



Acceptability and Outcomes of Transdiagnostic Guided Self-help Bibliotherapy for Internalizing Disorder Symptoms in Adults: A Fully Remote Nationwide Open Trial

Lorenzo Lorenzo-Luaces¹ · Jacqueline Howard¹ · Robinson De Jesús-Romero¹ · Allison Peipert¹ · John F. Buss¹ · Colton Lind¹ · Cassandra Botts¹ · Isabella Starvaggi¹

Accepted: 17 October 2022 / Published online: 12 December 2022

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022

Abstract

Introduction Doing What Matters in Times of Stress (DWM) is a five-module transdiagnostic guided self-help (GSH) intervention developed by the World Health Organization, originally in a group-based format. In a sample of individuals recruited from across the United States, we conducted an open trial to study the feasibility and acceptability of an adaptation of DWM in which guidance was provided individually and remotely via phone and videoconferencing.

Methods We assessed internalizing symptoms, psychological well-being, work and social functioning, usability of the intervention, and emotion regulation over the course of 6 weeks.

Results A total of 263 individuals completed our screening. Of those, 75.29% ($n = 198$) qualified for the intervention. We reached most participants who qualified (71.21%, $n = 141$) via phone to schedule a GSH session. Most of those scheduled attended a study session (84.4%, $n = 119$), and most of those who attended a session completed more than half the treatment (84.03%, $n = 100$). Retention rates were comparable to meta-analytic estimates of dropout rates in GSH. Participants showed improvement on internalizing symptoms, psychological well-being, work and social functioning, usability of the intervention, and emotion regulation.

Conclusion DWM is a freely available, seemingly efficacious transdiagnostic intervention for internalizing disorder symptoms.

Keywords Transdiagnostic CBT · Internalizing symptoms · Depression · Anxiety · Self-help

Internalizing disorder symptoms such as depression and anxiety are among the leading causes of disability worldwide (Kotov et al., 2017; Whiteford et al., 2013). Most individuals with internalizing disorder symptoms prefer, and are more likely to adhere to (Swift et al., 2017), treatment with psychological interventions than with medications (Löwe et al., 2006; McHugh et al., 2013). The most well-researched psychological interventions available for internalizing disorders are cognitive-behavioral therapies (CBTs; Lorenzo-Luaces, 2018; Lorenzo-Luaces, et al. 2021a; Cuijpers et al., 2019). CBTs are a family of interventions that aim to change behavior, cognition, and meta-cognition via the use of cognitive,

behavioral, and more recently acceptance and mindfulness-based interventions (DeRubeis & Lorenzo-Luaces, 2017; Hofmann & Hayes, 2019). Despite the existence of CBTs and other effective interventions, the burden of internalizing disorders has not decreased in the past four decades and most individuals still do not receive treatment (Jorm et al., 2017; Kazdin & Blase, 2011).

To increase the dissemination of psychological interventions, researchers have developed self-help approaches to the treatment of internalizing disorder symptoms. Self-help provides much of the same content as do individual psychological interventions (e.g., psychoeducation about stress, coping skills), but the content is delivered via websites or internet/phone applications (Lorenzo-Luaces et al., 2018a, b; Wasil et al., 2021b), books (i.e., bibliotherapy; De Jesús-Romero et al., 2022), other formats such as group lessons (Dolan et al., 2021), or some combination of these. In unguided self-help, an individual uses the self-help material

✉ Lorenzo Lorenzo-Luaces
lolorenz@indiana.edu

¹ Indiana University-Bloomington, Bloomington, IN 47405, USA

by themselves. In guided self-help (GSH), an individual uses self-help material with a professional or paraprofessional promoting adherence to the material along with emotional support. GSH-CBT appears to be roughly comparable to face-to-face CBT in its efficacy and both GSH-CBT, and individual face-to-face CBT is more effective than unguided self-help (Cuijpers et al., 2019). Given that GSH-CBT does not require the presence of a highly trained professional, it may be more scalable than face-to-face psychotherapy and therefore may have greater potential to reduce the public health burden of psychopathology (Kazdin & Blase, 2011). Despite the promise of GSH-CBT approaches, there are numerous barriers to their widespread adoption. Many popular books and apps have little to no research supporting their efficacy (Wasil et al., 2020). Those that are empirically supported are usually not available to the general public or to most clinicians (Wasil et al., 2021b) and clinicians do not frequently use GSH approaches (Peipert et al., 2022a, b). Even those interventions that both have been studied and are freely available are often not very accessible (e.g., they are written at a very high reading level; Martinez et al., 2008).

Beyond its scalability, the acceptability of an intervention may be a factor determining its successful dissemination (Michie et al., 2011). Wolf (1978) was one of the first to define acceptability, although they used the terms “social validity” and “social importance,” referring to the extent that an intervention introduces changes consistent with a client or society’s goals, deemed appropriate to use, and satisfactory to the client. Although GSH-CBT appears to have roughly equal outcomes to face-to-face individual CBT, it tends to be less acceptable to patients as evidenced by higher rates of dropout after treatment initiation (Cuijpers et al., 2019). Researchers have tried a number of strategies to increase the acceptability of GSH-CBT while maintaining its scalability, including changing the modality of the guidance (e.g., using text messages or e-mail instead of video calls), adjusting the dose of guidance, using eCoaches with varying qualifications (e.g., paraprofessionals versus experienced psychotherapists), and comparing synchronous versus asynchronous communication modes (Baumeister et al., 2014). By and large, the literature supports the importance of guidance in self-help CBT for achieving better outcomes and engagement, but does not provide strong support for other ways of increasing acceptability (Bur et al., 2022; Furukawa et al., 2021).

Doing What Matters in Times of Stress (DWM), previously called Self-Help Plus, is a GSH-CBT developed by the World Health Organization (WHO) based on principles of acceptance and commitment therapy (ACT), a form of CBT (Hofmann & Hayes, 2019). Originally, DWM consisted of five group-based psychoeducation sessions delivered through pre-recorded audio materials with a corresponding print guide. Two small trials supported the efficacy and feasibility

of DWM (Tol et al., 2018a, b). In a subsequent cluster randomized controlled trial (RCT) of Sudanese refugees in Uganda (Tol et al., 2020), 694 women from 14 villages were randomized to DWM or enhanced care as usual care (eCAU), which consisted of brief psychoeducation plus the provision of referrals. After 6 weeks, individuals in DWM reported lower rates of internalizing distress [standardized mean difference (*SMD*) = 0.72] and higher levels of psychological well-being (*SMD* = 0.51). In a preventive RCT with refugees and asylum seekers (*N* = 459), DWM appeared effective in reducing mental health symptoms among individuals with no mental health diagnosis, but did not predict a lower incidence of mental health diagnoses at a long-term follow-up (Purgato et al., 2021). A more recent RCT (*N* = 642) supported the preventive effects of DWM versus eCAU for Syrian refugees in Turkey (Acarturk et al., 2022). Taken together, these data support the efficacy of DWM as a group-based mindfulness and acceptance-focused GSH-CBT.

Because DWM is made freely available by the WHO (<https://www.who.int/publications/i/item/9789240003927>) and was developed at an accessible reading level (approximately fourth grade), it has the potential to be a widely scalable intervention. Following the COVID-19 pandemic, several translations of the DWM booklet were made available online. Given the logistic difficulties associated with group-based GSH-CBT (e.g., COVID-19 risk associated with gathering), the WHO adapted the DWM intervention materials such that they could be delivered remotely via telephone or videoconferencing. We conducted an open trial aiming to test the feasibility of conducting GSH-CBT using DWM with individual online coaching, instead of group-based meetings, for individuals across the United States. To our knowledge, this is the first published study of DWM implementation in the US. We assessed the acceptability of the intervention as evidenced by engagement with the guidance in GSH-CBT, self-reported usability, and self-reported satisfaction. In addition to tracking feasibility and acceptability of the intervention, we assessed internalizing symptoms, well-being, and purported emotion regulation mechanisms of change: cognitive reappraisal and expressive suppression (Aldao & Nolen-Hoeksema, 2010). We also measured quality of life and work and social functioning, which may be less responsive to psychosocial intervention than internalizing symptoms (Lorenzo-Luaces & Amsterdam, 2018; Peipert et al., 2022a, b).

Methods

This study was approved by Indiana University’s Institutional Review Board. Participant recruitment began October 17, 2020, and ended February 21, 2022. All recruitment was conducted online and all participants provided informed

consent for the study. The study was registered on ClinicalTrials.gov on July 1, 2021 (NCT04870099).

