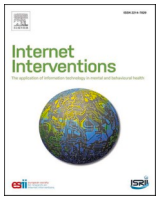
Internet Interventions 38 (2024) 100774

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent



Crowdsourcing integrated into a digital mental health platform for anxiety and depression: A pilot randomized controlled trial

Benjamin Kaveladze [a,b,c,*], Jane Shkel [c], Stacey Le [c], Veronique Marcotte [e], Kevin Rushton [f], Theresa Nguyen [f], Stephen M. Schueller [c,d]

[a] Department of Preventive Medicine, Northwestern University, Chicago, IL, United States
[b] Department of Medical Social Sciences, Northwestern University, Chicago, IL, United States
[c] Department of Psychological Science, University of California, Irvine, Irvine, CA, United States
[d] Department of Informatics, University of California, Irvine, Irvine, CA, United States
[e] School of Medicine, University of California, Irvine, Irvine, CA, United States
[f] Mental Health America, Alexandria, VA, United States

ARTICLE INFO

Keywords:
Digital mental health intervention
Depression
Anxiety
Crowdsourced support
Peer support
Public mental health

ABSTRACT

Background: Anxiety and depression are major public health concerns. Digital mental health interventions (DMHIs) are effective at reducing anxiety and depression, especially when they leverage human support. However, DMHIs that rely on human supporters tend to be less scalable. "Crowdsourced peer support," in which a "crowd" of many peers provides users support via structured and focused interactions, may enable DMHIs to provide some of human support's unique benefits at scale.

Objective: To conduct a pilot trial of two versions of a digital mental health intervention for anxiety and depression: one with crowdsourced peer support and one without.

Methods: We conducted a two-armed pilot randomized controlled trial examining two versions of the novel "Overcoming Thoughts" platform: crowdsourced (intervention) vs. non-crowdsourced (control). The crowdsourced version allowed participants to view and interact with other users' content. We randomly assigned 107 participants to use the crowdsourced ($n = 56$) or non-crowdsourced ($n = 51$) platform for 8 weeks. Participants completed assessments at baseline, 4 weeks, 8 weeks, and 16 weeks. At each time point, these assessments included measures of anxiety and depression, including the Depression, Anxiety, and Stress Scale (DASS, primary outcome), the Patient Health Questionnaire (PHQ-9, secondary outcome), and the Generalized Anxiety Disorder Questionnaire (GAD-7, secondary outcome). We also collected usage information, including the number of exercises started, and safety data.

Results: Using mixed models controlling for demographic factors, we compared the conditions' effectiveness in reducing depression and anxiety over time. Although we found significant drops over time in the DASS at both Week 8 and Week 16 ($ps < 0.01$), we did not find significant treatment x time interactions (Week 8, $p = 0.35$; Week 16, $p = 0.68$). The PHQ-9 and GAD-7 showed similar results. The median number of times participants used the platform was 3 (mean = 6.99, SD = 9.78). Greater platform use was not associated with a different change in DASS total score, PHQ-9 score, or GAD-7 score over eight weeks ($ps > 0.10$).

Conclusions: Neither version of the "Overcoming Thoughts" platform (crowdsourced or non-crowdsourced) reduced anxiety or depression significantly more than the other. Future work should investigate how digital platforms can better leverage crowdsourced support, and if crowdsourced support may be especially useful in certain kinds of systems, populations, or target areas. Optimizing intervention engagement and obtaining the large sample sizes needed for appropriate statistical power will be key challenges for similar studies.

NCT: 04226742

* Corresponding author at: Department of Preventive Medicine, Northwestern University, Chicago, IL, United States.
E-mail address: benjamin.kaveladze@northwestern.edu (B. Kaveladze).

https://doi.org/10.1016/j.invent.2024.100774
Received 16 July 2024; Received in revised form 2 September 2024; Accepted 10 September 2024
Available online 11 September 2024

1. Introduction

Anxiety and depression are common mental health problems globally, but there are not enough providers to meet the demand for high-quality care (Mongelli et al., 2020). Care-seeking has increased and evolved in recent years, with many people seeking guidance from technologies such as online screening tools (Jacobson et al., 2022; Johnson et al., 2022; Kaveladze et al., 2021; Murphy et al., 2018). Digital mental health interventions (DMHIs) are technology-based interventions used to prevent, assess, treat, and recover from mental health challenges, as well as to promote mental wellbeing. Although various terms exist, such as internet interventions and computerized therapies (see Smoktunowicz et al., 2020), we opt for "DMHI" to emphasize that the primary intervention is provided by the technology itself (with potential supplemental support coming from human involvement).

DMHIs have demonstrated an ability to disseminate effective evidence-based support for anxiety and depression in a variety of settings and contexts (Andersson et al., 2019; Karyotaki et al., 2021; Kim et al., 2023; Ramos et al., 2024). Successful implementation of DMHIs has occurred in several countries, such as Australia, Sweden, and Canada (Titov et al., 2019), as indicated by the DMHIs' integration into care settings, government support, and evidence of clinical value (i.e., Titov et al., 2020). However, such implementations have still reached only a small percentage of individuals who experience clinical levels of anxiety or depression, and they often rely on trained professionals to support care. Even successfully deployed DMHIs face numerous challenges sustaining and engaging users (Baumel et al., 2019; Park et al., 2022). Ultimately, to have the equitable global impact we seek, we need to continue to innovate on DMHIs, including identifying modes of interaction that can widely and cheaply scale. These DMHIs should meet people where they are, meaning they appeal to users with diverse backgrounds and preferences and are accessible in places people go to seek mental health support, not just traditional clinical pathways.

As mental health support-seeking evolves and expands, DMHIs can make the greatest impact by integrating into settings where people are likely to find them. Online mental health screening websites are one place to reach people who are seeking immediate support and motivated to take action. These spaces are frequently used as a first step toward improving mental health (Murphy et al., 2018); in 2022, over 6.3 million people worldwide completed a mental health screen on the website *Mental Health America* (Mental Health America, 2022). Rather than relying on typical DMHI dissemination strategies, such as standalone mobile apps that can be downloaded in a smartphone app store, integrating brief DMHIs into popular online spaces for help-seeking might better achieve the ideal of meeting people where they are.

1.1. Crowdsourced peer support in DMHIs

Making DMHIs more effective and engaging in real-world settings is a key goal for the field. One approach is to integrate human supporters (e.g., therapists, coaches, trained peer supporters, or untrained fellow users). Most research suggests that human support is useful for boosting engagement and impact (Werntz et al., 2023). However, human supporters are resource-intensive, which can decrease DMHIs' scalability. Additional work is needed to clarify which kinds of human support offer the most efficient combinations of effectiveness, appeal, and potential for scale (Schueller et al., 2017).

Human support in DMHIs is commonly provided by paid therapists or coaches with varying levels of expertise (e.g., professional education and credentialing, specific certifications like "mental health coach," and platform-specific training). Another form of support, however, is peer support, provided by people who share one's lived experience (i.e., peers). In some implementations, "peers" are trained and paid. For example, in one app, paid military veteran peer supporters helped app users (fellow veterans) via unstructured supportive check-ins (Nelson et al., 2014). Another DMHI used paid peers to help users navigate a recovery app and apply concepts to their lives (Gulliver et al., 2019). In other cases, peers are untrained fellow users (Fortuna et al., 2020), as in one app that leverages a pool of unpaid volunteer peers, who play an emotional supporter role in anonymous one-on-one chats with users (Baumel, 2015).

One form of online peer support with a particularly high potential for scale is crowdsourced peer support, in which members of a "crowd" (i.e., a large group of peers) provide support through structured and focused interactions. In crowdsourced peer support, one receives support from one or multiple crowd members. In some implementations, one instance of support provision can also reach many people, as in a system where exemplary support interactions are highlighted for future users. Thus, the dynamics of crowdsourced peer support are arguably more similar to those of anonymous online support discussion forums than traditional face-to-face therapy. DMHIs leveraging crowdsourced support have shown efficacy in multiple studies (Morris et al., 2015, 2018; Smith et al., 2021). Crowdsourced peer support is also a central component of platforms with large real-world user bases, such as Koko, Reddit, and TalkLife (De Choudhury and De, 2014; Morris, 2015; Sharma et al., 2020).

