

A crowdsourced megastudy of 12 digital single-session interventions for depression in US adults

Received: 28 July 2025

Accepted: 23 January 2026

Published online: 02 March 2026

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Digital, self-guided, single-session interventions (SSIs) deliver structured psychological support within one interaction. Here we crowdsourced 66 diverse 10-min SSIs for depression and, with input from researchers and lived-experience experts, selected 11 for testing in a preregistered online randomized controlled trial (ClinicalTrials.gov ID: [NCT06856668](https://clinicaltrials.gov/ct2/show/study/NCT06856668)). US adults ($N = 7,505$) experiencing elevated depressive symptoms were randomly assigned to 1 of the 11 crowdsourced SSIs, a validated behavioural activation SSI (active comparator) or a control condition without intervention content. Nearly all SSIs improved psychological outcomes immediately after completion ($d \leq 0.37$). However, only two SSIs significantly reduced depression at 4-week follow-up ($d = 0.14$ and 0.15). Unexpectedly, completing an SSI made participants feel less confident and less interested in making changes to overcome depression at 4 weeks, on average ($d = 0.05$). Future work should aim to leverage SSIs' immediate benefits to promote sustained behaviour change or service engagement.

Depression is a major global health concern and a profound burden for the 332 million people it affects each year^{1,2}. Evidence-based psychological interventions such as psychotherapy can empower people with the skills and confidence to prevent and treat their depression³. However, barriers such as cost, lack of providers and stigma make in-person psychological interventions difficult to access for most^{4,5}. Digital mental health interventions are more scalable than in-person options, offering users on-demand support at low or no cost via computers and mobile devices⁶. However, these tools have had limited impact due to challenges in reaching broad populations and engaging users effectively enough to produce lasting change⁷⁻⁹.

Single-session interventions (SSIs) are structured programmes designed to spark meaningful gains in mental health within a single interaction¹⁰. SSIs sidestep key barriers to accessing both in-person and digital mental health interventions, most notably the need for long-term engagement. SSIs can be delivered in various formats: in-person or online, self-guided or human-supported, and targeting outcomes in the SSI user or in others (for example, equipping parents with skills to better support their children)¹¹. Here, we use the term 'SSI' only to refer to SSIs that are digital, self-guided and focused on improving the user's mental health.

SSIs' lightweight nature makes them ideally suited for broad dissemination¹². SSIs have been implemented across diverse populations and settings, including psychotherapy waitlists, social media, schools and war zones¹¹. In many cases, these deployments reached people who were unlikely to have received an evidence-based intervention otherwise¹³.

SSI efficacy

SSIs have been shown to improve mental health outcomes in some populations, typically with small average effects that can, in some cases, persist for up to 9 months^{11,14,15}. However, evidence specifically regarding SSIs for depression remains mixed. A meta-analysis of 15 randomized controlled trials (RCTs) found that SSIs had a small overall effect on youth depressive symptoms ($g = -0.12$, 95% confidence interval (CI) -0.23 to -0.01), with substantial heterogeneity and a wide prediction interval¹⁶. However, multiple well-powered trials of SSIs with adult samples have not found substantial effects on depression^{17,18} or related outcomes such as anxiety¹⁹. Key questions remain about the populations for which SSIs are suitable (for example, adults versus adolescents) and which SSI designs are most effective.

✉ e-mail: bkavelad@uci.edu

SSIs are not a replacement for psychotherapy, nor are they sufficient to produce durable behaviour change in most cases. Rather, they offer a uniquely flexible alternative or adjunct to existing forms of mental health treatment, which are often inaccessible. Although SSIs' average effects are small at an individual level, if implemented at scale and in settings in need of on-demand support, they could reach many people underserved by existing care options and thereby improve public mental health^{20,21}. However, at this point, the global uptake and public health impact of SSIs remain uncertain.

SSI design

Current evidence-based SSIs typically focus on a core concept (for example, 'taking positive action can change your brain for the better'), reinforced through interactive exercises²². These SSIs aim to introduce evidence-based skills and immediately boost hope and agency—gains that can translate to longer-term change in mental health outcomes²³. SSIs' durations vary from 1 min (ref. 24) to 60 min (ref. 25), although most are between 5 min and 30 min. Some work has suggested that briefer SSIs (for example, 5–10 min) may be as useful as longer ones if they contain similar active elements²⁶.

Some best practices for SSI design have been proposed, building on evidence from longer digital mental health interventions, mindset interventions²⁷ and bibliotherapy self-help books²⁸. For example, the 'B.E.S.T.T.' elements of SSIs for youth mental health recommends brain science to normalize concepts and enhance credibility, empowering users to step into a 'helper' or 'expert' role, 'saying-is-believing' exercises to solidify learning, testimonials and evidence from valued others about SSIs' benefits, and targeting specific outcomes^{22,29}. Another model drawing from self-determination theory suggests that SSIs are most effective when they improve users' competence, autonomy and relatedness^{12,30}. SSIs aligning with these design principles provide solution-focused guidance in a clear format. Yet, these approaches may not be optimal for many problems and populations. Exploring novel design directions (that is, intervention strategies and stylistic elements) could advance SSIs towards greater efficacy and broader appeal.

Megastudies

An efficient way to investigate the potential of a diverse range of SSI designs is to conduct a 'megastudy'. Megastudies experimentally compare many interventions simultaneously using common outcomes, control conditions and sampling population³¹. This 'apples to apples' comparison enables faster, clearer insights than aggregating results across disparate observational studies and two-arm RCTs. By showing which approaches work and which do not, megastudies enable developers and researchers to direct their efforts towards better interventions, accelerating progress fieldwide³².

Megastudies have been conducted in several contexts. For example, a field megastudy that randomly assigned 61,293 members of a fitness chain to one of 53 interventions encouraging gym attendance found that offering small rewards for returning to the gym after a missed workout increased weekly visits by up to 27% (ref. 32). Another field megastudy ($n = 689,693$) showed that a text message reminder sent twice before a primary care visit boosted vaccination rates by 5% (ref. 33). A third megastudy tested 25 brief online interventions targeting anti-democratic attitudes and partisan animosity in a sample of 32,059 participants from the USA, finding the most successful interventions highlighted commonalities between likeable individuals with different political beliefs³¹.

The first step in organizing a megastudy is to identify interventions to test. Popular interventions and those with robust evidence from previous testing are useful starting points for this search, but they could overshadow innovative contributions from underrecognized teams without the means for rigorous testing. Crowdsourcing, in which many people are invited to submit interventions, can expand megastudies'

scopes to include these potential hidden gems. For example, the political megastudy noted above crowdsourced 252 submissions from global teams of academics and practitioners³¹.