Participants

Participants were adults (ages 18+) living in the United States who were recruited from social media platforms, primarily Facebook, as well as Instagram and Twitter. We promoted the study by writing a post in our lab's social media accounts directed to individuals "struggling with stress, depression, or anxiety." The post contained a survey link and a flyer describing major components of the study (e.g., the compensation rate). Given the limited reach of our lab's webpage, we used the paid features of the above social media platforms to promote the post. We sought to recruit participants using very minimal entry criteria. The only criterion for entry was the presence of at least mild psychological distress, as evidenced by a score ≥ 6 on the six-item Kessler Psychological Distress Scale (K6; see "K6" below). The only criterion for exclusion was the presence of recurrent death ideation/suicidality, which we initially operationalized as a score ≥ 1 (i.e., "several days") on item nine of the Patient Health Questionnaire-9 (PHQ-9; "thoughts that you would be better off dead or of hurting yourself"). We subsequently modified the suicidality criteria to exclude participants who scored ≥ 2 (i.e., "more than half the days"), thus allowing a PHQ-9 item nine score of one. We made this change given that our early data suggested that most individuals who were excluded from the study based on the original criterion only had infrequent death/suicidal ideation. Participants who were excluded on the basis of suicidality, regardless of how it was operationalized, were given resources for emergencies as well as for finding outpatient treatment providers.

Valid Respondents

Given that we conducted this study over the internet and across the United States, there was a possibility of fraudulent responses, including automated bots. To identify fraudulent respondents, we operationalized "bot-like" or fraudulent behavior as (a) responding that appeared implausibly fast (e.g., completing the roughly 100-question intake in 5 min or less), (b) filling out the study screening at a time in which we were not actively promoting the study, (c) providing duplicate contact information across entries (e.g., the same home address or e-mail), (d) providing seemingly fake contact information (e.g., the name "Deez Nutz"), or (e) other behavior that may be indicative of fraudulent responding (e.g., answers to questions that were actually "hidden" from survey participants), including inconsistent responding to survey items. Out of 486 "clicks" on our screening survey link, we deemed that 12.55% ($n = 61$) were fraudulent respondents. Of the purportedly human

respondents ($n = 425$), 162 individuals did not complete the screening survey, for a total of 263 individuals fully screened for trial entry (see Fig. 1).

Intervention

DWM is a five-chapter booklet that discusses principles of ACT and CBT, namely mindfulness (here called "grounding"), cognitive distancing ("unhooking"), value-based behavioral activation ("acting on your values"), gratitude ("being kind"), and acceptance ("making room"). Participants were given the option of reading a digital version of the DWM booklet, though most (88.89%) opted to receive the printed booklet by mail. Eligible participants were assigned an "eCoach" and contacted within a week of their qualification for the study.

eCoaches

The eCoaches were undergraduate/post-baccalaureate (JH, CL, and KB) and graduate students (RDJR, AP, and JB) at the beginning of the study. All were young adults and all were completing degrees in psychology. None of them had experience with ACT, though two had completed an introductory practicum in CBT prior to the start of the study (AP and RDJR). The undergraduate/post-baccalaureate eCoaches completed a four-hour training covering common mental disorders, principles of CBT, and technology-assisted GSH-CBT. All eCoaches (i.e., graduate and undergraduate students) completed a four-hour training over the course of four weeks that included reading the DWM booklet, completing the WHO's EQUIP training (EQUIP), reading a chapter on "basic helping skills" from the WHO's Problem Management Plus (PM+) guide, and reading a DWM-specific training manual provided by the WHO. Part of the training involved a safety plan for dealing with emergent suicidality. The safety plan involved administering the Columbia Suicide Severity Rating Scale (CSSRS; Posner et al., 2008) and giving crisis referrals for individuals who had acute suicidal ideation. The eCoaches also had to "pass" a roleplay with the principal investigator (LL-L).

Beyond passing the role-play with the PI, competence was not assessed systematically nor was adherence. For all interactions with participants, eCoaches were given semi-structured scripts. There were no statistically significant differences between the eCoaches in rates of study retention or changes any study outcome ($ps > 0.09$; see "Appendix").

Guidance

The first communication for the study was an onboarding or "welcome" call in which participants were provided

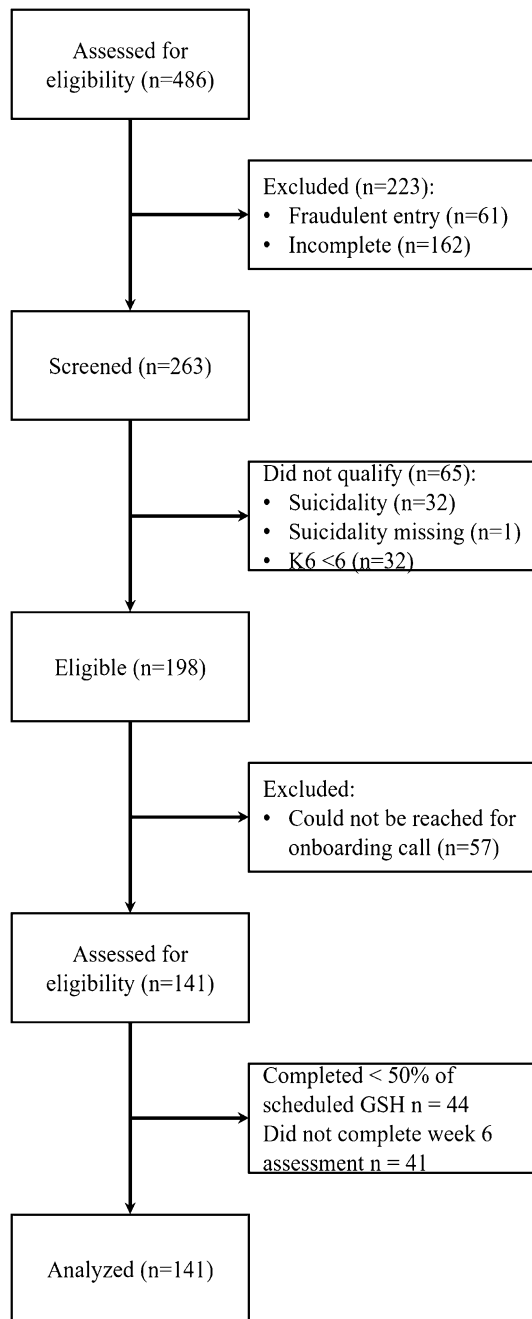


Fig. 1 Trial progression for individuals on a trial of transdiagnostic guided self-help for internalizing distress

information about the study, given a chance to ask questions, and asked to schedule three to six individual weekly sessions over a 6-week period. We let participants schedule fewer than six sessions aiming to maximize engagement for participants who thought six sessions would be too burdensome. During the welcome call, eCoaches helped participants create a plan to use the guide, primarily via a short exercise of mental contrasting with implementation intentions. As part of the exercise, eCoaches helped participants

identify how they wanted to use the guide, predict effects of using the guide, identify potential obstacles to adhering to the guide, and brainstorm potential solutions for the predicted obstacles. Subsequent meetings also followed a semi-structured script.

The guidance meetings began by asking participants to confirm they are still available to meet as well as to confirm their understanding of study confidentiality. The eCoaches ensured that study measures had been filled out and used a principle of measurement-based care: reviewing changes in internalizing symptoms, well-being, and emotion regulation with participants. Next, eCoaches assessed adherence to the reading, which was typically scheduled as one chapter per week. If the participant was able to use the guide with no or only a few challenges, they were praised for their efforts, asked about whether they practiced the specific exercises described in the chapter, and asked to comment on what they noticed about practicing the exercises. If the user was not able to use the guide, the eCoaches were instructed to show empathy and understanding and to review the participant's plan to use the book, including whether it adequately addresses challenges the participant encountered. Regardless of whether participants were able to read the book or not, if they were still interested in participating, the eCoach confirmed the next appointment. The PI provided weekly, as well as ad-hoc, supervision. Supervision focused on using concrete behavioral strategies (e.g., problem-solving) for promoting adherence to the intervention and dealing with resistance by using principles of motivational interviewing.

Outcome Measures

The main outcome measures we used to assess efficacy, potential mechanisms, acceptability, and feasibility are described below. All survey questionnaires were administered online using RedCap. For each survey questionnaire, we present a measure of internal consistency omega (ω), which is interpreted in the same manner as Cronbach's α but may have more desirable psychometric properties (Flora, 2020).