1.2. Current study

Integrating crowdsourced support into accessible online DMHIs might help to make them more engaging and effective. We conducted a pilot randomized controlled trial (RCT) examining the effects of crowdsourced support integrated into DMHIs. As a pilot (rather than confirmatory) trial, the study's primary aims were logistical: to identify problems and opportunities with the study design for subsequent large-scale, fully powered trials, and to identify opportunities to improve the intervention platform and crowdsourced feature (Conn et al., 2010). We predicted that the crowdsourced DMHI would be more effective at reducing depression and anxiety than an identical DMHI without crowdsourced support. We also explored how participants interacted with the platform over the eight-week study period.

2. Material and methods

2.1. Overview

In this pilot study, we compared the effectiveness of two versions of a platform built to reduce symptoms of depression and anxiety, with crowdsourced as the treatment condition and non-crowdsourced as the control condition. Participants experiencing clinically significant levels of depression and anxiety symptoms were recruited and randomly assigned (1:1) to a condition. Participants were instructed to use the platform for eight weeks. We collected outcome data at baseline, four, eight, and 16 weeks. The study was reviewed and approved by the Institutional Review Board at the University of California, Irvine, and was pre-registered as a clinical trial at clinicaltrials.gov (NCT: 04226742). The study was conducted remotely, and the chief investigator was in Irvine, California.

2.2. Intervention

This study used the "Overcoming Thoughts" platform, a novel platform for treating depression and anxiety symptoms. "Overcoming Thoughts" was based on principles of cognitive-behavioral therapy and heavily influenced by the unified protocol (Farchione et al., 2012). The platform focused on two specific techniques from cognitive-behavioral therapy – cognitive restructuring and behavioral experimentation. Both techniques were intended to teach users skills for coping with negative thoughts. These techniques were implemented in the platform as follows: On opening the platform, users were directed to input a thought they were struggling with. The platform then offered users a

choice between the two approaches to addressing that thought. The first was "Explore thoughts to feel better," which guided them through an activity based on cognitive restructuring. The activity used eight prompts guiding the user to identify unhelpful thoughts and core beliefs and replace them with more adaptive ones. Alternatively, users could select "Take steps to feel better," an option based on behavioral experimentation concepts. This option guided users through eight prompts focusing on identifying specific actions they could take to feel better and overcome obstacles to positive action. After users worked through all eight prompts in either activity, they were presented with a summary page showing all prompts and their responses. This platform was modeled after other cognitive-behavioral therapy-based DMHIs in that it included multiple components of CBT in a single intervention platform. Participants were encouraged to use the platform as frequently as they found valuable, with a recommendation of three times per week for the eight-week study period.

Our study focused on isolating the impact of crowdsourcing, thus our design compared two versions of the "Overcoming Thoughts" platform. The only feature that distinguished the two versions of the platform was crowdsourced support. The "crowdsourced" version of the platform was exactly as described above, except it also showed responses to the prompts generated by other platform users. The crowdsourced responses were displayed under the heading, "What others are saying." Users could look through a list of these responses and interact with a response by selecting "I relate," choosing to work through that response instead of their own, or reporting the response as inappropriate (Fig. 1).

2.3. Measures

Participants completed a brief screening form to determine eligibility (described below) and more detailed assessments at baseline, midpoint (week 4), end of treatment (week 8), and follow-up (week 16). The baseline, week 4, week 8, and week 16 assessments were identical and included the Depression, Anxiety, and Stress Scale (DASS), Patient Health Questionnaire - 9 (PHQ-9), Generalized Anxiety Disorder Questionnaire - 7 (GAD-7), as well as other measures, described on the clinicaltrials.gov page (NCT04226742). Here, we report only on changes in symptoms of depression and anxiety, with the DASS total score as our primary outcome and the PHQ-9 and GAD-7 as secondary outcome measures. We describe these measures below.

Fig. 1. Home screen of the Overcoming Thoughts platform "crowdsourced version."

2.3.1. Depression Anxiety and Stress Scale (DASS)

The DASS is a 42-item self-report instrument that measures three negative emotional constructs: depression, anxiety, and stress (Lovibond, 1998; Lovibond and Lovibond, 1995). Individuals use a 4-point Likert Scale to indicate the severity or frequency with which they experienced each item. The DASS yields three subscale scores. Scores range between 0 and 126 for the entire scale and 0–42 for each subscale. Higher scores on each of the constructs indicate a higher presence of depression, anxiety, or stress in participants' lives. The depression scale assesses dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest, and anhedonia. The anxiety scale measures autonomic arousal, skeletal muscle effects, situational anxiety, and anxious affect. The stress scale evaluates levels of chronic non-specific arousal. It gauges difficulty relaxing, nervous arousal, and being easily agitated, irritable, over-reactive, and impatient. The DASS has shown high internal consistency and concurrent validity in clinical trials and effectively distinguishes between features of depression, physical arousal, and psychological tension (Antony et al., 1998; Brown et al., 1997).

2.3.2. Patient Health Questionnaire - 9 (PHQ-9)

The PHQ-9 is a 9-item self-report instrument used to evaluate depressive symptoms (Kroenke et al., 2001). Respondents to the PHQ-9 questionnaire answer 9 questions indicating how often they have experienced depressive symptoms in the past 2 weeks and 1 question indicating how often any of the depressive symptoms impacted their ability to function in the last 2 weeks. Scores range from 0 to 27, higher scores on this instrument indicate a higher frequency of, or impairment from, depressive symptoms. The PHQ-9 has shown high validity and reliability in identifying symptoms of depression (Gilbody et al., 2007; Kroenke et al., 2001).

2.3.3. Generalized Anxiety Disorder - 7 (GAD-7)

The GAD-7 is a 7-item self-report instrument used to assess the severity of generalized anxiety disorder (Spitzer et al., 2006). The GAD-7 questionnaire includes 7 questions indicating how often one experienced anxiety symptoms in the past two weeks. Response options are “not at all,” “several days,” “more than half the days” and “nearly every day.” Scores range from 0 to 21, with higher scores indicating more severe anxiety symptoms. The GAD-7 has shown validity in primary care patients and the general population (Löwe et al., 2008; Mossman et al., 2017).

2.3.4. Demographics and mental health history

Participants completed a brief demographics and mental health history form, including age, gender, race, and ethnicity, as well as modules from the Structured Clinical Interview for DSM-5 (SCID-5; Shankman et al., 2018) asking if they had ever received diagnoses of bipolar, dissociative, substance use, or psychotic disorders.

2.3.5. Platform usage

We collected data on participants' usage of the “Overcoming Thoughts” platform during the eight-week study period. Usage data includes the number of “thoughts” each participant started and the number of questions answered for each thought. A “thought” on the platform refers to a user responding to the initial prompt on the home screen and selecting either “explore thoughts to feel better” or “take steps to feel better.” "Question responses" refers to how many answers users generated for the subsequent prompts that “work through” the initial thoughts (between zero and eight). We determined platform usage using log data resulting from user behavior while logged into the platform. The platform also collected usage behavior for users who were not logged into the platform. To mitigate the possibility that some users might use the platform while not logged in, we matched participants' identities across sessions using their device IP addresses. However, we could not detect participants' usage if they were not logged in and accessed the platform on a device with a different IP address. As such, our usage metrics might underestimate some participants' usage.

2.3.6. Safety

We tracked participant safety during the trial in two ways. First, we tracked self-reported suicidality, risk assessments conducted, and the determinations from the risk assessments. Self-reported suicidality was identified by evaluating any participant who reported a PHQ-9 item 9 $>$ 0 at any of the assessment time points, expressed suicidality to a member of our study team, or entered content into the platform expressing suicidality. Second, we calculated the percentage of participants in each condition who experienced clinically meaningful worsening using the meaningful change criteria proposed by Jacobson and Truax (1991); i.e. 1.96 times the standard error of the change score.

2.4. Eligibility criteria

Participants were eligible to participate in the study if they met the following criteria: (1) displayed significant mood and anxiety symptoms as defined by PHQ-9 or GAD-7 scores of at least 10, (2) were able to speak and read English as the intervention platform and all assessments were administered in English, (3) were at least 18 years of age at baseline, and (4) were currently residing in the United States. Eligibility criteria for PHQ-9 and GAD-7 were based on scores corresponding to moderate levels of depression or anxiety respectively (Kroenke et al., 2001; Spitzer et al., 2006), and therefore commonly used in clinical practice to identify individuals for whom treatment is warranted.