The current study

Given SSIs' capacity for scale, finding innovative ways to increase their efficacy and appeal may bring them closer to their promise of global impact. We conducted a crowdsourced megastudy to identify and evaluate a diverse set of SSIs for depression in US adults reached via online participant recruitment platforms. We compared the efficacy and appeal of these SSIs with an evidence-based active comparator SSI, as well as an attention-matched control that provided information unrelated to mental health³⁴. Analyses examined which SSIs significantly reduced depression and improved related outcomes, and which features characterized the most and least helpful SSIs.

Results

Data collection occurred from March to June 2025. After exclusions, of 7,505 randomized participants, 7,338 (97.8%) completed the first session and 5,824 (77.6%) completed the 4-week follow-up. The sample's mean age was 36.6 years (s.d. 11.8), 59% were women and 36% were men, 74% were white and 14% were Black, 10.9% were of Hispanic, Latino or Spanish origin, and 60% had received some form of mental health support in the past month (Table 1). Mental health outcome distributions at baseline are shown in Fig. 1. The trial CONSORT diagram is available online (<https://osf.io/b7c86>).

Preliminary analyses

In the control condition, depressive symptoms (Patient Health Questionnaire-9, PHQ-9) decreased from baseline to week 4 ($b = -2.99$; $t(1,355) = -19.76$; $P < 0.001$; Cohen's $d = -0.58$; 95% CI -0.52 to -0.64). Among those in the control condition who completed both timepoints, depression at baseline correlated with depression at week 4 at $r = 0.48$; $t(1,206) = 18.80$; $P < 0.001$; 95% CI 0.43 to 0.52 . A total of 2.0% of participants dropped out of the study while completing the SSIs or control; the dropout rate differed across conditions ($X^2(12) = 71.65$, $P < 0.001$) and was below 2.6% for all conditions except mindful acceptance (6.2%). The dropout rate at 4-week follow-up was 21.0% and did not differ across conditions ($X^2(12) = 16.23$; $P = 0.181$). A total of 99.9% of participants passed the attention check item, and 99.8% spent at least 5 min on the main survey (median duration 20 min).

Change in outcomes across conditions at immediate post-test

The 12 SSIs (available at <https://osf.io/nqgte>) were each compared with the control, a 10-min educational programme unrelated to mental health, in their effects on psychological outcomes. Immediately after completing the treatment, participants assigned to 11 of the 12 SSIs had greater increases in agency than those assigned to the control ($d = 0.12$ to 0.37 ; $P = 0.003$ to <0.001); those assigned to 9 of the SSIs had greater reductions in hopelessness ($d = -0.11$ to -0.18 ; $P = 0.006$ to <0.001); those assigned to 11 of the SSIs had greater increases in depression change expectancies ($d = 0.12$ to 0.35 ; $P = 0.009$ to <0.001); and those assigned to 11 had greater increases in readiness to change ($d = 0.11$ to 0.28 ; $P = 0.014$ to <0.001); refer to Fig. 2.

In a preregistered secondary analysis, compared with the active comparator Behavioural Activation SSI, those assigned to two SSIs (Personalized Intervention Recommender ($b = -0.78$; $t(13,275) = -3.52$; $P < 0.001$; $d = -0.17$; 95% CI -0.08 to -0.27) and Inner Child Healing Walk ($b = -0.56$; $t(13,272) = -2.53$; $P = 0.012$; $d = -0.12$; 95% CI -0.03 to -0.22) had smaller increases in agency at post-test, while one SSI had a larger increase in agency (Interactive Cognitive Reappraisal, $b = 0.60$; $t(13,298) = 2.66$; $P = 0.008$; $d = 0.13$; 95% CI 0.03 to 0.23). Those assigned to one SSI had a smaller reduction in hopelessness at post-test (Personalized Intervention Recommender ($b = 0.30$; $t(13,283) = 2.11$; $P = 0.035$; $d = 0.10$; 95% CI 0.01 to 0.20)). Those assigned to four SSIs had smaller

Table 1 | Study participants’ demographic characteristics at baseline (N=7,505)

Characteristic	Mean (s.d.); n (%)
Age	36.64 (11.77)
Race	
White	5,541 (74%)
Black or African American	1,049 (14%)
Chinese	94 (1.3%)
Filipino	89 (1.2%)
American Indian or Alaska Native	87 (1.2%)
Asian Indian	69 (0.9%)
Prefer not to say	67 (0.9%)
Vietnamese	65 (0.9%)
Korean	59 (0.8%)
Japanese	34 (0.5%)
Hawaiian	6 (<0.1%)
Guamanian	3 (<0.1%)
Samoaan	2 (<0.1%)
Asian or Pacific Islander—Other	14 (0.2%)
Other	323 (4.3%)
Ethnicity	
No, not of Hispanic, Latino or Spanish origin	6,612 (88%)
Yes, Mexican, Mexican American, Chicano	423 (5.6%)
Yes, Puerto Rican	121 (1.6%)
Yes, Cuban	43 (0.6%)
Yes, another Hispanic, Latino or Spanish origin	229 (3.1%)
Prefer not to say	77 (1.0%)
Gender	
Woman	4,427 (59%)
Man	2,709 (36%)
Non-binary	242 (3.2%)
Agender	41 (0.5%)
Two-spirit	10 (0.1%)
Additional gender category/identity not listed	49 (0.7%)
Prefer not to say	27 (0.4%)
Disability or health condition	
No	3,917 (52%)
Yes	3,330 (44%)
Prefer not to answer	257 (3.4%)
Education	
No formal education	5 (<0.1%)
Less than a high school diploma	77 (1.0%)
High school graduate—high school diploma or the equivalent (for example, GED)	1,041 (14%)
Some college, but no degree	1,929 (26%)
Associate degree (for example, AA or AS)	803 (11%)
Bachelor’s degree (for example, BA, AB or BS)	2,451 (33%)
Master’s degree (for example, MA, MS, MEng, MEd, MSW or MBA)	912 (12%)
Doctorate degree (for example, PhD or EdD)	170 (2.3%)
Professional degree (for example, MD, DDS, DVM, LLB or JD)	102 (1.4%)
Prefer not to say	15 (0.2%)
Political party	
Democrat	3,159 (42%)
Independent	2,081 (28%)
Republican	1,561 (21%)

Table 1 (continued) | Study participants’ demographic characteristics at baseline (N=7,505)