Acceptability and Feasibility

We report the number of participants who (1) completed our baseline assessment, (2) qualified for the study, (3) we were able to reach for an onboarding call, (4) completed one session of GSH-CBT, (5) completed more than 50% of the scheduled assessments, and (6) completed the post-treatment (week 6) assessments. In addition to these metrics, we administered the Systems Usability Scale (SUS; Brooke et al., 1996), a brief measure for assessing the usability of given systems (e.g., websites). The SUS has ten items and is rated on a five-point scale with responses ranging from 0

(“strongly agree”) to 4 (“strongly disagree”). Scale scores are multiplied by 2.5 to produce scores ranging from 0 to 100. Prior work supports the reliability and validity of the SUS (Lewis, 2018). We administered the SUS at week 6, but we also administered a slightly modified version of the SUS at baseline to control for between-individual differences (e.g., the tendency to give high ratings). Participants were asked to rate the usability of the book at baseline “based on the small amount of information [they] have on the study.” In the current study, the baseline scores on the SUS appeared internally consistent ($\omega = 0.85$, 95% CI 0.81–0.89).

K6

The K6 (Kessler et al., 2002) is a six-item self-report scale that measures internalizing distress (e.g., nervousness, depression). Its items are rated on a scale of 0 (“none of the time”) to 4 (“all of the time”), producing scores ranging from 0 to 24 where higher scores indicate greater psychological distress. Previous studies support the reliability and validity of the K6 (Batterham et al., 2018; Staples et al., 2019) and it has previously been used as an outcome measure in GSH-CBT studies (Lorenzo-Luaces et al., 2018a, b). A score of 13 indicates “severe” symptoms of internalizing distress. In the current study, the baseline scores on the K6 appeared internally consistent ($\omega = 0.7$, 95% CI 0.62–0.79). The K6 was administered at baseline and every week thereafter, up to the week 6 post-treatment assessment.

WHO Well-Being Index (WHO-5)

The WHO-5 (Topp et al., 2015) is a five-item self-report scale that measures subjective well-being, an aspect of positive mental health. Its items are rated on a scale of 0 (“at no time”) to 5 (“all of the time”). The raw total scores (0 to 25) are multiplied by 4, producing final scores ranging from 0 to 100, where higher scores indicate greater well-being. Prior work supports the reliability and validity of the WHO-5 (Topp et al., 2015), and it has previously been studied as an outcome measure in GSH-CBT studies (Tol et al., 2020). A score of 50 is considered a useful cutoff for screening for major depression. In the current study, the baseline scores on the WHO-5 appeared internally consistent ($\omega = 0.79$, 95% CI 0.73–0.86). The WHO-5 was administered at baseline and every week thereafter, up to the week 6 post-treatment assessment.

Work and Social Adjustment Scale (WSAS)

The WSAS (Mundt et al., 2002) is a five-item self-report measure that assesses impairment in work, relationships, household, and leisure activities as a result of a specific problem; in our study we queried the effect of “stress” on functioning

(e.g., “[b]ecause of my *stress* my ability to form and maintain close relationships with others, including those I live with, is impaired”). Each item is rated on a nine-point Likert ranging from zero (“not at all”) to eight (“very severely”), producing scores ranging from zero to 40 where higher scores indicate greater impairment. Scores over 10 and over 20 are considered to indicate moderate and severe impairment, respectively. Prior work supports the reliability and validity of the WSAS across various patient populations (Zahra et al., 2014). In the current study, the baseline scores on the WSAS appeared internally consistent ($\omega = 0.82$, 95% CI 0.77–0.87).

Emotion Regulation Scale (ERQ)

The ERQ (Gross & John, 2003) is a ten-item self-report measure of individual differences in the use of two emotion regulation strategies: cognitive reappraisal (ERQ-reappraisal; items 1, 3, 5, 6, 8, and 10) and expressive suppression (ERQ-suppression; items 2, 4, 6, and 9). Prior work supports the reliability and validity of the ERQ in community samples (Preece et al., 2019, 2021). The ERQ items are rated on a seven-point Likert scale with responses ranging from 1 (“strongly disagree”) to 7 (“strongly agree”). We averaged item scores to produce final scores on the same metric of the original items (i.e., 1 to 7) in order to make the ERQ-reappraisal and ERQ-suppression subscales (with differing numbers of items) comparable. In the current study, the baseline scores on the ERQ-reappraisal ($\omega = 0.83$, 95% CI 0.77–0.88) and ERQ-suppression appeared internally consistent ($\omega = 0.87$, 95% CI 0.83–0.91). The ERQ was administered at baseline and every week thereafter, up to the week 6 post-treatment assessment.

Other Assessments

We also administered (1) the PROMIS Depression Short Form (also known as the Cross-Cutting DSM Severity Measure for Depression; American Psychiatric Association, 2013), an eight-item self-report measure of depression symptoms that produces scores ranging from 8 to 40, and (2) the Alcohol Use Disorder Identification Test (AUDIT; Allen et al., 1997), a self-report screening scale for problematic drinking that produces scores ranging from 0 to 40, and (3) a single-item measure of self-rated health (Jylhä, 2009) assessed on a five-point scale. Prior work supports the reliability and validity of both the PROMIS (Cella et al., 2010, 2019; Pilkonis et al., 2014) and the AUDIT (Meneses-Gaya et al., 2009). We included these measures to provide a better characterization of the sample, for example, regarding the specific types of internalizing distress experienced (i.e., depression versus other), comorbid externalizing symptoms (alcohol being the most common type of externalizing disorder), as well as physical health and somatoform symptoms

(Brissette et al., 2003). An additional rationale for including these assessments is that these variables have been found to predict treatment outcomes (Kessler et al., 2017).

Sample Size Justification and Power

We initially powered the trial to be able to detect a statistically significant difference in engagement relative to the 65.1% engagement rate reported in Van Ballegeooijen et al.'s (2014) meta-analysis. We used an online (<https://sample-size.net/sample-size-conf-interval-proportion/>) calculator (Kohn & Senyak, 2021) to estimate the sample size required to test for a statistically significant difference between a sample proportion and the expected value (i.e., 65.1%) at a p value $< .05$. This analysis suggested we needed to recruit at least 95 individuals to have an adequately powered acceptability trial. We ultimately recruited more participants than was necessary ($n = 141$) partially due to uncertainty about our ability to retain all individuals. In addition, this was possible because the PI (LL-L) obtained additional funding for an extension assessing the effects of GSH-CBT on natural language metrics of social media data. That subcomponent of the study (i.e., whether GSH-CBT produces effects detectable via social media) is not being used in the current analysis. We used G*Power 3.1 (Faul et al., 2007) to estimate what magnitude of within-person changes we had adequate powered to detect in our 141-participant sample. The results of this analysis suggested that we could detect small-to-medium within-person changes ($d = 0.25$, at $p < .05$ and power of 80%) with the achieved sample size (i.e., $n = 141$).

Analytic Strategy

All analyses were conducted in R version 4.1.2 (see R Core Team, 2013) using the R Studio Graphic User Interface. All code and deidentified data are available on the Open Science Foundation (OSF) site (<https://osf.io/j32uw/>). First, we report the percentage of participants who progressed from beginning the survey to the end of the study. We followed an intent-to-treat (ITT) approach, analyzing data from all individuals who reached the onboarding call to confirm trial participation.

Multiverse Analysis of Engagement

There are at least two other definitions of trial entry possible in online trials like ours. One is less conservative, considering to be true study subjects only those individuals who completed at least one post-baseline assessment after the baseline eligibility survey. The other is more conservative,

considering to be true study participants all individuals who qualified for the study regardless of whether or not they were reachable following the baseline eligibility survey. We also analyzed two different definitions of trial completion: completing 100% of the scheduled GSH-CBT sessions, and completing at least 50% of the scheduled GSH-CBT sessions. Given variability in how study entry and completion could be defined, we conducted a “multiverse” analysis to assess the acceptability of the intervention (see Steegen et al., 2016). In a multiverse analysis, a researcher conducts and presents all possible ways of analyzing data (i.e., here, the engagement rate). Multiverse analysis has been recommended as a way of increasing transparency in psychological sciences by presenting readers multiple versions of the data, as opposed to simply presenting the analyses that give the result with the most favorable effects. We chose a non-inferiority margin of 10% to determine if observed engagement rates were equivalent to the rates from the meta-analysis.