Participants were ineligible if they had severe suicidality, as indicated by ideation, plan, and intent. Current suicidality was assessed through both the PHQ-9 item 9 and a suicide severity assessment based on the Columbia-Suicide Severity Rating Scale (Chung et al., 2023; Posner et al., 2011). All participants who scored the PHQ-9 item 9 $>$ 0 were contacted for a subsequent risk assessment.

2.5. Recruitment

Participants were recruited from Mental Health America's Screening-to-Supports platform (S2S), an online screening platform that conducts approximately 3000 screens per day. Individuals who screened positive for depression (PHQ-9) or anxiety (GAD-7) on the screening platform received online study advertisements to test an online platform for depression and anxiety. Clicking these advertisements led to a REDCap form with screening questions, including demographics, contact information, age, diagnostic status, and English proficiency. Individuals deemed eligible for the study received a consent form, a baseline outcome questionnaire, and a link to schedule a telephone diagnostic assessment involving SCID-5 modules and a brief onboarding orientation.

Individuals were determined to have entered the study if they completed the baseline questionnaire, diagnostic assessment, and onboarding session. Participants were randomly assigned to a condition at consent using a predetermined randomization scheme (1:1); because study enrollment was based on completing the baseline and the onboarding session, this led to slightly unequal assignment to conditions, with 56 participants in the crowdsourced group and 51 participants in the non-crowdsourced group. All interactions between participants and the research team were fully blinded with regard to experimental conditions.

2.6. Analysis plan

We ran linear mixed models examining the change over time in the DASS, PHQ-9, and GAD-7, controlling for age, sex, race, and ethnicity, with a participant identifier as a random intercept (Preacher et al., 2006). We analyzed all 107 participants according to intent-to-treat principles. We did not impute missing outcome data because linear

mixed models' estimates for longitudinal clinical trial data are not improved by data imputation (Chakraborty and Gu, 2009). However, if <25 % of the items in the scale were missing, we imputed missing data within scales using the mean of the non-missing items for that participant for that scale (this occurred four times, combined across the DASS, PHQ-9, and GAD-7 and all time points).

We aimed to have 50 participants complete most study time points in each condition, allowing the main analysis to detect differences across conditions in change over time of Cohen's d $\geq$ 0.50 with 80 % power. While we expected to observe a smaller effect size than d = 0.50, this was a pilot study intended to determine feasibility and preliminary effectiveness, rather than a confirmatory effectiveness trial. To obtain Cohen's d from a binary predictor's regression coefficient (or the interaction coefficient between two binary predictors) we divided the coefficient by its standard deviation at baseline (Feingold, 2009). We used the $p < 0.05$ criterion for statistical significance. We also assessed clinically meaningful improvement during the study, defined as a statistically significant improvement; i.e. 1.96 times the standard error of the change score (Jacobson and Truax, 1991).

2.7. Openness

To analyze and visualize data, we used R version 4.3.1 (R Core Team, 2015), the *lme4* v1.1-35.1 package (Bates et al., 2015), and the *tidyverse* package v2.0.0 (Wickham and RStudio, 2021). We prospectively preregistered the study (clinicaltrials.gov: NCT04226742). The study data and analysis code are available at https://osf.io/5bq2r/ (Kaveladze, 2024).

Table 1
Baseline characteristics among participants meeting criteria for inclusion in analyses.

Characteristic *Mean* (*SD*); *n* (%)	Non-crowdsourced, *n* = 51	Crowdsourced, *n* = 56	*p*-Value
Age	34.59 (13.36)	33.30 (11.43)	0.77
Sex			0.77
Female	32 (63 %)	32 (57 %)	
Male	18 (35 %)	23 (41 %)	
Non-binary/other	1 (2.0 %)	1 (1.8 %)	
Race			0.40
White	31 (61 %)	37 (66 %)	
Black or African American	4 (7.8 %)	7 (13 %)	
More than one race	6 (12 %)	5 (8.9 %)	
Asian	4 (7.8 %)	6 (11 %)	
Unknown/not reported	5 (9.8 %)	1 (1.8 %)	
American Indian/ Alaska Native	1 (2.0 %)	0 (0 %)	
Ethnicity			0.12
Hispanic or Latino	13 (25 %)	8 (14 %)	
NOT Hispanic or Latino	37 (73 %)	48 (86 %)	
Unknown/not reported	1 (2.0 %)	0 (0 %)	
Diagnosis (SCID-5)	5 (9.8 %)	4 (7.1 %)	0.73
Mental Health Literacy Score	133.39 (12.70)	133.19 (11.54)	0.95
DASS Depression Subscale	24.35 (11.35)	24.50 (10.15)	0.96
DASS Anxiety Subscale	16.36 (9.68)	16.93 (9.82)	0.69
DASS Total Scale	63.24 (24.26)	65.68 (23.78)	0.66
GAD-7	12.47 (5.18)	13.58 (4.84)	0.31
PHQ-9	16.47 (5.49)	16.37 (5.36)	0.94

Means were compared using Wilcoxon rank sum tests, and frequencies were compared using Fisher's exact tests. Scores on the Mental Health Literacy Scale range from 35 to 160 (O'Connor and Casey, 2015).

3. Results

We enrolled 107 participants from February to November 2021. Baseline values of all variables did not significantly differ by condition (Table 1, *ps* > 0.10), indicating acceptable randomization. The proportion of participants who dropped out of the study at any time point also did not differ across conditions (*ps* > 0.62), as shown in the CONSORT diagram in Fig. 2.

3.1. Change in depression and anxiety across conditions

On average, adjusting for covariates and experimental condition assignment, DASS total scores decreased between baseline and each time point; *ps* < 0.01. Contrary to our hypotheses, participants assigned to the crowdsourced (intervention) condition did not report significantly greater decreases in DASS total scores between baseline and any follow-up than those assigned to the non-crowdsourced (control) condition. Fig. 3 visualizes differences across conditions and Table 2 shows the model output. Similarly, participants assigned to the crowdsourced condition did not report significantly greater decreases in GAD-7 scores between baseline and four-week ($b = 0.61$; $d = 0.11$; 95 % CI, −1.64, 2.85; $p = 0.60$), eight-week ($b = -0.61$; $d = -0.11$; 95 % CI, −2.95, 1.73; $p = 0.61$), or 16-week follow-ups ($b = 0.32$; $d = 0.06$; 95 % CI, −2.03, 2.67; $p = 0.79$) than those assigned to the non-crowdsourced condition. Finally, those assigned to the crowdsourced condition did not report significantly greater decreases in PHQ-9 scores between baseline and four-week ($b = 1.16$; $d = 0.19$; 95 % CI, −1.06, 3.38; $p = 0.31$), eight-week ($b = -0.26$; $d = -0.04$; 95 % CI, −2.57, 2.05; $p = 0.83$), or 16-week follow-ups ($b = 0.24$; $d = 0.04$; 95 % CI, −2.09, 2.57; $p = 0.84$) than those assigned to the non-crowdsourced condition.

Most participants experienced a clinically meaningful improvement in DASS total score, including 64.6 % at week 4, 72.6 % at week 8, and 70.0 % at week 16. Similarly, 53.7 % experienced a clinically meaningful improvement in their GAD-7 score at week 4, 56.2 % did at week 8, and 59.7 % did at week 16. 72.8 % experienced a clinically meaningful improvement in their PHQ-9 score at week 4, 68.5 % did at week 8, and 71.8 % did at week 16. None of these improvement rates differed significantly by intervention condition (*ps* > 0.20).

3.2. Platform usage

Fig. 4 shows the total number of thoughts entered into the tool per participant. Participants completed all eight questions for 74 % of the thoughts they entered and zero out of eight questions for 15 % of the thoughts. Cumulatively, participants chose the “take steps to feel better” path for 43 % of their thoughts, and the “explore thoughts to feel better” path for 57 % (excluding the 94 initial thoughts for which participants did not answer any questions as we cannot know which condition they chose). Among those who completed all eight questions, the median duration spent working with each thought was 6 min. Within the 56-day intervention period, participants used the tool a median of 2 days (mean = 4.79, SD = 5.96). The mean number of days between users' first interaction with the tool and their last was 28.00 (SD = 21.35). Participants who entered more thoughts into the platform did not experience greater or weaker reductions in DASS total score, PHQ-9 score, or GAD-7 score over eight weeks (ps > 0.10). Finally, the number of thoughts entered into the tool did not differ across conditions ($p = 0.51$).