Characteristic	Mean (s.d.); n (%)
Something else	477 (6.4%)
Prefer not to say	227 (3.0%)
Relationship/marital status	
Single	2,467 (34%)
Married	2,342 (32%)
In a relationship	1,688 (23%)
Divorced	454 (6.2%)
In a civil union/partnership	125 (1.7%)
Separated	117 (1.6%)
Widowed	90 (1.2%)
Prefer rather not say	29 (0.4%)
Employment status	
Full-time	3,563 (48%)
Part-time	1,294 (17%)
Unemployed	827 (11%)
Not in paid work (for example, homemaker or disabled)	664 (8.9%)
Student	555 (7.4%)
Business owner	355 (4.7%)
Retired	158 (2.1%)
Prefer not to say	84 (1.1%)
Household income	
Less than US\$10,000	459 (6.1%)
US\$10,000–US\$19,999	489 (6.5%)
US\$20,000–US\$29,999	632 (8.4%)
US\$30,000–US\$39,999	717 (9.6%)
US\$40,000–US\$49,999	700 (9.3%)
US\$50,000–US\$59,999	669 (8.9%)
US\$60,000–US\$69,999	536 (7.1%)
US\$70,000–US\$79,999	528 (7.0%)
US\$80,000–US\$89,999	341 (4.5%)
US\$90,000–US\$99,999	440 (5.9%)
US\$100,000–US\$124,999	670 (8.9%)
US\$125,000–US\$149,999	432 (5.8%)
US\$150,000–US\$174,999	262 (3.5%)
US\$175,000–US\$199,999	124 (1.7%)
US\$200,000–US\$224,999	85 (1.1%)
US\$225,000–US\$249,999	43 (0.6%)
US\$250,000 or more	104 (1.4%)
Prefer not to say	269 (3.6%)
Social ladder (1–10)	4.37 (1.73)
Mental health support received in past month (multiple selections permitted)	
No formal supports for mental health accessed	3,026 (40.3%)
Received some form of formal support (any of the below options)	4,475 (59.6%)
A therapist, coach or social worker (in person or online)	2,290 (30.5%)
A medical doctor (including psychiatrists)	2,088 (27.8%)
An online mental health app	1,146 (15.3%)
An offline mental health resource (a book or video)	1,043 (13.9%)
An online mental health support resource	815 (10.9%)
A support group (in person or online)	789 (10.5%)
Phone or text support	688 (9.2%)
Other	168 (2.2%)

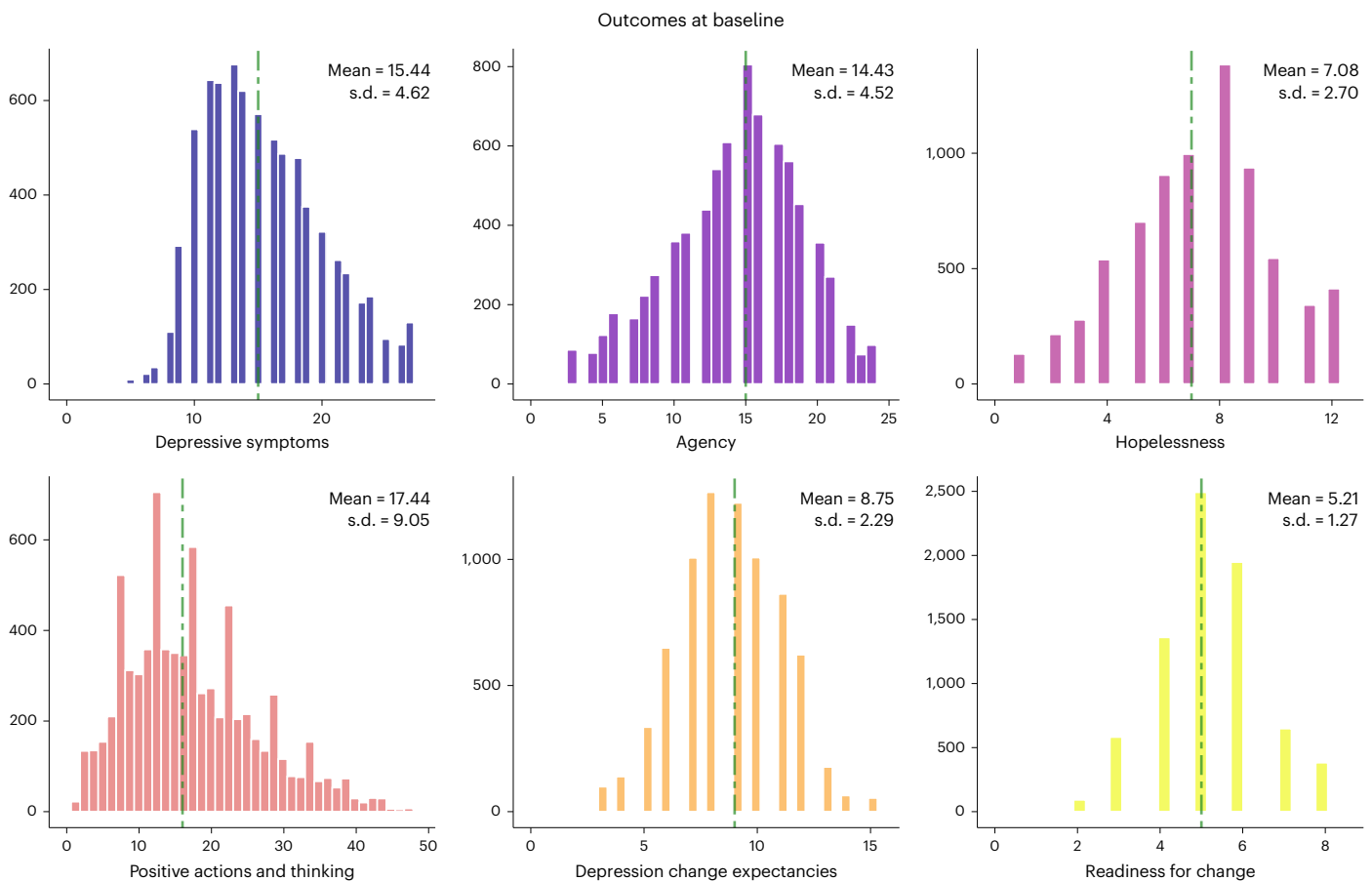


Fig. 1 | Distributions of mental health outcomes at baseline in a megastudy of 12 SSIs for depression ($N = 7,505$). The height of the coloured area shows the number of participants who reported the score on the horizontal axis at baseline. The vertical dot-dash green line shows the median value for each outcome.