Demographics and Change Over Time

We report baseline demographic and clinical characteristics by presenting means and standard deviations for continuous variables and percentages for categorical variables (see Table 1). To assess the preliminary efficacy of the intervention, we conducted mixed regression models using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages in R to regress internalizing symptoms (K6), well-being (WHO-5), ERQ-reappraisal, and ERQ-suppression on time in GSH-CBT. We coded the time variable by dividing each week by six, the total number of weeks in the study. This creates a variable ranging from 0 to 1 (e.g., week 1 is 0.17, week 2 is 0.33, week 3 is 0.50, etc.) where a one-point increase in “time” represents change from baseline to the end of treatment. We calculated effect sizes for the results of these mixed models by using the framework proposed by Feingold (2009) wherein estimated change over the course of the intervention is divided by the baseline standard deviation of the measure. Additionally, we assessed pre–post change in the perceived usability of DWM (i.e., the SUS score) as well as in psychosocial functioning (i.e., WSAS) by conducting a paired-sample t -test comparing the scores at baseline with the scores at week 6. We calculated the effect size of pre–post change in the SUS and WSAS by dividing change in these measures by the standard deviation of their change scores. We calculated 95% confidence intervals (CIs) for both of these d -type effect sizes by using the “cohen.d.ci” function in the psych package in R (Revelle, 2016).

In our analyses of the SUS and WSAS, we used a last-observation carried forward (LOCF) imputation approach to deal with missing session-six scores. In our analyses of the K6, WHO-5, and ERQ, we used hierarchical linear modeling (HLM) to handle missing data (Hox, 2000). We refer to these analyses as the “LOCF-imputed” data.

Table 1 Sociodemographic and clinical features of 141 individuals in a trial of transdiagnostic guided self-help

Characteristic	N	N = 141
Age (years)	138	40.36 (13.95)
Age onset distress	141	16.47 (9.54)
Distress (K6: 0–24)	141	10.99 (3.59)
Well-being (WHO5: 0–100)	139	29.55 (14.65)
Functioning (WSAS: 0–40)	141	22.48 (8.77)
Usability (SUS: 0–100)	135	71.94 (13.59)
Reappraisal (ERQ: 1–7)	136	4.21 (1.14)
Suppression (ERQ: 1–7)	138	3.71 (1.33)
Predicted satisfaction	139	6.88 (2.03)
Depression (PROMIS: 8–40)	136	23.97 (5.83)
Alcohol use (AUDIT: 0–40)	139	2.91 (3.69)
Health (1–5)	140	
Excellent		6 (4.3%)
Good		56 (40%)
Average		58 (41%)
Poor/terrible		20 (14%)
Gender identity	141	
Male/TGNC		21 (15%)
Female		120 (85%)
Race/ethnicity	141	
Non-Hispanic White		110 (78%)
Non-Hispanic Black		5 (3.5%)
Hispanic		11 (7.8%)
Asian		8 (5.7%)
Other/multiracial		7 (5.0%)
Married or dating	141	
Single		52 (37%)
Married or dating		89 (63%)
Unemployed	141	
Employed		101 (72%)
Unemployed		40 (28%)
Antidepressant use	141	
Current or past		100 (71%)
Never		41 (29%)
Educational attainment	141	
HS diploma or lower		5 (3.5%)
Some college		30 (21%)
Associate's degree		17 (12%)
Bachelor's degree		49 (35%)
Master's degree		31 (22%)
Postgraduate		9 (6.4%)
Income	140	
< \$15,000		18 (13%)
\$15,000–\$24,999		8 (5.7%)
\$25,000–\$34,999		15 (11%)
\$35,000–\$49,999		11 (7.9%)
\$50,000–\$74,999		31 (22%)
\$75,000–\$99,999		26 (19%)
\$100,000–\$149,999		23 (16%)
\$150,000–\$199,999		8 (5.7%)

¹Mean (SD); *n* (%)

K6 Kessler 6 Scale for Psychological Distress, WHO-5 WHO Well-

Table 1 (continued)

being Index 5, ERQ Emotion Regulation Questionnaire, WSAS Work and Social Adjustment Scale, SUS Systems Usability Scale, TGNC transgender or gender non-conforming, HS high school

Missing Data Imputation

Missing data in demographic covariates was minimal, with most individuals providing complete information for most variables. A maximum of six individuals did not fully answer the SUS (4.3%). There was more longitudinal data missing for the K6, WHO-5, and ERQ related to dropout or missing assessments, up to a maximum of 33.3%, over the course of the intervention. Additionally, several WSAS and SUS ratings were also missing, as not all participants completed the week 6 assessment. To address missing data, including missing outcome data, we imputed all missing data using a machine learning algorithm: non-parametric missing value imputation using random forests, with the R package *missForest* (Stekhoven & Bühlmann, 2012). To preserve the association between the variables (Ginkel et al., 2020), we did not pre-process variables and only minimally recoded them (see OSF). Imputation models that use multiple variables are preferred to LOCF-type imputation models like the one we describe above (Kenward & Molenberghs, 2009; Lachin, 2016). The variables in the imputation model included baseline to week 6 data on the K6, WHO-5, ERQ, WSAS, SUS, PROMIS depression severity, and baseline demographic and clinical data including age, gender, race, marital status, unemployment status, education, income, self-rated health, antidepressant history, and the age at which participants first struggled with internalizing disorder symptoms. We refer to these data as the RF-imputation data. For all of the analyses described above, we present the random forest imputation (RF-imputation) results and the LOCF-imputed results.

Results

Feasibility

A total of 425 individuals started our screening survey, of whom 263 completed it (61.88%). Of these, 198 met our entry criteria. A large number ($n = 32$, 49.23%) of those who did not meet the entry criteria had symptoms that were too mild (i.e., $K6 < 6$). A similarly large number did not meet the criteria because of the suicidality exclusion ($n = 32$, 49.23%). One participant was excluded for not responding to the item probing suicidality. Of the 198 participants who qualified for the study, we were able to reach 141 individuals for an onboarding call (71.21%) to confirm participation. Of those,

119 (84.4%) initiated GSH. Of participants who we reached for an onboarding call, 100 (70.92%) completed at least half of scheduled sessions. 97 (68.79%) Individuals attended the post-treatment assessment. We conducted the “multiverse” analysis calculating the retention rates from all possible definitions of those who entered the study (i.e., qualified, $n = 198$; reached for onboarding call, $n = 141$; initiated GSH, $n = 119$) as well as the two definitions of those who completed the study (i.e., who attended the post-treatment assessment $n = 97$, or who completed at least half of agreed-upon sessions $n = 100$). Overall, four of the six engagement rates we calculated were non-inferior to the 65.1% engagement rate from the meta-analysis by Van Ballegooijen et al. (2014) and one was inconclusive (see Fig. 2).

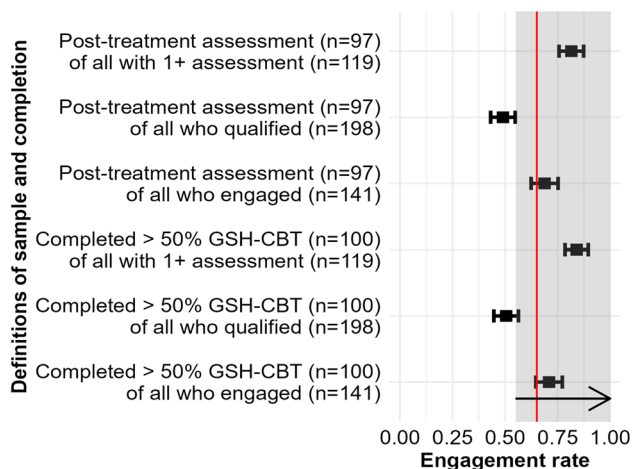


Fig. 2 Multiverse analysis of engagement rates for different definitions of trial entry and trial completion for participants undergoing transdiagnostic guided self-help CBT

Sample Demographics

Most participants in the sample identified as cisgender women. Although different income categories were relatively well represented, the sample was somewhat more educated than one would expect of a representative US sample (see Table 1). On average, participants reported struggling with internalizing disorder symptoms very early in life ($M = 16.47, SD = 9.54$). Although our internalizing symptom entry criterion was relatively low ($K6 \geq 6$), the average participant reported internalizing distress scores substantially higher than the cutoff ($K6: M = 10.99, SD = 3.59$), as well as relatively low well-being ($WHO-5: M = 29.55, SD = 14.65$).