3.3. Safety

During the trial, 123 instances of suicidality were reported on the PHQ-9 item 9. Of these instances, 102 were determined to be at low risk of suicidality based on their PHQ-9 item 9 score being no higher than 1 at each previous measurement point, as well as the suicidality risk evaluation we conducted during the screening process. We followed up with the remaining 21 participants to complete a suicide risk assessment

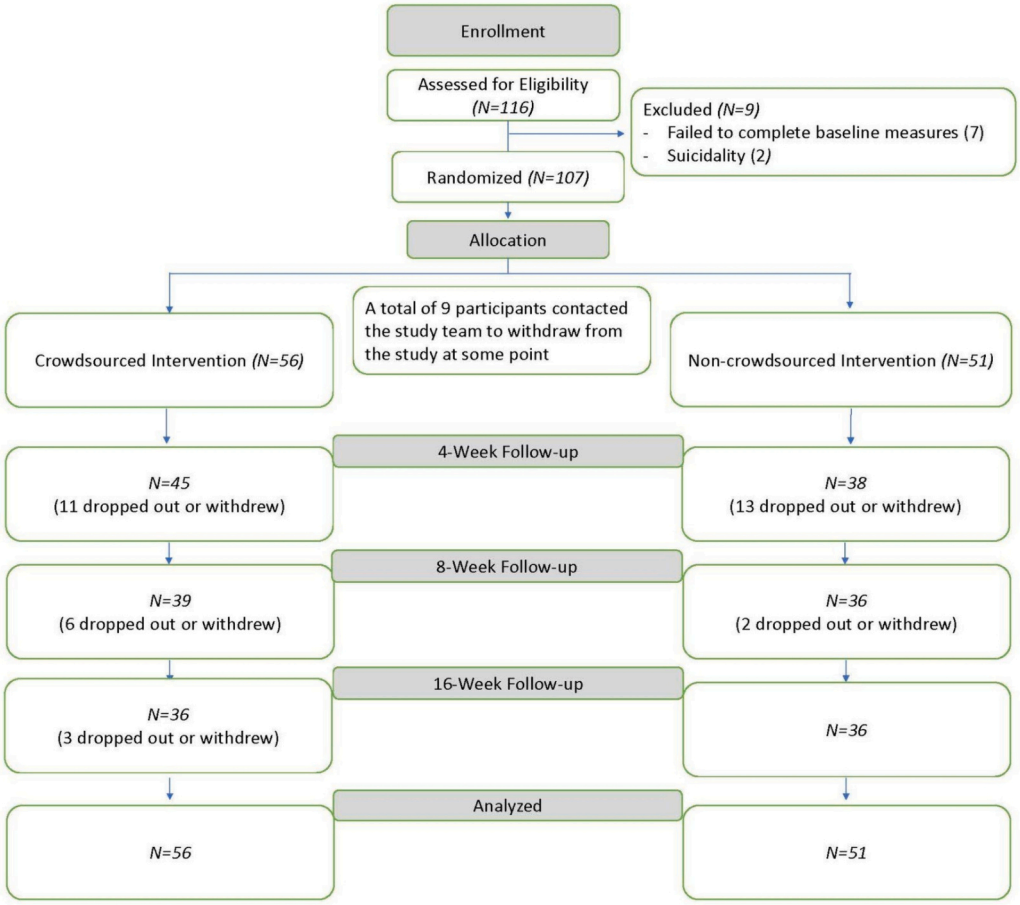


Fig. 2. CONSORT diagram.

as indicated by our protocol; this included completing the Columbia Suicide Severity Rating Scale as described in Section 2.2. In all of these instances, participants were found not to be at imminent risk for suicide and no additional intervention was required. In one instance, a participant e-mailed the study team to report suicidal ideation, but without intent or plan, and the participant was determined to be at low risk. No other adverse events were reported.

During the eight-week trial period, we monitored individuals for clinically meaningful worsening and compared rates between the conditions at weeks 4 and 8. For the DASS total score, 19.5 % had a clinically meaningful worsening at week 4 and 13.7 % did at week 8. For the GAD-7, 14.6 % had a clinically meaningful worsening at week 4 and 23.3 % did at week 8. For the PHQ-9, 11.0 % had a clinically meaningful worsening at week 4 and 7.0 % did at week 8. None of these rates of worsening differed significantly by intervention condition ($ps > 0.20$).

4. Discussion

Advancing innovative models of human support could enhance future DMHIs. We conducted a two-armed pilot randomized controlled trial of crowdsourced peer support integrated into a novel, highly accessible DMH platform. Our goals were to identify problems and opportunities with the study design for future fully powered trials and to explore ways to improve the intervention platform and crowdsourced support feature. We also collected data on target outcomes and piloted an analytic strategy that would apply to a fully-powered trial. Although participants' self-reported anxiety and depression decreased on average, the crowdsourced version of the platform was not significantly more effective than the non-crowdsourced version, against our prediction. Because this was a pilot study with a small sample, these analyses lacked statistical power to detect small or moderately large differences across conditions. Power was further reduced by poor engagement with the intervention platform; the number of times participants interacted with the tools ranged widely, but was lower than intended on average.

Low engagement is the norm among DMHIs (Baumel et al., 2019), and "Overcoming Thoughts" was no exception. Most study participants used the platform much less often than instructed during the eight-week study period, suggesting that few people struggling with anxiety or depression are motivated to interact with DMHIs like "Overcoming Thoughts" consistently. This is an important problem for future effectiveness trials of DMHIs to consider; participants who do not use the interventions to which they are assigned are not useful for comparisons across conditions. Trials should have clear prior estimates of interventions' intended "dose" (i.e., how much time and attention one pays to an intervention) and ensure that their interventions are engaging enough for most participants to receive that dose.

One reason for this low repeated usage could be that "Overcoming Thoughts" was not sufficiently designed for repeat engagement. Indeed, in qualitative work stemming from this trial, several participants reported frustration with the platform's repetitiveness and simplicity (Shkel et al., 2023). Engagement-focused design elements, such as push notifications and a compelling onboarding experience, might increase platform use (Boucher and Raiker, 2024). Another possible reason for the low engagement is that participants were inadequately incentivized. Participants were paid to complete study measurements, but not to interact with the tool. Additional monetary and alternative incentives (i.e., encouraging messages) may have boosted engagement and impact (Griffith Fillipo et al., 2022). On the other hand, too much outside-intervention incentive may reduce findings' generalizability to real-world settings. A final reason for the low engagement might be that

