to improve or maintain good levels of mental health, elicited the belief in individuals that they are able to change or take control of their life and can adapt to circumstances. This is important as self-efficacy, a sense of control, and cognitive flexibility have been identified as protective factors for good mental health [104-106].

Participant Program Engagement and Satisfaction With the Program

Overall, engagement with the Be Well Plan was strong. Participants who began the program attended, on average, over two-thirds of the 5 sessions; of those who completed the postsession feedback, over 90% (n=68-75, 91.2%-100%) felt engaged during the sessions. The field of internet-based mental health interventions has been grappling with high attrition and dropout rates [107,108], particularly with fully self-guided programs and open-access trials. For this study, it is likely that the Be Well Plan did not directly target symptoms of depression or anxiety, almost half of the sample (46.7% for depression, 44.0% for anxiety) showed a reliable change in the respective outcomes.

Strengths and Limitations

A major strength of this study was the pre-registered, rigorous RCT design. The RCT was conducted in a vulnerable population with high attendance rates and whose mental health benefitted from the intervention. This is 1 of the first studies to rigorously evaluate an online group-facilitated mental health intervention via teleconferencing that aims to improve both mental health and well-being without targeting specific symptoms or a specific group. The intervention is unique as it allows individuals to experiment with a variety of activities that can be tailored to their individual needs and circumstances and encourages habit formation. A particular strength of the study was that symptoms of depression and anxiety were reduced, even though the Be Well Plan was not developed to specifically address these outcomes, nor were participants provided with any traditional psychoeducational information about these mental health problems.

Another strength was the web-based format of the intervention, which is particularly interesting in vast countries, such as Australia, where internet access is sufficient in regional and rural areas and mental health services are accepted and actively sought out [109]. Web-based interventions have previously been well accepted, allowing participants to interact while remaining in their own homes [110]. This modality has been investigated in the COVID-19 era, with interventions demonstrating efficacy, participant satisfaction, and engagement, while removing barriers and inconveniences related to attending in-person sessions [111,112].

Although there were several strengths of the study, some limitations need to be discussed. First, the study used a waitlist control group for comparison; although waitlist groups are cost-effective and ethical alternative control conditions, they might exaggerate effects sizes compared to other control conditions (eg, no intervention or active psychological placebo conditions) [113]. Future studies should test the Be Well Plan against an active, psychological placebo control group. Second, although the study population (ie, university students) was a vulnerable group, it does limit the generalizability of the findings to the general public. For instance, the general public typically reports better mental health, which in turn means that effect sizes in the general public may be considerably lower. Digital literacy may also be higher in a student population, potentially affecting accessibility and scalability for a broader range of specific or general community groups. Future work should be conducted to further test the Be Well Plan in different population sizes in the general public.
cohorts. Third, the study was not sufficiently powered to find significant effects for depression and anxiety in the completer analysis, despite medium effect sizes. Although almost half of participants showed reliable improvements in depression and anxiety outcomes, future studies should include larger samples to allow to test for small-to-medium effects sizes in psychological distress with clinical samples. Fourth, there was a high number of participants who registered their interest, completed the baseline measures, and were randomized to the intervention group but did not commence the first session. Although a limitation, this is common with internet-based programs; for example, a variety of studies in the area have noted similar engagement (uptake and adherence) challenges [114-116]. Fifth, our findings are short-term results only as our postassessment was taken at 1 week after the final session. The literature clearly indicates that the impact of mental health interventions diminishes over time, particularly in general well-being programs. Although investigating the long-term impact might look like an interesting question, the literature on diminishing returns is well established [30]. Rather, it is arguably more important to invest effort in designing and testing sustainable booster development. Current work is currently underway to develop ongoing topical booster sessions that aim to both reinforce core program learnings and introduce new content and activities over time. Another limitation of this study was that we did not collect data on which activities individuals used during their participation in the program. Future studies should examine intervention processes (eg, which activities were used and how frequently) to better understand mechanisms of change. Furthermore, most participants were female. Although males who did participate in the program benefited equally as female participants, some caution is required when generalizing results in males, due to a small sample size. Future studies need to attract and evaluate more males in programs such as the Be Well Plan.

Finally, as is the case with most psychological treatment studies, the outcome measures were all based on subjective reports. Future studies could feasibly undertake evaluations using behavioral or other objective measures. For instance, the use of technology can now facilitate evaluations using objective measures, such as activity levels, sleep patterns and other physiological outcomes, the use of health services, and prescribed psychopharmaceuticals. Knowing whether the Be Well Plan also advances improvements in such outcomes would add to its utility as a prevention and early intervention program.

**Conclusion**

This is 1 of the first studies to rigorously evaluate a live-facilitated (eg, via teleconferencing software), online intervention, the Be Well Plan, that aims to improve both mental health and well-being without targeting specific symptoms or a particular target group. The intervention is unique as it allows individuals to experiment with a variety of activities that can be tailored to their individual needs and circumstances and encourages habit formation. A particular strength of the study was that symptoms of depression and anxiety reduced (alongside improvements in well-being and resilience), even though the Be Well Plan was not developed to specifically address these outcomes directly, nor were participants provided with any traditional psychoeducational information about these mental health problems. This points to the program having value in both mental illness prevention and early intervention settings where current offerings are limited.

**Acknowledgments**

The authors would like to acknowledge the contributions of South Australian Health and Medical Research Institute staff members, including Monique Newberry, Natalie Tuckey, Kim Seow, Stuart Freebairn, Belinda Hall, Ben Ashley, Karen Reilly, and Laura Lo, as well as Flinders University PhD student Alexis Howard. The delivery of the Be Well Plan was supported with funding from the College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia, as well as funding from the James and Diana Ramsay Foundation.

**Conflicts of Interest**

The South Australian Health and Medical Research Institute, which employs JvA and MI, receives financial compensation from providing the Be Well Plan to organizations and the community.

Multimedia Appendix 1
Weekly content of the 5-week Be Well Plan program.
[DOCX File, 15 KB - mental_v9i5e37292_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 56149 KB - mental_v9i5e37292_app2.pdf]

**References**


52. Australian Communications and Media Authority. Communications and Media in Australia: How We Use the Internet. Canberra, Australia: Australian Communications and Media Authority; 2022.


61. Randomness and Integrity Services Ltd. Random Integer Generator. URL: https://www.random.org/integers/ [accessed 2022-04-26]


69. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.01600960.x] [Medline: 11556941]


96. Fassnacht et al. JMIR Ment Health 2022 | vol. 9 | iss. 5 | e37292 | p.178 https://mental.jmir.org/2022/5/e37292 (page number not for citation purposes)


Abbreviations

ACT: acceptance- and commitment-based therapy
CBT: cognitive behavior therapy
CD-RISC-10: 10-item Connor-Davidson Resilience Scale
CFS: Cognitive Flexibility Scale
GAD-7: 7-item General Anxiety Disorder
mITT: modified intention to treat
NGSES: New General Self-Efficacy Scale
PHQ-9: 9-item Patient Health Questionnaire
RCI: reliable change index
RCT: randomized controlled trial
SCS: Sense of Control Scale
WEMWBS: Warwick Edinburgh Mental Well-Being Scale

© Daniel B Fassnacht, Kathina Ali, Joep van Agteren, Matthew Iasiello, Teri Mavrangelos, Gareth Furber, Michael Kyrios. Originally published in JMIR Mental Health (https://mental.jmir.org), 05.05.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on https://mental.jmir.org/, as well as this copyright and license information must be included.