For depressive symptoms, scores from 5 to 9 indicate mild depression, 10 to 14 indicate moderate depression, 15 to 19 indicate moderately severe depression and 20 to 27 indicate severe depression.

increases in depression change expectancies at post-test (Personalized Intervention Recommender, 5 Habits to Beat Depression, Inner Child Healing Walk and AI-Assisted Personalized Narrative; $d = -0.12$ to -0.21 , $P = 0.023$ to <0.001). Lastly, those assigned to two SSIs (Personalized Intervention Recommender ($b = -0.16$; $t(13,273) = -2.28$; $P = 0.023$; $d = -0.12$; 95% CI -0.02 to -0.22) and Inner Child Healing Walk ($b = -0.14$; $t(13,270) = -1.99$; $P = 0.047$; $d = -0.11$; 95% CI 0.00 to -0.21) had smaller increases in readiness to change at post-test. For full model output, refer to ‘Secondary Analysis H3’ at <https://osf.io/agvh6/files/r5kyb>

Change in outcomes across conditions at 4-week follow-up

At 4-week follow-up, compared with those assigned to the control, participants assigned to Interactive Cognitive Reappraisal ($b = -0.76$; $t(6,813) = -2.46$; $P = 0.014$; $d = -0.15$; 95% CI -0.03 to -0.26) and Mindful Attention Skills ($b = -0.72$; $t(6,704) = -2.37$; $P = 0.018$; $d = -0.14$; 95% CI -0.02 to -0.25) had greater reductions in depression. Participants assigned to six of the SSIs had greater increases in agency ($d = 0.08$ to 0.16 , $P = 0.048$ to <0.001), and those assigned to Interactive Cognitive Reappraisal had greater gains in positive actions and thinking ($b = 0.98$; $t(6,407) = 2.04$; $P = 0.041$; $d = 0.10$; 95% CI 0.00 to 0.20). Yet, participants assigned to Personalized Intervention Recommender showed smaller increases in readiness to change ($b = -0.18$; $t(13,595) = -2.90$; $P = 0.004$; $d = -0.14$; 95% CI -0.04 to -0.23); refer to Figs. 3–5. There were no differences by condition in change in hopelessness ($P > 0.083$) or depression change expectancies ($P > 0.057$).

None of the SSIs reduced or increased mental health outcomes at 4-week follow-up more or less than the active comparator Behavioural

Activation SSI ($P > 0.06$; refer to ‘Secondary Analysis H3’ at <https://osf.io/agvh6/files/r5kyb>). Secondary analyses examined the average change in outcomes between the mean of all SSIs and the control condition. At post-test, SSIs outperformed the control on every outcome ($d = 0.12$ to 0.21 ; $P < 0.001$); at week 4, SSIs showed greater reductions in depression ($b = -0.35$; $t(6,689) = -2.05$; $P = 0.041$; $d = -0.07$; 95% CI 0.00 to -0.13) and greater increases in agency ($b = 0.36$; $t(13,537) = 3.26$; $P = 0.001$; $d = 0.08$; 95% CI 0.03 to 0.13), but smaller increases in readiness to change ($b = -0.07$; $t(13,588) = -2.01$; $P = 0.045$; $d = -0.05$; 95% CI 0.00 to -0.10), while hopelessness, positive actions and thinking, and depression change expectancies did not differ (refer to ‘Secondary Analysis H6’ at <https://osf.io/agvh6/files/r5kyb>). Sensitivity analyses that added covariates to the depression analysis produced similar results to the primary model (refer to ‘Sensitivity Analyses’ at <https://osf.io/agvh6/files/r5kyb>).

SSIs’ effects on outcomes at post-test were correlated with their effects at week 4 in agency ($r(10) = 0.66$; $P = 0.020$; 95% CI 0.13 to 0.89), but not hopelessness ($r(10) = 0.34$; $P = 0.279$; 95% CI -0.29 to 0.77), depression change expectancy ($r(10) = 0.36$; $P = 0.246$; 95% CI -0.27 to 0.78), or readiness for change ($r(10) = 0.55$; $P = 0.063$; 95% CI -0.03 to 0.86), although this analysis had weak power with a sample size of only 12; refer to ‘Correlation Between Post-test and Week Four Effects’ at <https://osf.io/agvh6/files/r5kyb>).

Robustness checks. We conducted several checks to test if our findings were robust to varying analytic approaches. When adjusting for multiple comparisons between SSIs and the control (refer to ‘Robustness Check: Correcting for multiple comparisons’ at