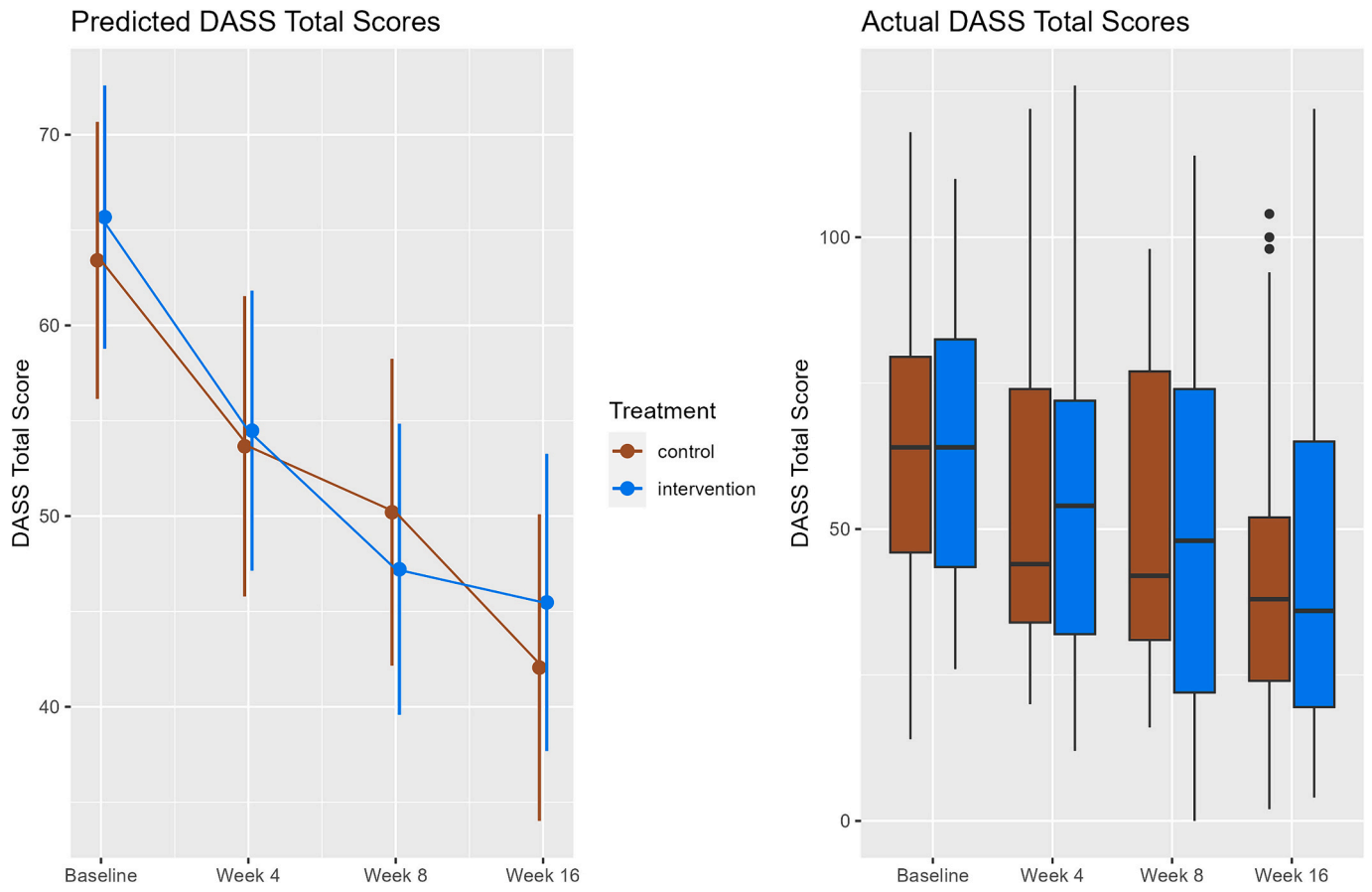


Fig. 3. Change in depression, anxiety, and stress (DASS total score) across conditions, shown with marginal effects and boxplots of actual sample data.
The left panel shows estimates from the linear mixed model, adjusting for age, sex, race, ethnicity, and diagnostic status; the error bars show the estimates' 95 % confidence interval. The right panel illustrates the observed data; the boxes' midpoints show the sample medians and their upper and lower bounds show the 25th and 75th percentiles. Baseline $n = 107$, week 4 $n = 83$, week 8 $n = 75$, week 16 $n = 72$.

the tool required too much effort (i.e., responding to several written prompts with little feedback). In previous work, asking participants to provide written responses within a DMHI led many to drop out (Dobias et al., 2022). Yet, in this study, once participants began to work with a thought, they responded to all eight prompts 74 % of the time, which might suggest that they did not find the writing prompts to be overly burdensome.

A different way to interpret the low platform usage rate in our study, rather than the intervention lacking appeal, is that many participants were satisfied with, and sufficiently benefitted from, just one or a few sessions. Several participants reported using the platform less as their symptoms alleviated and as they became more familiar with the questions and could complete the exercises in their heads, as is discussed in more detail in Shkel et al. (2023). Maybe simple online DMHIs like "Overcoming Thoughts" should be primarily viewed as single-session interventions, rather than multi-session journeys (Bunge et al., 2016; Lokman et al., 2017; Schleider et al., 2020). If so, maximizing engagement and effectiveness within a single session may be a more productive design goal than increasing the number of sessions participants complete.

Despite the low rates of return use, many participants interacted with the platform deeply. Interviews with a subset of study participants revealed that they perceived substantial benefits of the intervention, like increasing their coping skills, prompting self-reflection, and slowing down their thinking to break thoughts down (these data are presented in Shkel et al., 2023). The primary outcomes we used may not capture some of the benefits that can come from brief DMHIs like "Overcoming Thoughts." More work might consider the kinds of change that people who experience anxiety and depression are most interested in when using DMHIs, aiming to tailor DMHIs to users' valued outcomes and goals (Chevance et al., 2020; Cuijpers, 2019).

4.1. *Strengths and limitations*

The study had several strengths, including a randomized controlled trial design with full blinding. The sample was more demographically diverse, in terms of age, gender, race, and baseline mental health status, than most participant samples in studies of DMHIs (Gunawardena et al., 2024), and participants' average mental health literacy was typical of community samples (O'Connor and Casey, 2015). Also, the study dropout rate was typical of clinical trials of DMHIs; we observed 33 % dropout over 16 weeks, while average dropout rates are 25 % for guided DMHIs and 29 % for unguided (Karyotaki et al., 2021). Another strength was that the sample was recruited from an online mental health screening platform; this improves generalizability to settings where interventions like "Overcoming Thoughts" might be implemented for broad reach.

One limitation of this work was the lack of a passive control group. A passive control would have allowed us to compare each condition to no treatment, separately from other potential influences like regression to the mean. Participants in both conditions showed statistically significant decreases in symptoms of depression and anxiety at all follow-up time points, compared to baseline. Yet, this does not necessarily mean the intervention was successful, as similar decreases in depressive symptoms are typical of control condition participants in trials of DMHIs (Tong et al., 2023). However, there were advantages to using an active

Table 2
Mixed-effects model predicting DASS total score.

DASS total score				
Predictors	Estimates	*CI*	*p*	*d*
(Intercept)	107.32	60.93–153.71	<**0.001**	
Treatment [crowdsourced]	2.57	−7.57–12.70	0.617	0.10
Time [Week 4]	−10.07	−16.84 to −3.30	**0.004**	−0.39
Time [Week 8]	−13.53	−20.50 to −6.57	<**0.001**	−0.53
Time [Week 16]	−21.91	−28.88 to −14.94	<**0.001**	−0.85
Age	−0.23	−0.62–0.17	0.261	
Sex [Male]	0.80	−8.83–10.43	0.869	0.03
Sex [non-binary/other]	−12.65	−45.83–20.53	0.450	−0.49
Race [Asian]	−22.91	−69.65–23.84	0.333	−0.89
Race [Black or African American]	−40.84	−87.44–5.76	0.085	−1.59
Race [more than one race]	−38.81	−85.66–8.03	0.103	−1.51
Race [unknown/not reported]	−41.39	−91.80–9.03	0.106	−1.61
Race [White]	−39.97	−85.24–5.30	0.083	−1.55
Ethnicity [NOT Hispanic or Latino]	−0.39	−13.20–12.41	0.951	−0.02
Ethnicity [unknown/not reported]	−14.05	−66.88–38.79	0.599	−0.55
Diagnostic Status	23.65	5.81–41.49	**0.010**	0.92
Treatment [crowdsourced] × time [Week 4]	−0.91	−10.10–8.28	0.846	−0.03
Treatment [crowdsourced] × time [Week 8]	−4.60	−14.17–4.97	0.345	−0.18
Treatment [crowdsourced] × time [Week 16]	2.01	−7.70–11.72	0.684	0.08

Random effects	
σ^2	232.17
$\tau_{00\ study\_id}$	428.20
ICC	0.65
$N_{\ study\_id}$	107
Observations	334
Marginal R^2/Conditional R^2	0.186/0.714

A linear mixed model testing how change in the DASS total score over 16 weeks differed across conditions. The dependent variable, DASS total score, is scored from 0 to 120. There are two experimental conditions and four time points (baseline, 4-week follow-up, 8-week follow-up, and 16-week follow-up). Race estimates are relative to the American Indian/Alaska Native group.