Attesting to the clinical severity of our sample, 131 (92.91%) of individuals met the WSAS cutoff for moderate impairment, 125 (88.65%) individuals met the WHO-5 cutoff for depression screening, 87 (61.7%) met the PROMIS depression cutoff for moderate depression, 88 (62.41%) met the WSAS cutoff for severe impairment, and 40 (28.37%) met the K6 cutoff for severe distress.

Acceptability

After reading the consent, participants’ expectations of the usability of DWM were relatively high (RF-imputed: $M = 72.18, SD = 13.41$; LOCF-imputed: $M = 71.94, SD = 13.59$). About two thirds of individuals reported scores of 68 or higher (RF-imputed: 62.41%, LOCF-imputed: 60.74%). After 6 weeks of DWM, acceptability (RF-imputed: $M = 84.1, SD = 8.599$; LOCF-imputed: $M = 84.43, SD = 10.17$) increased substantially [RF-imputation: $M_{diff} = 11.89, t(140) = 11.16, p < .001$; LOCF-imputed: $M_{diff} = 8.259, t(134) = 8.088, p < .001$; see Table 2]. Of those queried at week 6, almost all reported the intervention was usable (i.e., $SUS \geq 68$; RF-imputed: 94.33%, LOCF-imputed: 91.75%).

Table 2 Standardized mean differences (SMDs) of within-person changes for 141 participants with internalizing distress treated with transdiagnostic guided self-help CBT

Outcome	LCI (RF)	SMD (RF)	UCI (RF)	LCI (LOCF)	SMD (LOCF)	UCI (LOCF)
Internalizing (K6)	-1.42	-1.21	-0.99	-1.52	-1.30	-1.08
Well-being (WHO-5)	0.68	0.88	1.07	0.74	0.94	1.14
Reappraisal (ERQ)	0.54	0.73	0.92	0.60	0.79	0.98
Suppression (ERQ)	-0.50	-0.33	-0.16	-0.55	-0.38	-0.21
Functioning (WSAS)	-1.18	-0.98	-0.78	-0.88	-0.70	-0.51
Usability (SUS)	0.74	0.93	1.13	0.51	0.70	0.88
Satisfaction	0.74	0.94	1.14	0.54	0.72	0.91

LCI lower 95% confidence interval, UCI upper 95% confidence interval, RF random forest imputation, LOCF last observation carried forward imputation, K6 Kessler 6 Scale for Psychological Distress, WHO-5 WHO Well-being Index 5, ERQ Emotion Regulation Questionnaire, WSAS Work and Social Adjustment Scale, SUS Systems Usability Scale

Symptom Reduction

Over the course of GSH-CBT, there were large reductions in internalizing disorder symptoms [RF-imputation: $B = -4.37$, $SE = 0.29$, $t(140) = -14.96$, $p < .001$; LOCF-imputed: $B = -4.67$, $SE = 0.4$, $t(94.38) = -11.77$, $p < .001$]. Similarly, over the course of GSH-CBT, there were large improvements in well-being [RF-imputation: $B = 12.7$, $SE = 1.44$, $t(140) = 8.8$, $p < .001$; LOCF-imputed: $B = 13.8$, $SE = 1.95$, $t(107.4) = 7.055$, $p < 0.001$]. From baseline to the post-treatment assessment, there were improvements in work and social functioning [RF-imputation: $M_{diff} = -7.87$, $t(140) = -11.91$, $p < 0.001$; LOCF-imputed: $M_{diff} = -5.48$, $t(140) = -8.256$, $p < 0.001$]. These changes were large in magnitude (see Table 2; Fig. 3). There were no reports of treatment-emergent suicidality.

Emotion Regulation

Compared to nationally representative samples (Preece et al., 2021), participants in our study had relatively low levels of self-reported reappraisal use (RF-imputation: $M = 4.2$, $SD = 1.12$; LOCF-imputed: $M = 4.21$, $SD = 1.14$) but average levels of suppression use (RF-imputation: $M = 3.73$, $SD = 1.32$; LOCF-imputed: $M = 3.71$, $SD = 1.33$). Over the course of GSH-CBT, there were large increases in cognitive reappraisal use (RF-imputation: $B = 0.823$, $SE = 0.078$, $t(140) = 10.59$, $p < .001$; LOCF-imputed: $B = 0.891$, $SE = 0.11$, $t(93.11) = 8.399$, $p < .001$). The reductions in expressive suppression were more modest [RF-imputation: $B = -0.437$, $SE = 0.09$, $t(140) = -4.867$, $p < .001$; LOCF-imputed: $B = -0.513$, $SE = 0.12$, $t(101) = -4.339$, $p < .001$; see Table 2].

Discussion

We conducted a fully remote nationwide clinical trial to assess the feasibility, acceptability, and preliminary efficacy of DWM, a GSH-CBT, in a United States sample. Our results demonstrate that DWM can be delivered in a fully remote fashion, consistent with other studies suggesting that nationwide recruitment of individuals with internalizing symptoms is feasible (Arean et al., 2016). Our multiverse analysis suggested that we were able to retain anywhere from 48 to 79% of individuals. The most conservative retention estimate (48%) was calculated with study entry defined as simply qualifying for the study after completing the baseline survey and study completion defined as completing more than 50% of scheduled GSH-CBT sessions. This conservative approach suggested that we lost half of study participants to dropout, resulting in lower retention rates than we

might expect based on meta-analyses of GSH-CBT. All other definitions of retention rate that we tested suggested yielded comparable with or higher than in prior work (70–79%).

Nonetheless, we observed that with every “step” in the process of engaging individuals in GSH-CBT, there was a sizeable reduction in participants who remained engaged with the intervention. This has also been observed in studies of individual face-to-face psychotherapy (Krendl & Lorenzo-Luaces, 2021). These findings imply that in order to maximize the reach of GSH-CBT, we need to reduce as many of the barriers of its initiation as possible (e.g., shorter screening times, more relaxed entry criteria). While human engagement generally facilitates desired outcomes and promotes adherence to self-help approaches like bibliotherapy or internet-based treatment (Cuijpers et al., 2019), requiring human engagement (e.g., as in reaching the welcome call for the current trial) may actually be a barrier for some individuals. Over the 6-week study period, individuals who engaged in the DWM GSH-CBT reported very large decreases in internalizing distress, large increases in well-being, large improvements in functioning, and improved perceptions of the usability of the intervention. Additionally, we also found that participants experienced large increases in cognitive reappraisal and medium decreases in expressive suppression. These findings continue to underscore the promise of GSH-CBT to address the burden of untreated psychopathology. For example, it may be beneficial to target dissemination of self-help CBT material at individuals currently on waiting lists for more traditional psychological services (Peipert et al., 2022a, b).

The level of attrition we reported underscores the need to optimize engagement in GSH-CBT. Given that human support has been the only replicable predictor of engagement in GSH-CBT (Bur et al., 2022; Cavanagh et al., 2010; Furukawa et al., 2021), it would behoove the field to explore ways of optimizing the human element of GSH-CBT. For example, in face-to-face CBT, greater session frequency (e.g., twice-weekly instead of once-weekly) improves outcomes (Bruijnicks et al., 2020). It is possible that increasing the level of engagement in GSH-CBT may improve outcomes or engagement. In our study, most dropout occurred after individuals qualified for the study when they could not be reached for an onboarding call. The moment in which individuals seek psychological services may be a moment when they are most receptive for such services, and waiting time following that moment (e.g., the wait for the onboarding call) may represent an engagement obstacle (Krendl & Lorenzo-Luaces, 2021). One promising way to increase engagement, then, may be to immediately make very brief interventions, such as single-session interventions (SSIs), immediately available to treatment seekers. SSIs have been widely studied for youth psychopathology (Schleider & Weisz, 2017), and emerging research supports their use