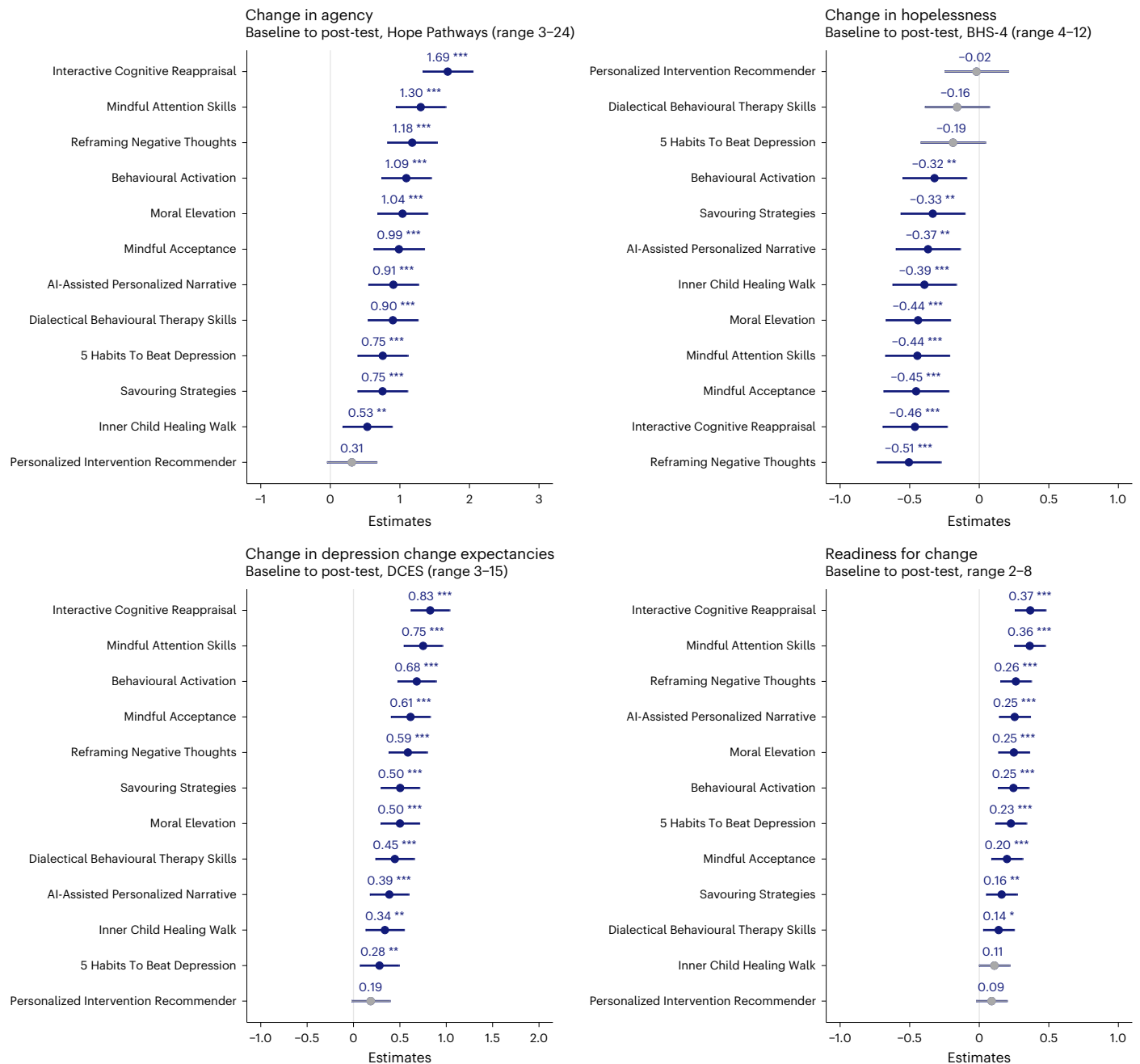


Fig. 2 | Intervention effects at immediate post-test in a megastudy (N = 7,505) of 12 SSIs for depression. Points reflect model estimates of a condition’s average change from baseline to immediate post-test, compared with the control condition. Error bars reflect 95% CIs; non-overlapping CIs are not evidence of

statistically significant differences between SSIs. Blue reflects a difference from the control in the desired direction. Grey bars reflect statistically insignificant differences from the control. *P < 0.05, **P < 0.01, ***P < 0.001.

<https://osf.io/agvh6/files/r5kyb>), all of the differences in mental health outcomes at post-test remained statistically significant. Among the differences at 4-week follow-up, the drop in depressive symptoms was no longer significant for both Interactive Cognitive Reappraisal and Mindful Attention Skills, the change in agency was no longer significant for Inner Child Healing Walk, and the change in positive actions and thinking was no longer significant for Interactive Cognitive Reappraisal. Rerunning the analyses with the complete sample (that is, including participants who were probably randomized multiple times) produced similar results to the original models, as did rerunning the analyses with the most exclusive sample (that is, removing all responses with a repeated IP address at baseline). Analyses accounting for differential

attrition across conditions using inverse probability weighting also produced similar results to the original models.

As measurement invariance is crucial to understanding intervention effects³⁵, we conducted non-preregistered invariance tests on the PHQ-9³⁶ at week 4, as well as agency and hopelessness scores at immediate post-test using alignment optimization³⁷. All conditions except Reframing Negative Thoughts and Personalized Intervention Recommender had higher agency than the control. The following groups showed higher levels of hope than the control (in mean-difference size order): Behavioural Activation, 5 Habits to Beat Depression, Mindful Acceptance, Savouring Strategies, Reframing Negative Thoughts, Inner Child Healing Walk, Moral Elevation, AI-Assisted Personalized

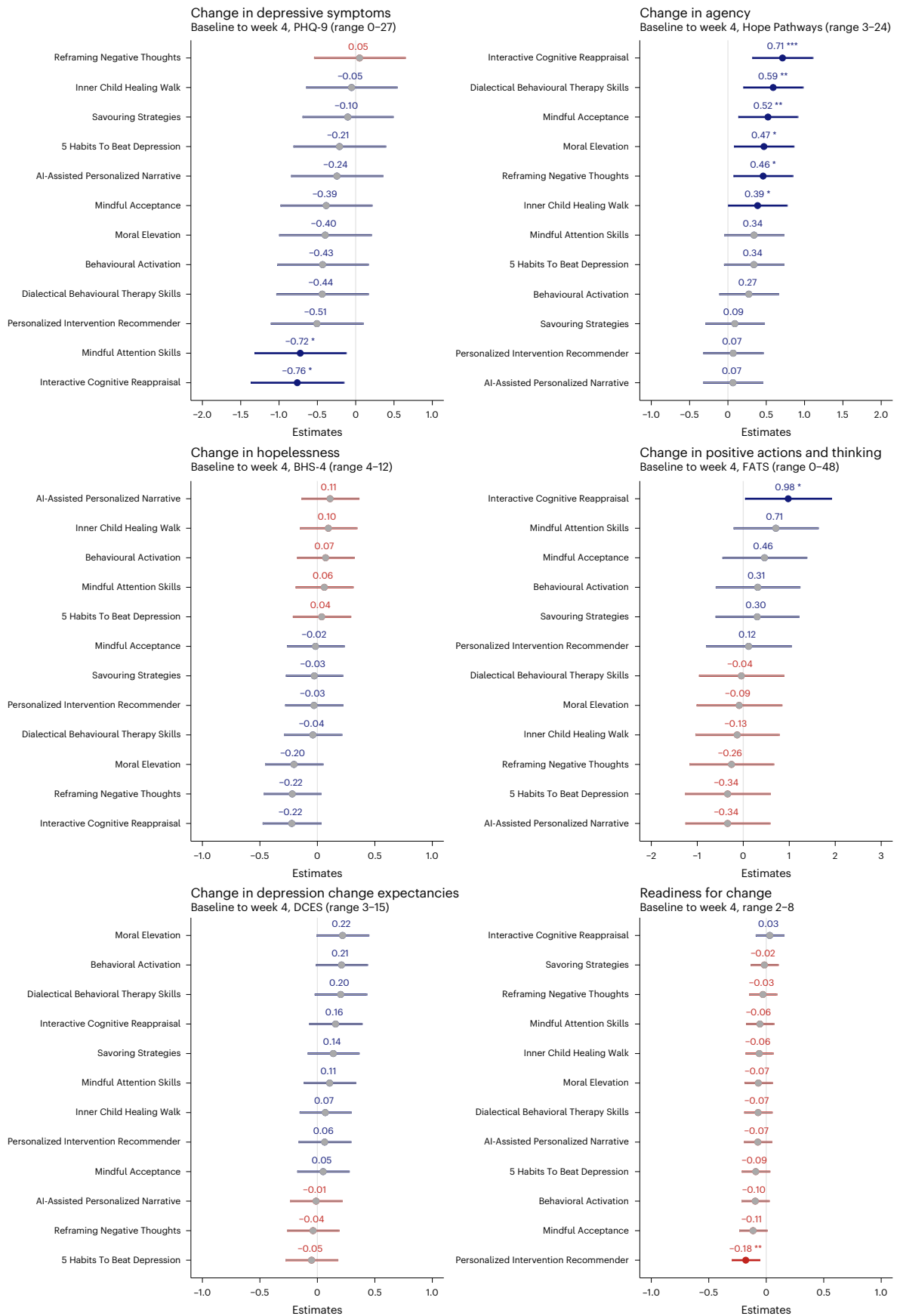


Fig. 3 | Intervention effects at 4-week follow-up in a trial (N = 7,505) of 12 SSIs for depression. Points reflect model estimates of a condition’s average change from baseline to 4-week follow-up compared with the control condition. Error bars reflect 95% CIs; non-overlapping CIs are not evidence of statistically

significant differences between SSIs. Blue reflects a difference from the control in the desired direction, and red reflects the opposite. Grey bars reflect statistically insignificant differences from the control. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