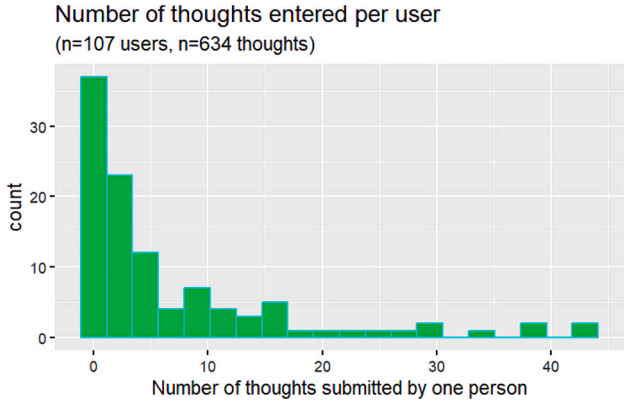


Fig. 4. Number of thoughts entered per participant during the study.
The histogram shows the number of thoughts each user entered into the platform during the 8-week intervention period. Participants entered a median of 3 thoughts and the mean was 6.99 (SD = 9.78).

comparison condition (the non-crowdsourced version of the platform). For one, it allowed us to isolate the effect of crowdsourcing, our intervention element of interest. An active comparison condition also enabled us to treat a DMHI with potential for real-world implementation as a starting point for future improvement (Goldberg et al., 2023).

Another limitation was that, as a pilot, the study was not sufficiently powered to detect the effect sizes we might realistically expect. However, even if our application of crowdsourced peer support was perfect and engagement was ideal, our sample size ($n = 107$) would still likely be too small to detect the effect of the crowdsourced intervention component, if one exists. The differences in user experience between conditions were slight (especially relative to studies comparing DMHIs to passive controls) and were made even smaller by participants' low frequency of interacting with the intervention. For context, trials of digital single-session interventions often aim to identify effects of $d < 0.20$ (Schleider et al., 2022), and trials examining a specific intervention component of such an intervention would require smaller effect sizes than that. As a result, our study design needs substantial changes to be feasible, such as facilitating the participant enrollment and data collection process to enroll more participants without a large team of research staff.

A final limitation was that we conducted our assessments separately from the intervention platform. Had we built assessments into the platform, we might have learned more about its in-the-moment impacts. Our intervention usage data also had limitations; for example, although we instructed users to log in to their accounts when using the platform, they did not always do so. We partly addressed this issue by matching users' login information to their devices' IP Addresses, but this likely missed times when users accessed the platform from other devices.

4.2. Future directions

Few participants in our study used "Overcoming Thoughts" as we intended it to be used. Future trials need to be more thoughtful about optimizing user engagement. To make DMHIs that are flexible to different users' preferences, we need to understand the ways and cadences in which people want to interact with these tools. For example, some users might have preferred a more social experience than the crowdsourcing feature in "Overcoming Thoughts" provided. In contrast, others may have felt that the crowdsourcing feature was already too socially interactive for such a personal task. Our qualitative analysis of study participant interviews revealed that users of the non-crowdsourced platform were more likely to comment on the usefulness of the summary feature, which provided an overview of their entries into the platform, than users of the crowdsourced platform (Shkel et al., 2023). This suggests that crowdsourcing features might draw users away from other aspects of the platform. Adding more features and tools to a DMHI may not always be beneficial, and designers should consider tradeoffs between increased content and a tool's functionality and appeal. People might differ in these preferences based on demographic factors, diagnostic criteria, or personal interests.

Our platform used a relatively minimalistic application of crowdsourcing: showing responses from previous users and allowing one to build on those responses, add an "I relate" to that thought's count, or report the thought as inappropriate. We may have been able to better integrate crowdsourcing into this platform. For one, we could have designed the platform to increase interaction between participants (e.g., allowing users to comment on others' contributions). We also could have included more social cues in the system (e.g., displaying users' names and profile pictures). Future implementations of crowdsourcing in DMHIs should pay more attention to how design choices might help provide users with a sense of community and validation. These interpersonal elements might make DMHIs more impactful and encourage more frequent use (Kwok et al., 2024).

Integrating crowdsourced peer support into DMHIs is an opportunity with case-specific advantages and risks. Future work should explore the

costs and benefits of different forms of crowdsourced support within DMHIs. While our implementation of crowdsourced support did not demonstrate utility, other forms of crowdsourced support warrant further study. It is also important to examine how crowdsourced support may be inferior to other kinds of peer support. Like other forms of support provided online via anonymous peers, crowdsourced support's quality (both informational and empathic) is unpredictable, it might feel less personal, and it often relies on many motivated individuals' participation (Zhu and Stephens, 2019).

In addition to improving human support, it is important to explore other innovative intervention strategies that might increase the reach of accessible public mental health tools like "Overcoming Thoughts." Since developing the tool for this study, Mental Health America has continued to iterate on "Overcoming Thoughts." Recently, they integrated a large language model (LLM) into the "Overcoming Thoughts" platform to facilitate cognitive restructuring. The LLM-supported version showed encouraging results regarding user satisfaction and success in cognitive restructuring (Sharma et al., 2024).

As noted above, full-power trials aiming to identify the effects of particular intervention elements, as ours did, require massive sample sizes. Because popular online spaces like Mental Health America reach so many people (between 2021 and 2022, "Overcoming Thoughts" was accessed via Mental Health America over 100,000 times), they would be invaluable settings for testing future DMHIs. Researchers interested in conducting full-powered trials with study designs like ours should consider collaborating with such popular settings. In addition, "Overcoming Thoughts" might offer a useful example for other studies aiming to isolate particular components of brief targeted interventions.

4.3. Conclusion

Accessible digital mental health tools delivered in popular online help-seeking spaces can offer users an important first step to improvement. This pilot test aimed to provide insights into how future full-powered trials examining the effects of crowdsourced support (and other intervention components) in DMHIs can be conducted. Our hypothesis, that a crowdsourced peer support feature in a DMHI would cause greater reductions in self-reported depression and anxiety symptoms, was not supported, although this comparison was substantially underpowered. The number of times that participants interacted with the tool was low on average, suggesting that its design needs to be made more engaging. Alternatively, it may make more sense as a single-session intervention than as a tool for repeated use. Full-powered trials using similar designs to ours should pay particular attention to optimizing engagement with interventions and the feasibility of attaining sufficient sample size for appropriate statistical power.

Abbreviations

DASS	Depression Anxiety and Stress Scale
PHQ-9	Patient Health Questionnaire
GAD-7	Generalized Anxiety Disorder Questionnaire
SCID-5	Structured Clinical Interview for DSM-5

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Stephen Schueller reports financial support was provided by National Institute of Mental Health of the National Institutes of Health. Benjamin Kaveladze reports financial support was provided by National Institute of Mental Health of the National Institutes of Health. Stephen Schueller reports a relationship with Headspace and Kooth that includes: board membership and consulting or advisory. Stephen Schueller reports a relationship with Boehringer Ingelheim that includes: consulting or advisory. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research reported in this publication was supported by the National Institutes of Health including the National Institute of Mental Health under the award number R34MH113616 (PI: Schueller) and the National Center for Research Resources and the National Center for Advancing Translational Sciences through Grant UL1 TR001414. BK received funding from the NIMH Multidisciplinary Training Program in Digital Mental Health T32 fellowship (1T32MH115882-01) during the preparation of this manuscript. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH. Data used in the preparation of this manuscript are available in the National Institute of Mental Health (NIMH) Data Archive (NDA), dataset identifier(s): 10.15154/g49h-d486. The authors would also like to acknowledge Zoë Dodge-Rice, who assisted with the study management and recruitment, and Gavin Green, who assisted in data cleaning and analysis. Conflict of interest: SMS serves on the Scientific Advisory Board for Headspace and Kooth, for which he receives compensation, and has consulted for Otsuka Pharmaceuticals and Boehringer Ingelheim for work unrelated to this project.

References

Andersson, G., Carlbring, P., Titov, N., Lindefors, N., 2019. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can. J. Psychiatry 64 (7), 465–470. https://doi.org/10.1177/0706743719839381.

Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W., Swinson, R.P., 1998. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol. Assess. 10 (2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Baumel, A., 2015. Online emotional support delivered by trained volunteers: users' satisfaction and their perception of the service compared to psychotherapy. Journal of Mental Health (Abingdon, England) 24 (5), 313–320. https://doi.org/10.3109/09638237.2015.1079308.