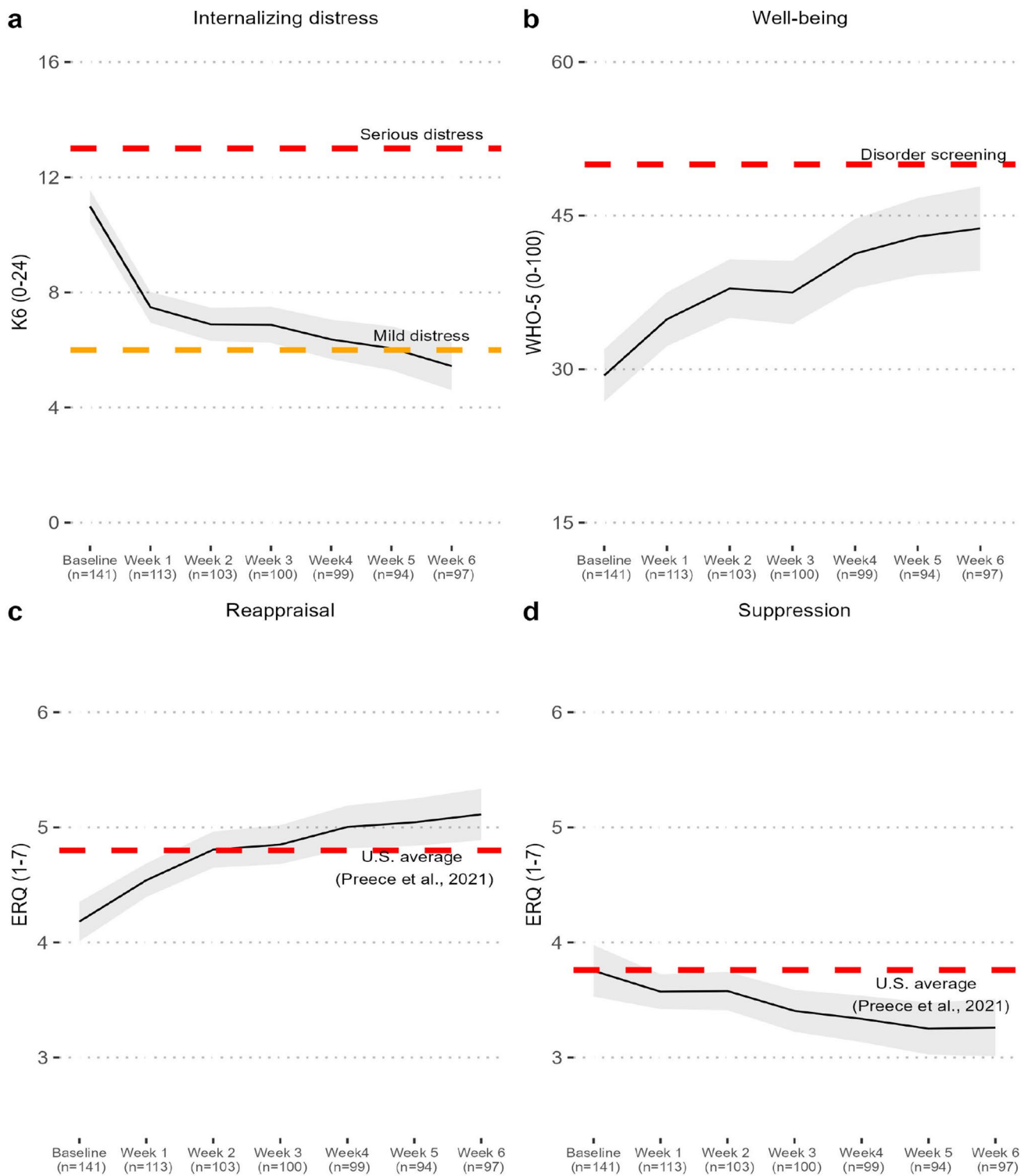


Fig. 3 Changes in **a** internalizing distress, **b** well-being, **c** cognitive reappraisal, and **d** expressive suppression in 141 individuals in GSH-CBT

in adults (Wasil et al., 2021a, b, c). A logical future direction is to study the combination of SSI and other forms of GSH-CBT in a “stepped care” fashion (Lorenzo-Luaces et al., 2017) to investigate whether immediate access to an

SSI increases subsequent engagement with treatment (the GSH-CBT). Given individual differences in engagement and response to these interventions, it will also be worthwhile to explore predictors of engagement and response that could be

used for the purposes of risk stratification (Lorenzo-Luaces et al., 2017, 2020; Lorenzo-Luaces et al., 2021b). Male gender seems to be a predictor of low engagement, as does lower education level and younger age (Karyotaki et al., 2015). Future larger-scale studies should explore a greater number of potential predictors (Kessler et al., 2017). Another option, then, may be to target GSH-CBT efforts at the individuals most likely to adhere to them, allocating treatment resources on the basis of expected engagement.

The major limitation of the current study is its lack of a control group; however, the changes we observed in the K6 and WHO-5 were of the same magnitude as in the large trial by Tol et al. (2020). An RCT is nonetheless warranted, because it is possible that the observed changes in symptoms and emotion regulation are attributable to the natural passage of time, placebo effects, or participant characteristics. Another limitation is that the sample consisted primarily of women. The formative work by the WHO regarding DWM also saw difficulty with recruiting men (Tol et al., 2020). Men are less likely to meet the criteria for internalizing distress than women, and large-scale studies suggest that even when men have internalizing distress, they are substantially less likely to seek treatment when compared to women (Rayner et al., 2021). In addition, non-Hispanic Black, Hispanic, Asian, and other racial-ethnic minority individuals were underrepresented relative to the general US population. Although these individuals were underrepresented in our sample, a meta-analysis by our group (De Jesús-Romero & Lorenzo-Luaces, 2002) suggests that these rates of representation of racial-ethnic diversity in our trial are consistent with other trials of digital interventions. Future work should explore how to increase the engagement of cisgender men and racial-ethnic minoritized individuals in GSH studies. Finally, we measured the acceptability of the intervention by using (1) self-rated usability, (2) self-rated satisfaction, and (3) compliance with the guidance. Thus, a limitation of the study is that we did not measure acceptability of the ACT-based content nor did we systematically address adherence or engagement with the GSH-CBT material (e.g., reading comprehension, percent of time practicing mindfulness).

Several strengths of the study are worth discussing. First, we recruited individuals from all over the US to maximize generalizability. Additionally, we used very relaxed entry criteria to maximize the representativeness of our sample (Lorenzo-Luaces et al., 2018a, b). We measured a variety of outcomes including a transdiagnostic measure of mental health disorder symptoms, two features of emotion regulation, aspects of positive mental health functioning like well-being and functioning, and explicit assessments of participants' perceptions of the usability of the intervention as well as their satisfaction with it. Almost all participants had moderate functional impairment, most met cut-off criteria for disorder screening in the various measures we used to

characterize psychopathology, and about two thirds met the criterion for severe impairment on the WSAS. These results suggest we were able to recruit individuals with relatively severe clinical profiles.

Interestingly, during the intervention, participants reported more change in cognitive reappraisal than in expressive suppression. Cognitive reappraisal is a basic emotion regulation strategy that is the equivalent of cognitive restructuring in CBT (Lorenzo-Luaces et al., 2015, 2016). In most forms of CBT, restructuring is a recommended way of reducing cognitive distortions to improve mood but is not formally addressed by DWM or ACT-based interventions except via distancing or defusion exercises. Expressive suppression is analogous to the idea of experiential avoidance in ACT and antithetical to acceptance (Hofmann & Asmundson, 2008). Thus, it was unexpected that suppression changed less than reappraisal. It may be that a more direct measure of ACT-relevant processes would capture change better than the ERQ. However, in the larger DWM trial by the WHO, improvement in an ACT-specific measure in DWM versus treatment as usual was relatively modest ($d=0.42$; Tol et al., 2020) and not maintained over a follow-up period ($d=0.09$). One possibility is that DWM may facilitate symptom change through processes other than those implied by the ACT model. Process-outcome research clarifying the association between changes in emotion regulation and changes in symptoms is needed, including exploration of alternative processes that may explain changes in DWM such as normalization of distress, working alliance, behavioral activation, or other potential mechanisms. The efficacy of GSH-CBT is well established relative to control conditions and, to some extent, relative to individual therapy (Cuijpers et al., 2019). The areas of uncertainty remaining in relation to GSH-CBT include its dissemination in real world settings, best staging practices for GSH-CBT relative to other treatments, identifying who engages and responds to these interventions, and isolating mechanisms through which these interventions achieve their effects.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10608-022-10338-5>.

Author Contributions Conceptualization: LL-L, JH; Data curation: LL-L, JH, IS; Formal analysis: LL-L, IS; Funding acquisition: LL-L; Investigation: LL-L, JH, RDJ-R, AP, JB, CL, KB, IS; Methodology: LL-L; Project administration: LL-L, JH; Resources: LL-L; Software: LL-L, IS; Supervision: LL-L; Validation: LL-L; Visualization: LL-L; Writing—original draft: LL-L; Writing—review and editing: LL-L, JH, RDJ-R, AP, JB, CL, KB, IS.