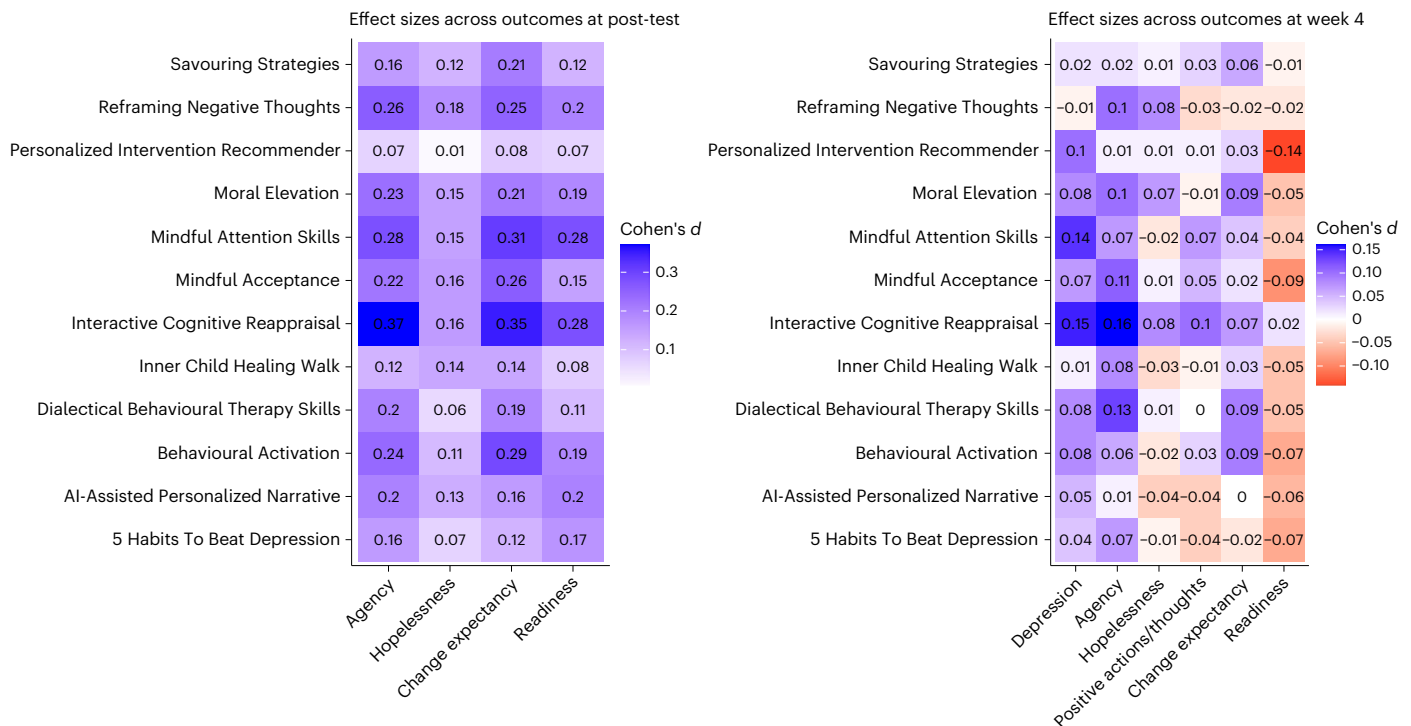


Fig. 4 | Effect sizes by outcome and condition at post-test and week 4 in a trial ($N = 7,505$) of 12 SSIs for depression. Values reflect Cohen's d effect sizes expressing the difference in the change from baseline to the stated timepoint

compared with the control condition. Blue reflects change in the desired direction, and red reflects the opposite. Hopelessness and depression were reversed to standardize effect direction.

Narrative and Mindful Attention Skills. For PHQ-9, only Mindful Acceptance showed longer-term improvements in depression than the control.

SSI acceptability and expected benefit

Figure 5 shows participant ratings of SSIs' acceptability and expected benefit at post-test. At 4-week follow-up, none of the SSIs differed from the active comparator SSI in user-rated overall quality (that is, star rating); $d \leq 0.14$, $P > 0.08$) and Inner Child Healing Walk had a lower likelihood of producing an Aha! moment than the active comparator ($d = 0.23$, $P = 0.002$). Refer to 'Secondary Analysis H4' at <https://osf.io/agvh6/files/r5kyb> for more information on acceptability and expected benefit comparisons between the active comparator SSI (Behavioural Activation) and all other SSIs.

Moderators of change

A preregistered analysis showed participants who expected they could improve their depression more at baseline had greater decreases in depression by week 4 ($b = -0.08$, $t(6,707) = -2.85$, $P = 0.004$). An exploratory analysis limited to participants who were assigned to an SSI found those who had not accessed any form of formal mental health support in the past month had a greater decrease in depression by week 4, on average ($b = -0.62$; $t(5,327) = -4.02$; $P < 0.001$; $d = -0.12$; 95% CI -0.06 to -0.18).

Additional analyses

As an exploratory analysis, we examined the intervention features (strategies and stylistic elements) of the most and least efficacious and acceptable SSIs. Among the 12 features we coded for the SSIs before data collection, none predicted SSI efficacy or acceptability, except that meditation practice predicted lower credibility and expected benefit (refer to 'Additional Analysis I' at <https://osf.io/agvh6/files/r5kyb>). Several other additional analyses are also presented at <https://osf.io/agvh6/files/r5kyb>.