Baumel, A., Muench, F., Edan, S., Kane, J.M., 2019. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J. Med. Internet Res. 21 (9), e14567 https://doi.org/10.2196/14567.

Boucher, E.M., Raiker, J.S., 2024. Engagement and retention in digital mental health interventions: a narrative review. BMC Digital Health 2 (1), 52. https://doi.org/10.1186/s44247-024-00105-9.

Brown, T.A., Chorpita, B.F., Korotitsch, W., Barlow, D.H., 1997. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. Behav. Res. Ther. 35 (1), 79–89. https://doi.org/10.1016/s0005-7967(96)00068-x.

Bunge, E.L., Williamson, R.E., Cano, M., Leykin, Y., Muñoz, R.F., 2016. Mood management effects of brief unsupported internet interventions. Internet Interv. 5, 36–43. https://doi.org/10.1016/j.invent.2016.06.001.

Chakraborty, H., Gu, H., 2009. A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials with Missing Values, RTI Press. http://www.ncbi.nlm.nih.gov/books/NBK538904/.

Chevance, A., Ravaud, P., Tomlinson, A., Le Berre, C., Teufer, B., Touboul, S., Fried, E.I., Gartlehner, G., Cipriani, A., Tran, V.T., 2020. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. Lancet Psychiatry 7 (8), 692–702. https://doi.org/10.1016/S2215-0366(20)30191-7.

Chung, T.H., Hanley, K., Le, Y.-C., Merchant, A., Nascimento, F., De Figueiredo, J.M., Wilcox, H.C., Coryell, W.H., Soares, J.C., Selvaraj, S., 2023. A validation study of PHQ-9 suicide item with the Columbia Suicide Severity Rating Scale in outpatients with mood disorders at National Network of Depression Centers. J. Affect. Disord. 320, 590–594. https://doi.org/10.1016/j.jad.2022.09.131.

Conn, V.S., Algase, D.L., Rawl, S.M., Zerwic, J.J., Wyman, J.F., 2010. Publishing pilot intervention work. West. J. Nurs. Res. 32 (8), 994–1010. https://doi.org/10.1177/0193945910367229.

Cuijpers, P., 2019. Targets and outcomes of psychotherapies for mental disorders: an overview. World Psychiatry 18 (3), 276–285. https://doi.org/10.1002/wps.20661.

De Choudhury, M., De, S., 2014. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. In: Proceedings of the International AAAI Conference on Web and Social Media, 8(1). https://doi.org/10.1609/icwsm.v8i1.14526. Article 1.

Dobias, M.L., Morris, R.R., Schleider, J., 2022. Single-session interventions embedded within Tumblr: acceptability, feasibility, and utility study. JMIR Formative Research 6 (7), e39004. https://doi.org/10.2196/39004.

Farchione, T.J., Fairholme, C.P., Ellard, K.K., Boisseau, C.L., Thompson-Hollands, J., Carl, J.R., Gallagher, M.W., Barlow, D.H., 2012. Unified protocol for transdiagnostic