Funding This research was partially funded by the National Institute of Mental Health Grant Number T32 MH103213-06 which provided support for Peipert, De Jesús-Romero, Buss, and Starvaggi and Grant Numbers KL2TR002530 and UL1TR002529 (A. Shekhar, PI) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award which

provided support for Professor Lorenzo-Luaces. The work was also supported the 2020 Global Mental Health Fellowship through the APA and IUPsys.

Declarations

Conflict of Interest Professor Lorenzo-Luaces has received consulting fees from Happify Health, Inc. who had no role in the current research. Jacqueline Howard, Robinson De Jesús-Romero, Allison Peipert, John Buss, Colton Lind, Kassandra Botts Isabella Starvaggi declare that they have no conflict of interest.

Research Involving Human Participants and/or Animals All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional and/or National Research Committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Indiana University IRB.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

- Acarturk, C., Uygun, E., Ilkkursun, Z., Carswell, K., Tedeschi, F., Batu, M., Eskici, S., Kurt, G., Anttila, M., Au, T., et al. (2022). Effectiveness of a WHO self-help psychological intervention for preventing mental disorders among Syrian refugees in Turkey: A randomized controlled trial. *World Psychiatry*, *21*(1), 88–95.
- Aldao, A., & Nolen-Hoeksema, S. (2010). Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy*, *48*(10), 974–983.
- Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research*, *21*(4), 613–619.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Publishing.
- Arean, P. A., Hallgren, K. A., Jordan, J. T., Gazzaley, A., Atkins, D. C., Heagerty, P. J., & Anguera, J. A. (2016). The use and effectiveness of mobile apps for depression: Results from a fully remote clinical trial. *Journal of Medical Internet Research*, *18*(12), e6482.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv Preprint [arXiv:1506.04967](https://arxiv.org/abs/1506.04967)
- Batterham, P., Sunderland, M., Slade, T., CEAR, A., & Carragher, N. (2018). Assessing distress in the community: Psychometric properties and crosswalk comparison of eight measures of psychological distress. *Psychological Medicine*, *48*(8), 1316–1324.
- Baumeister, H., Reichler, L., Munzinger, M., & Lin, J. (2014). The impact of guidance on Internet-based mental health interventions—A systematic review. *Internet Interventions*, *1*(4), 205–215.
- Brissette, I., Leventhal, H., & Leventhal, E. A. (2003). Observer ratings of health and sickness: Can other people tell us anything about our health that we don't already know? *Health Psychology*, *22*(5), 471.
- Brooke, J., et al. (1996). SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.
- Brujiniks, S. J., Lemmens, L. H., Hollon, S. D., Peeters, F. P., Cuijpers, P., Arntz, A., Dingemans, P., Willems, L., Van Oppen, P., Twisk, J. W., et al. (2020). The effects of once- versus twice-weekly sessions on psychotherapy outcomes in depressed patients. *The British Journal of Psychiatry*, *216*(4), 222–230.
- Bur, O. T., Krieger, T., Moritz, S., Klein, J. P., & Berger, T. (2022). Optimizing the context of support of web-based self-help in individuals with mild to moderate depressive symptoms: A randomized full factorial trial. *Behaviour Research and Therapy*, *152*, 104070.
- Cavanagh, K., et al. (2010). Turn on, tune in and (don't) drop out: Engagement, adherence, attrition, and alliance with Internet-based interventions. *Oxford Guide to Low Intensity CBT Interventions*, *227*, 232.
- Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., et al. (2019). PROMIS adult health profiles: Efficient short-form measures of seven health domains. *Value in Health*, *22*(5), 537–544.
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., et al. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, *63*(11), 1179–1194.
- Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: A network meta-analysis. *JAMA Psychiatry*, *76*(7), 700–707. <https://doi.org/10.1001/jamapsychiatry.2019.0268>
- De Jesús-Romero, R., & Lorenzo-Luaces, L. (2022). Representation of racial and ethnic diversity in clinical trials of Internet-based self-help treatments for depression. *Presented at the 2022 Society of Consulting Psychology*.
- De Jesús-Romero, R., Wasil, A., & Lorenzo-Luaces, L. (2022). Willingness to use internet-based versus bibliotherapy interventions in a representative U.S. Sample: Cross-sectional survey study. *JMIR Formative Research*, *6*(8), e39508.
- de Meneses-Gaya, C., Zuardi, A. W., Loureiro, S. R., & Crippa, J. A. S. (2009). Alcohol Use Disorders Identification Test (AUDIT): An updated systematic review of psychometric properties. *Psychology and Neuroscience*, *2*(1), 83.
- DeRubeis, R. J., & Lorenzo-Luaces, L. (2017). Recognizing that truth is unattainable and attending to the most informative research evidence. *Psychotherapy Research*, *27*(1), 33–35.
- Dolan, N., Simmonds-Buckley, M., Kellett, S., Siddell, E., & Delgado, J. (2021). Effectiveness of stress control large group psychoeducation for anxiety and depression: Systematic review and meta-analysis. *British Journal of Clinical Psychology*, *60*(3), 375–399.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.
- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*(1), 43.
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using r to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484–501.
- Furukawa, T. A., Suganuma, A., Ostinelli, E. G., Andersson, G., Beevers, C. G., Shumake, J., Berger, T., Boele, F. W., Buntrock, C., Carlbring, P., et al. (2021). Dismantling, optimising, and personalising Internet cognitive behavioural therapy for depression: A systematic review and component network meta-analysis using individual participant data. *The Lancet Psychiatry*, *8*(6), 500–511.
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348.

- Hofmann, S. G., & Asmundson, G. J. (2008). Acceptance and mindfulness-based therapy: New wave or old hat? *Clinical Psychology Review*, 28(1), 1–16.
- Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science*, 7(1), 37–50.
- Hox, J. J. (2000). Multilevel analysis of grouped and longitudinal data. In *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Routledge.
- Jorm, A. F., Patten, S. B., Brugha, T. S., & Mojtabai, R. (2017). Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry*, 16(1), 90–99.
- Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science and Medicine*, 69(3), 307–316.
- Karyotaki, E., Kleiboer, A., Smit, F., Turner, D. T., Pastor, A. M., Andersson, G., Berger, T., Botella, C., Breton, J., Carlbring, P., et al. (2015). Predictors of treatment dropout in self-guided web-based interventions for depression: An ‘individual patient data’ meta-analysis. *Psychological Medicine*, 45(13), 2717–2726.
- Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science*, 6(1), 21–37.
- Kenward, M. G., & Molenberghs, G. (2009). Last observation carried forward: A crystal ball? *Journal of Biopharmaceutical Statistics*, 19(5), 872–888.
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976.
- Kessler, R. C., Van Loo, H., Wardenaar, K., Bossarte, R., Brenner, L., Ebert, D., de Jonge, P., Nierenberg, A., Rosellini, A., Sampson, N., et al. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, 26(1), 22–36.
- Kohn, M. A., & Senyak, J. (2021, December 20). *Sample size calculators*. UCSF CTSI. Retrieved December 28, 2021, from <https://www.sample-size.net/>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., et al. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454.
- Krendl, A. C., & Lorenzo-Luaces, L. (2021). Identifying peaks in attrition after clients initiate mental health treatment in a university training clinic. *Psychological Services*, 19, 519–526.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26.
- Lachin, J. M. (2016). Fallacies of last observation carried forward analyses. *Clinical Trials*, 13(2), 161–168.
- Lewis, J. R. (2018). The system usability scale: Past, present, and future. *International Journal of Human-Computer Interaction*, 34(7), 577–590.
- Lorenzo-Luaces, L. (2018). The evidence for cognitive behavioral therapy. *JAMA*, 319(8), 831–832.
- Lorenzo-Luaces, L., & Amsterdam, J. D. (2018). Effects of venlafaxine versus lithium monotherapy on quality of life in bipolar II major depressive disorder: Findings from a double-blind randomized controlled trial. *Psychiatry Research*, 259, 455–459.
- Lorenzo-Luaces, L., DeRubeis, R. J., van Straten, A., & Tiemens, B. (2017). A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. *Journal of Affective Disorders*, 213, 78–85. <https://doi.org/10.1016/j.jad.2017.02.010>
- Lorenzo-Luaces, L., German, R. E., & DeRubeis, R. J. (2015). It’s complicated: The relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clinical Psychology Review*, 41, 3–15. <https://doi.org/10.1016/j.cpr.2014.12.003>
- Lorenzo-Luaces, L., Johns, E., & Keefe, J. R. (2018a). The generalizability of randomized controlled trials of self-guided Internet-based cognitive behavioral therapy for depressive symptoms: Systematic review and meta-regression analysis. *Journal of Medical Internet Research*, 20(11), e10113. <https://doi.org/10.2196/10113>
- Lorenzo-Luaces, L., Keefe, J. R., & DeRubeis, R. J. (2016). Cognitive-behavioral therapy: Nature and relation to non-cognitive behavioral therapy. *Behavior Therapy*, 47(6), 785–803.
- Lorenzo-Luaces, L., Lemmens, L. H., Keefe, J. R., Cuijpers, P., & Bockting, C. L. (2021a). The efficacy of cognitive behavioral therapy for emotional disorders. In *Handbook of cognitive behavioral therapy: Overview and approaches* (Vol. 1, pp. 51–89). American Psychological Association.
- Lorenzo-Luaces, L., Peipert, A., Romero De Jesús, R., Rutter, L. A., & Rodríguez-Quintana, N. (2021b). Personalized medicine and cognitive behavioral therapies for depression: Small effects, big problems, and bigger data. *International Journal of Cognitive Therapy*, 14(1), 59–85.
- Lorenzo-Luaces, L., Rodríguez-Quintana, N., Riley, T. N., & Weisz, J. R. (2020). A placebo prognostic index (PI) as a moderator of outcomes in the treatment of adolescent depression: Could it inform risk-stratification in treatment with cognitive-behavioral therapy, fluoxetine, or their combination? *Psychotherapy Research*, 31(1), 5–18. <https://doi.org/10.1080/10503307.2020.1747657>
- Lorenzo-Luaces, L., Zimmerman, M., & Cuijpers, P. (2018b). Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants? *Journal of Affective Disorders*, 234, 8–13.
- Löwe, B., Schulz, U., Gräfe, K., & Wilke, S. (2006). Medical patients’ attitudes toward emotional problems and their treatment. *Journal of General Internal Medicine*, 21(1), 39–45.
- Martinez, R., Whitfield, G., Dafters, R., & Williams, C. (2008). Can people read self-help manuals for depression? A challenge for the stepped care model and book prescription schemes. *Behavioural and Cognitive Psychotherapy*, 36(1), 89–97.
- McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013). Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: A meta-analytic review. *The Journal of Clinical Psychiatry*, 74(6), 13979.
- Michie, S., Van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1), 1–12.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, 180(5), 461–464.
- Peipert, A., Krendl, A. C., & Lorenzo-Luaces, L. (2022a). Waiting lists for psychotherapy and provider attitudes toward low-intensity treatments as potential interventions: Survey study. *JMIR Formative Research*, 6(9), e39787.
- Peipert, A., Rodríguez-Quintana, N., & Lorenzo-Luaces, L. (2022b). Outcomes of student trainee-delivered cognitive behavioral therapy (CBT) on internalizing symptoms, CBT skills, and life satisfaction. *International Journal of Cognitive Therapy*, 15(1), 94–113.
- Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Maihoefer, C. C., & Lawrence, S. M. (2014). Validation of the depression item bank