Discussion

SSIs are beginning to fill gaps in care, enabling more people to access reliable support in moments of need¹². We crowdsourced 66 diverse SSIs for depression from global teams and conducted an experiment ($N = 7,505$) to evaluate the 11 most promising ones against an evidence-based active comparator and a control. Many SSIs improved depression-relevant outcomes in the immediate term relative to the control ($d \leq 0.37$), but these effects had substantially weakened 4 weeks later ($d \leq 0.16$). At 4-week follow-up, only two of the SSIs (Interactive Cognitive Reappraisal ($d = 0.15$) and Mindful Attention Skills ($d = 0.14$)) statistically significantly alleviated depression, although others had effect sizes that were nearly as large. After correcting for multiple comparisons, all of the statistically significant differences at post-test remained significant, but several of the differences at 4-week follow-up lost statistical significance, including both significant SSI effects on depression (refer to 'Robustness Check: Correcting for multiple comparisons' at <https://osf.io/agvh6/files/r5kyb>). Results highlight the efficacy and appeal of several kinds of SSIs, some reflecting current design norms and others introducing novel approaches. For example, several video-based SSIs—a previously unexplored SSI format—were efficacious and well liked.

Interpreting results

In our main analysis, the SSI that most reduced participants' depressive symptoms at 4-week follow-up (Interactive Cognitive Reappraisal) did so by 0.76 PHQ-9 points more on average than the control ($d = 0.15$, $P = 0.014$). This effect is close to the average effect size of SSIs for youth depression $g = -0.12$ (ref. 16) and larger than the effects reported in previous studies of SSIs in adults^{17,18}. Yet, the average drop in depression across all 12 SSIs was only slightly greater than the control's ($d = 0.07$, $P = 0.041$). Also, this 'winning' effect may reflect a modest winner's curse inflation³⁸.

Although these 10-min SSIs' effects were substantially smaller than those of longer-term interventions for depression (which have

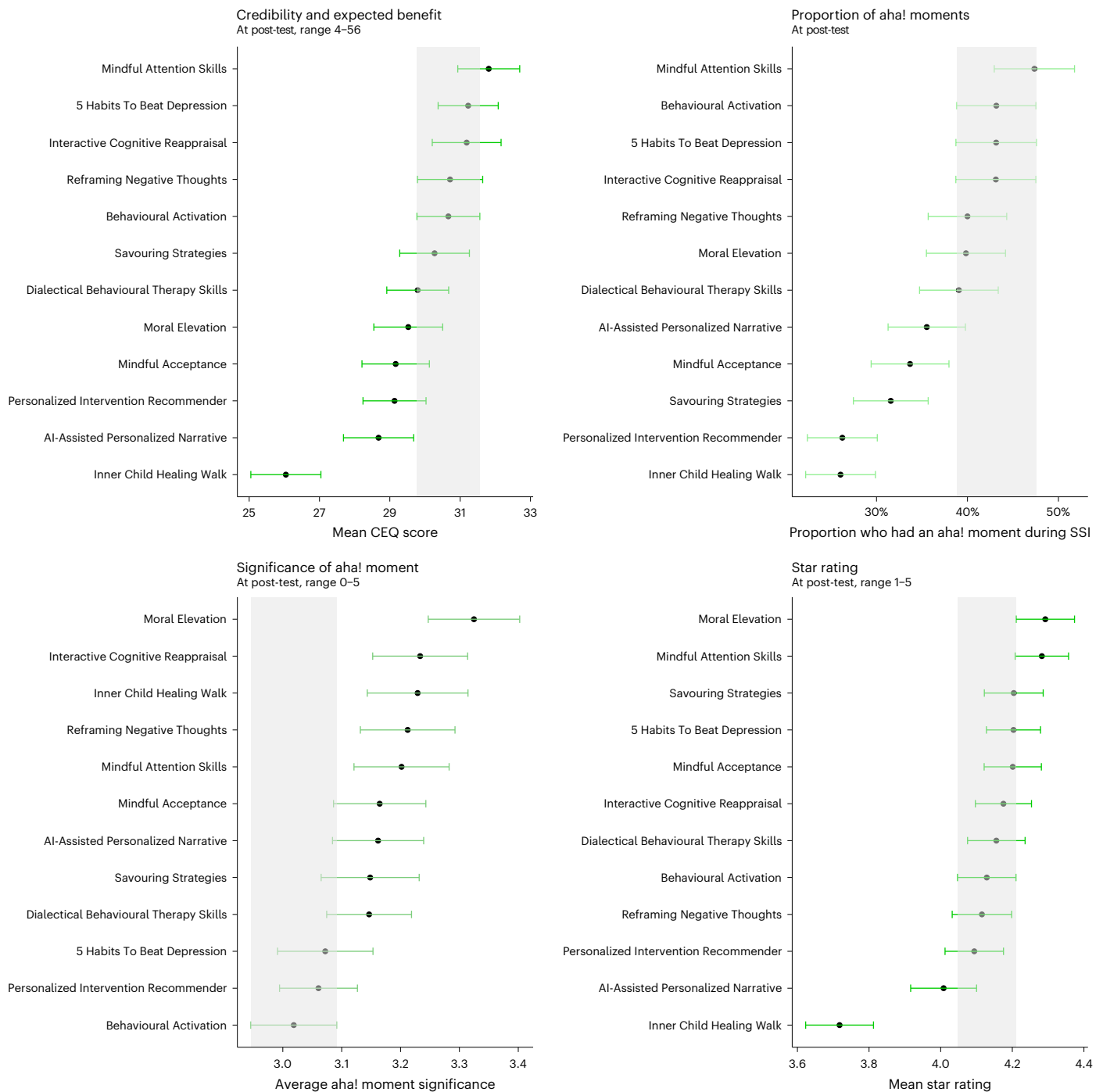


Fig. 5 | Acceptability outcomes in a trial (N = 7,505) of 12 SSIs for depression. Points reflect the average rating on each outcome, except for those in the top right plot, which refer to the percentage of participants who experienced an

‘aha!’ moment. Error bars reflect 95% CIs. Shaded regions reflect the error bar of the behavioural activation SSI. Non-overlapping error bars differ significantly ($P < 0.05$).

an average effect size $g = 0.43$, ref. 39), their value proposition (that is, a measurable reduction in depression from an on-demand 10-min programme) may still be compelling to many people seeking instant support, especially if those 10 min are enjoyable. As noted, because 10-min SSIs are low-cost, highly scalable and minimally burdensome, these small individual effects could add up to a public health impact (refer to the ‘Implementation’ section below).

Secondary analyses comparing each submitted SSI with an active comparator SSI (Behavioural Activation) found no significant differences in mental health outcomes at 4-week follow-up, although there

were several differences at immediate post-test. Overall, these results suggest that the current standard of SSI design compares decently to a range of other SSI designs in efficacy and appeal, although these comparisons would certainly benefit from greater statistical power.

The SSIs’ effects on outcomes at post-test correlated moderately with their effects at week 4 (refer to ‘Additional Analysis 2’ at <https://osf.io/agvh6/files/r5kyb>), although there were several inconsistencies across timepoints. For example, Reframing Negative Thoughts decreased hopelessness more than any other SSI at post-test but had nearly the weakest effect on it at week 4.