treatment of emotional disorders: a randomized controlled trial. Behav. Ther. 43 (3), 666–678. https://doi.org/10.1016/j.beth.2012.01.001.
Feingold, A., 2009. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol. Methods 14 (1), 43–53. https://doi.org/10.1037/a0014699.
Fortuna, K.L., Naslund, J.A., LaCroix, J.M., Bianco, C.L., Brooks, J.M., Zisman-Ilani, Y., Deegan, P., 2020. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR mental health 7 (4), e16460.
Gilbody, S., Richards, D., Brealey, S., Hewitt, C., 2007. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J. Gen. Intern. Med. 22 (11), 1596–1602. https://doi.org/10.1007/s11606-007-0333-y.
Goldberg, S.B., Sun, S., Carlbring, P., Torous, J., 2023. Selecting and describing control conditions in mobile health randomized controlled trials: a proposed typology. Npj Digital Medicine 6 (1), 1–8. https://doi.org/10.1038/s41746-023-00923-7.
Griffith Fillipo, I.R., Pullmann, M.D., Hull, T.D., Zech, J., Wu, J., Litvin, B., Chen, S., Arean, P.A., 2022. Participant retention in a fully remote trial of digital psychotherapy: comparison of incentive types. Frontiers in Digital Health 4, 963741. https://doi.org/10.3389/fdgth.2022.963741.
Gulliver, A., Banfield, M., Morse, A.R., Reynolds, J., Miller, S., Galati, C., 2019. A peer-led electronic mental health recovery app in a community-based public mental health service: pilot trial. JMIR Formative Research 3 (2), e12550. https://doi.org/10.2196/12550.
Gunawardena, T., Bartholmae, M.M., Karpov, M.V., Dod, R., Ahuja, K., Rajendran, A., Kathrotia, M., Dodani, S., 2024. Predictors for interest to participate in digital mental health therapy: a cross-sectional survey of individuals with anxiety and depression. BMC Digital Health 2 (1), 21. https://doi.org/10.1186/s44247-024-00080-1.
Jacobson, N.C., Yom-Tov, E., Lekkas, D., Heinz, M., Liu, L., Barr, P.J., 2022. Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: evidence from internet search behavior in a large U.S. cohort. J. Psychiatr. Res. 145, 276–283. https://doi.org/10.1016/j.jpsychires.2020.11.010.
Jacobson, N.S., Truax, P., 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol. 59 (1), 12–19. https://doi.org/10.1037//0022-006x.59.1.12.
Johnson, J.A., Sanghvi, P., Mehrotra, S., 2022. Technology-based interventions to improve help-seeking for mental health concerns: a systematic review. Indian J. Psychol. Med. 44 (4), 332–340. https://doi.org/10.1177/02537176211034578.
Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F.M.G., Furukawa, T.A., Cuijpers, P., Riper, H., Patel, V., Mira, A., Gemmil, A.W., Yeung, A.S., Lange, A., Williams, A.D., Mackinnon, A., Geraedts, A., van Straten, A., Meyer, B., Björkelund, C., Forsell, Y., Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, 2021. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 78 (4), 361–371. https://doi.org/10.1001/jamapsychiatry.2020.4364.
Kaveladze, B., 2024. Crowdsourcing Integrated in a Digital Mental Health Platform for Anxiety and Depression: A Pilot Randomized Controlled Trial. https://osf.io/5bq2r/.
Kaveladze, B., Nguyen, T., Schueller, S.M., 2021. Online Mental Health Diagnostic Tests Can Provide Users Better Guidance: Examining Data from 1,045,562 Screens and Surveys. Association for Psychological Science Conference. https://imgur.com/a/xtvAght.
Kim, J., Aryee, L.M.D., Bang, H., Prajogo, S., Choi, Y.K., Hoch, J.S., Prado, E.L., 2023. Effectiveness of digital mental health tools to reduce depressive and anxiety symptoms in low- and middle-income countries: systematic review and meta-analysis. JMIR Mental Health 10, e43066. https://doi.org/10.2196/43066.
Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
Kwok, I., Freedman, M., Kamsickas, L., Lattie, E.G., Yang, D., Moskowitz, J.T., 2024. The SoCAP (Social Communication, Affiliation, and Presence) taxonomy of social features: scoping review of commercially available eHealth apps. J. Med. Internet Res. 26, e49714.
Lokman, S., Leone, S.S., Sommers-Spijkerman, M., van der Poel, A., Smit, F., Boon, B., 2017. Complaint-directed mini-interventions for depressive complaints: a randomized controlled trial of unguided web-based self-help interventions. J. Med. Internet Res. 19 (1), e4 https://doi.org/10.2196/jmir.6581.
Lovibond, P.F., 1998. Long-term stability of depression, anxiety, and stress syndromes. J. Abnorm. Psychol. 107 (3), 520–526. https://doi.org/10.1037//0021-843x.107.3.520.
Lovibond, P.F., Lovibond, S.H., 1995. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav. Res. Ther. 33 (3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U.
Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., Herzberg, P.Y., 2008. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med. Care 46 (3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093.
Mental Health America, 2022. Lessons from MHA Screening, Mental Health America. https://mhanational.org/lessons-MHA-Screening.
Mongelli, F., Georgakopoulos, P., Pato, M.T., 2020. Challenges and opportunities to meet the mental health needs of underserved and disenfranchised populations in the United States. FOCUS 18 (1), 16–24. https://doi.org/10.1176/appi.focus.20190028.
Morris, R., 2015. Crowdsourcing mental health and emotional well-being [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/97972.
Morris, R.R., Kouddous, K., Kshirsagar, R., Schueller, S.M., 2018. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J. Med. Internet Res. 20 (6) https://doi.org/10.2196/10148.
Morris, R.R., Schueller, S.M., Picard, R.W., 2015. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. J. Med. Internet Res. 17 (3), e72 https://doi.org/10.2196/jmir.4167.
Mossman, S.A., Luft, M.J., Schroeder, H.K., Varney, S.T., Fleck, D.E., Barzman, D.H., Gilman, R., DelBello, M.P., Strawn, J.R., 2017. The Generalized Anxiety Disorder 7-item scale in adolescents with generalized anxiety disorder: signal detection and validation. Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists 29 (4), 227–234A.
Murphy, J.M., Nguyen, T., Lucke, C., Chiang, C., Plasencia, N., Jellinek, M., 2018. Adolescent self-screening for mental health problems; demonstration of an Internet-based approach. Acad. Pediatr. 18 (1), 59–65. https://doi.org/10.1016/j.acap.2017.08.013.
Nelson, C.B., Abraham, K.M., Walters, H., Pfeiffer, P.N., Valenstein, M., 2014. Integration of peer support and computer-based CBT for veterans with depression. Comput. Hum. Behav. 31, 57–64. https://doi.org/10.1016/j.chb.2013.10.012.
O'Connor, M., Casey, L., 2015. The Mental Health Literacy Scale (MHLS): a new scale-based measure of mental health literacy. Psychiatry Res. 229 (1–2), 511–516. https://doi.org/10.1016/j.psychres.2015.05.064.
Park, S.Y., Sigmon, C.N., Boeldt, D., Park, S.Y., Sigmon, C.A.N., Boeldt, D., 2022. A framework for the implementation of digital mental health interventions: the importance of feasibility and acceptability research. Cureus 14 (9). https://doi.org/10.7759/cureus.29329.
Posner, K., Brown, G.K., Stanley, B., Brent, D.A., Yershova, K.V., Oquendo, M.A., Currier, G.W., Melvin, G.A., Greenhill, L., Shen, S., Mann, J.J., 2011. The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am. J. Psychiatry 168 (12), 1266–1277. https://doi.org/10.1176/appi.ajp.2011.10111704.
Preacher, K.J., Curran, P.J., Bauer, D.J., 2006. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. J. Educ. Behav. Stat. 31 (4), 437–448. https://doi.org/10.3102/10769986031004437.
R Core Team, 2015. R: A Language and Environment for Statistical Computing [Computer Software]. R Foundation for Statistical Computing.
Ramos, G., Hernandez-Ramos, R., Taylor, M., Schueller, S.M., 2024. State of the science: using digital mental health interventions to extend the impact of psychological services. Behav. Ther. https://doi.org/10.1016/j.beth.2024.04.004.
Schleider, J., Dobias, M.L., Sung, J.Y., Mullarkey, M.C., 2020. Future directions in single-session youth mental health interventions. J. Clin. Child Adolesc. Psychol. 49 (2), 264–278. https://doi.org/10.1080/15374416.2019.1683852.
Schleider, J., Mullarkey, M.C., Fox, K.R., Dobias, M.L., Shroff, A., Hart, E.A., Roulston, C. A., 2022. A randomized trial of online single-session interventions for adolescent depression during COVID-19. Nature Human Behaviour 6 (2), 2. https://doi.org/10.1038/s41562-021-01235-0.
Schueller, S.M., Tomasino, K.N., Mohr, D.C., 2017. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin. Psychol. Sci. Pract. 24 (1), 27–45. https://doi.org/10.1037/h0101740.
Shankman, S.A., Funkhouser, C.J., Klein, D.N., Davila, J., Lerner, D., Hee, D., 2018. Reliability and validity of severity dimensions of psychopathology assessed using the Structured Clinical Interview for DSM-5 (SCID). Int. J. Methods Psychiatr. Res. 27 (1), e1590 https://doi.org/10.1002/mpr.1590.
Sharma, A., Choudhury, M., Althoff, T., Sharma, A., 2020. Engagement patterns of peer-to-peer interactions on mental health platforms. In: Proceedings of the International AAAI Conference on Web and Social Media, 14, pp. 614–625. https://doi.org/10.1609/icwsm.v14i1.7328.
Sharma, A., Rushton, K., Lin, I.W., Nguyen, T., Althoff, T., 2024. Facilitating Self-Guided Mental Health Interventions Through Human-Language Model Interaction: A Case Study of Cognitive Restructuring (arXiv:2310.15461). arXiv. https://doi.org/10.48550/arXiv.2310.15461.
Shkel, J., Green, G., Le, S., Kaveladze, B., Marcotte, V., Rushton, K., Nguyen, T., Schueller, S.M., 2023. Understanding users' experiences of a novel web-based cognitive behavioral therapy platform for depression and anxiety: qualitative interviews from pilot trial participants. JMIR Formative Research 7, e46062. https://doi.org/10.2196/46062.
Smith, C.E., Lane, W., Miller Hillberg, H., Kluver, D., Terveen, L., Yarosh, S., 2021. Effective strategies for crowd-powered cognitive reappraisal systems: a field deployment of the flip*doubt web application for mental health. Proceedings of the ACM on Human-Computer Interaction 5 (CSCW2), 417:1–417:37. https://doi.org/10.1145/3479561.
Smoktunowicz, E., Barak, A., Andersson, G., Banos, R.M., Berger, T., Botella, C., Dear, B. F., Donker, T., Ebert, D.D., Hadjistavropoulos, H., Hodgins, D.C., Kaldo, V., Mohr, D. C., Nordgreen, T., Powers, M.B., Riper, H., Ritterband, L.M., Rozental, A., Schueller, S.M., Carlbring, P., 2020. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. Internet Interv. 21, 100331 https://doi.org/10.1016/j.invent.2020.100331.
Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166 (10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092.
Titov, N., Dear, B.F., Nielssen, O., Wootton, B., Kayrouz, R., Karin, E., Genest, B., Bennett-Levy, J., Purtell, C., Bezuidenhout, G., Tan, R., Minissale, C., Thadhani, P., Webb, N., Willcock, S., Andersson, G., Hadjistavropoulos, H.D., Mohr, D.C., Kavanagh, D.J., Staples, L.G., 2020. User characteristics and outcomes from a national digital mental health service: an observational study of registrants of the

Australian MindSpot Clinic. The Lancet Digital Health 2 (11), e582–e593. https://doi.org/10.1016/S2589-7500(20)30224-7.
Titov, N., Hadjistavropoulos, H.D., Nielssen, O., Mohr, D.C., Andersson, G., Dear, B.F., 2019. From research to practice: ten lessons in delivering digital mental health services. Journal of Clin. Med. 8 (8), 8 https://doi.org/10.3390/jcm8081239.
Tong, A.C., Ho, F.S., Chu, O.H., Mak, W.W., 2023. Time-dependent changes in depressive symptoms among control participants in digital-based psychological intervention studies: meta-analysis of randomized controlled trials. J. Med. Internet Res. 25, e39029 https://doi.org/10.2196/39029.
Werntz, A., Amado, S., Jasman, M., Ervin, A., Rhodes, J.E., 2023. Providing human support for the use of digital mental health interventions: systematic meta-review. J. Med. Internet Res. 25, e42864 https://doi.org/10.2196/42864.
Wickham, H., RStudio, 2021. tidyverse: Easily Install and Load the "Tidyverse" (Version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=tidyverse.
Zhu, Y., Stephens, K.K., 2019. Online support group participation and social support: incorporating identification and interpersonal bonds. Small Group Res. 50 (5), 593–622. https://doi.org/10.1177/1046496419861743.

Glossary

Engagement: Interacting with core components of an intervention
Intervention: a program provided to someone to bring about a desired effect
Digital mental health intervention: an intervention delivered through a digital medium, such as a smartphone or computer, that is intended to improve a mental health outcome in the user
Crowdsourced peer support: a form of human support in which members of a "crowd" (i.e., a large group of people who share a user's lived experience) provide support through structured and focused interactions.