- from the patient-reported outcomes measurement information system (PROMIS) in a three-month observational study. *Journal of Psychiatric Research*, 56, 112–119.
- Posner, K., Brent, D., Lucas, C., Gould, M., Stanley, B., Brown, G., Fisher, P., Zelazny, J., Burke, A., Oquendo, M., et al. (2008). *Columbia Suicide Severity Rating Scale (C-SSRS)* (p. 10). Columbia University Medical Center.
- Preece, D. A., Becerra, R., Hasking, P., McEvoy, P. M., Boyes, M., Sauer-Zavala, S., Chen, W., & Gross, J. J. (2021). The Emotion Regulation Questionnaire: Psychometric properties and relations with affective symptoms in a United States general community sample. *Journal of Affective Disorders*, 284, 27–30.
- Preece, D. A., Becerra, R., Robinson, K., & Gross, J. J. (2019). The Emotion Regulation Questionnaire: Psychometric properties in general community samples. *Journal of Personality Assessment*, 102(3), 348–356.
- Purgato, M., Carswell, K., Tedeschi, F., Acarturk, C., Anttila, M., Au, T., Bajbouj, M., Baumgartner, J., Biondi, M., Churchill, R., et al. (2021). Effectiveness of self-help plus in preventing mental disorders in refugees and asylum seekers in Western Europe: A multinational randomized controlled trial. *Psychotherapy and Psychosomatics*, 90(6), 403–414.
- Rayner, C., Coleman, J. R., Purves, K. L., Carr, E., Cheesman, R., Davies, M. R., Delgadillo, J., Hübel, C., Krebs, G., Peel, A. J., et al. (2021). Sociodemographic factors associated with treatment-seeking and treatment receipt: Cross-sectional analysis of UK Biobank participants with lifetime generalised anxiety or major depressive disorder. *BJPsych Open*, 7(6), e216.
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Revelle, W. (2016). *How to: Use the psych package for factor analysis and data reduction*. Northwestern University.
- Schleider, J. L., & Weisz, J. R. (2017). Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 56(2), 107–115.
- Staples, L. G., Dear, B. F., Gandy, M., Fogliati, V., Fogliati, R., Karin, E., Nielssen, O., & Titov, N. (2019). Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: The PHQ-2, GAD-2, and K-6. *General Hospital Psychiatry*, 56, 13–18.
- Stegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712.
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.
- Swift, J. K., Greenberg, R. P., Tompkins, K. A., & Parkin, S. R. (2017). Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: A meta-analysis of head-to-head comparisons. *Psychotherapy*, 54(1), 47.
- Tol, W. A., Augustinavicius, J., Carswell, K., Brown, F., Adaku, A., Leku, M., García-Moreno, C., Ventevogel, P., White, R., & Van Ommeren, M. (2018a). Translation, adaptation, and pilot of a guided self-help intervention to reduce psychological distress in South Sudanese refugees in Uganda. *Global Mental Health*, 5, e25.
- Tol, W. A., Augustinavicius, J., Carswell, K., Leku, M. R., Adaku, A., Brown, F. L., García-Moreno, C., Ventevogel, P., White, R. G., Kogan, C. S., et al. (2018b). Feasibility of a guided self-help intervention to reduce psychological distress in South Sudanese refugee women in Uganda. *World Psychiatry*, 17(2), 234.
- Tol, W. A., Leku, M. R., Lakin, D. P., Carswell, K., Augustinavicius, J., Adaku, A., Au, T. M., Brown, F. L., Bryant, R. A., García-Moreno, C., et al. (2020). Guided self-help to reduce psychological distress in South Sudanese female refugees in Uganda: A cluster randomised trial. *The Lancet Global Health*, 8(2), e254–e263.
- Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84(3), 167–176.
- Van Ballegooijen, W., Cuijpers, P., Van Straten, A., Karyotaki, E., Andersson, G., Smit, J. H., & Riper, H. (2014). Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: A meta-analysis. *PLoS ONE*, 9(7), e100674.
- van Ginkel, J. R., Linting, M., Rippe, R. C., & van der Voort, A. (2020). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of Personality Assessment*, 102(3), 297–308.
- Wasil, A. R., Gillespie, S., Patel, R., Petre, A., Venturo-Conerly, K. E., Shingleton, R. M., Weisz, J. R., & DeRubeis, R. J. (2020). Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. *Journal of Consulting and Clinical Psychology*, 88(11), 983–993.
- Wasil, A. R., Gillespie, S., Schell, T., Lorenzo-Luaces, L., & DeRubeis, R. J. (2021a). Estimating the real-world usage of mobile apps for mental health: Development and application of two novel metrics. *World Psychiatry*, 20(1), 137.
- Wasil, A. R., Palermo, E., Lorenzo-Luaces, L., & DeRubeis, R. (2021b). Is there an app for that? An overview of popular mental health and wellness apps. *Cognitive Behavioural Practice*. <https://doi.org/10.31234/osf.io/su4ar>
- Wasil, A. R., Taylor, M. E., Franzen, R. E., Steinberg, J. S., & DeRubeis, R. J. (2021c). Promoting graduate student mental health during COVID-19: Acceptability, feasibility, and perceived utility of an online single-session intervention. *Frontiers in Psychology*, 12, 1167.
- Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., et al. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, 382(9904), 1575–1586.
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11(2), 203–214.
- Zahra, D., Qureshi, A., Henley, W., Taylor, R., Quinn, C., Pooler, J., Hardy, G., Newbold, A., & Byng, R. (2014). The Work and Social Adjustment Scale: Reliability, sensitivity and value. *International Journal of Psychiatry in Clinical Practice*, 18(2), 131–138.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.