Potential iatrogenic effects on longer-term readiness to change. Surprisingly, at 4-week follow-up, participants assigned to SSIs expressed slightly lower readiness to make changes towards overcoming depression than those assigned to the control, on average. This pattern raises the possibility of mild iatrogenic effects. It is possible that, once the initial boosts in confidence and motivation from an SSI faded, participants felt let down because they had hoped for a change that never materialized. If this interpretation is correct, it would be the second published instance of an iatrogenic effect from an SSI⁴⁰.

Alternatively, this finding might reflect an issue with participants' interpretation of the items on the readiness to change measure. Specifically, participants in the control condition may have expressed higher readiness to change because they had only received information about trout so far, and remained ready to receive an intervention because the study was advertised as an opportunity to 'help test an online mental health resource'. Because the readiness-to-change scale was developed for this study, its psychometric validity remains uncertain. This uncertainty precludes strong conclusions on whether this reflects an important iatrogenic effect.

Why some SSIs may have outperformed others. Because the experiment lacked statistical power for high-resolution comparisons between SSIs, efforts to generalize about the SSI features that were more efficacious and appealing than others are extremely tentative. However, here we note some possible patterns across SSIs that may be worth further investigation.

The most efficacious SSIs across outcomes and timepoints, such as Interactive Cognitive Reappraisal and Mindful Attention Skills, focused on a single practical skill with a compelling call to action and an emphasis on helping users apply learnings into their daily lives. By contrast, SSIs such as Inner Child Healing Walk and Personalized Intervention Recommender may have felt too abstract because they lacked specific targets. And while 5 Habits to Beat Depression was well liked for its engaging and professional presentation, its rapid presentation of five habits might have left participants unclear on their best path to change.

Although Interactive Cognitive Reappraisal and Mindful Attention Skills were similar in their focus on an evidence-based skill and their targeting of negative thought patterns, they also differed substantively. For one, Interactive Cognitive Reappraisal was text-based, while Mindful Attention Skills centred on cinematic video lessons. Notably, Mindful Attention Skills' key message—that training attention gives greater agency over negative thoughts—was similar to that of Reframing Negative Thoughts, the SSI with the weakest effect on depression at 4 weeks. However, Mindful Attention Skills featured engaging videos and interactive exercises, while Reframing Negative Thoughts was purely audio. Perhaps Reframing Negative Thoughts' lack of interactive exercises applying concepts to users' daily lives held back its long-term impact.

These (speculative) observations mostly align with existing SSI design principles, such as the B.E.S.T.T. elements^{22,29} and empowering users towards self-determination^{12,30}. Yet, our findings also highlight the value of novel approaches such as video-based intervention content. Ultimately, SSIs' effectiveness is driven by a complex interplay of factors, including alignment with the target population's preferences, delivery format, emotional resonance and user burden. Some important design factors may be difficult to operationalize—both of the leading SSIs in this study, Interactive Cognitive Reappraisal and Mindful Attention Skills, are the products of nearly 10 years of iterative design that probably amounts to more than the sum of their intervention components^{41,42}.

Strengths and limitations of the study

This study had several strengths, including low attrition, a large and diverse sample achieved through broad entry criteria and nationwide recruitment, and the inclusion of two comparator conditions: a control

and an active comparator SSI. In addition, the SSIs included for testing reflected a global crowdsourcing effort and an organized selection process. Yet, factors such as the study's relatively short follow-up (4 weeks), lack of intermediary timepoints, all SSIs being under 10 min, and minimal adverse event monitoring limit our conclusions.

Our choice to conduct our study using the online recruitment platforms CloudResearch Connect and Prolific had several advantages. Recruiting such a large sample on social media would have been difficult given the problem of fraudulent research participants⁴³. The rates of attrition, within and between sessions, on the platforms we used are much lower than typical for social media recruiting, which alleviated differential dropout^{44,45}. However, this choice of sample also had weaknesses. Online workers differ from the broader populations SSIs aim to serve in several ways⁴⁶; for example, many participants were probably not actively seeking mental health support when they found the study and might differ from real-world help-seekers in their motivation and attention to intervention content. Overall, caution is needed when generalizing our results beyond the population sampled: non-probability opt-in panels of adult online workers experiencing elevated (mostly moderately elevated) depressive symptoms.

We aimed to include a wide range of intervention strategies and styles, a choice that had both advantages and limitations. One strength was that it made the study more reflective of all possible designs and increased our odds of finding unexpectedly fruitful SSIs. Yet, this variety also made it harder to know why an SSI succeeded. With only 12 SSIs to compare, we could not be confident that a given feature led to greater efficacy or acceptability. Increasing the number of SSIs tested and reducing their variety may have improved the study's chances of producing strong SSI design recommendations.

Relatedly, to expand the range of SSIs in the study, we took efforts to crowdsource SSI submissions from diverse teams, including teams with no involvement in research on SSIs or mental health. Yet, most submissions came from teams connected to academic research, probably because we shared the project via our core team's online social networks (that is, mostly researchers). More extensive advertising and monetary prizes for successful submissions may have helped to reach a broader range of submitters. In addition, although we took efforts to make the crowdsourcing and SSI selection process as fair as possible, after selecting nine SSIs through our planned process, we added two SSIs (Dialectical Behavioural Therapy Skills and Reframing Negative Thoughts) through separate processes that were more influenced by submitters' reputations and relationships with our team (refer to Methods).

The study's statistical power was another source of strengths and weaknesses. On the one hand, the trial had a large sample and was appropriately powered to detect small effect sizes similar to those observed in prominent SSI trials³⁴. On the other hand, most of the comparisons in the study were still not sufficiently powered to detect effects that might be valuable. A useful target would be the average effect size of SSIs for youth depression observed in Ball et al.'s meta-analysis, $g = -0.12$ (ref. 16). However, detecting such a small effect would require a much larger sample than we had in this study.

Future directions

One avenue for future work is to further study the SSIs that showed the greatest promise in this megastudy; while this study showed which SSIs are efficacious, future work can more closely examine why. For example, coming megastudies might hold elements such as style or content constant across SSIs to pinpoint active elements and mechanisms of change. A related direction is to apply our findings to psychological interventions more broadly, for example, by integrating SSIs into longer interventions.

Implementation. A crucial future direction is implementing SSIs into systems that can maximize their impact. Even if SSIs are effective and appealing, simply making them freely available online is unlikely

to spark organic virality in today's competitive attention economy. Thoughtful implementation into settings with wide reach, such as social media and schools, or high need, such as therapy waitlists, is key.

Despite their successes, existing SSIs have not achieved digital mental health's holy grail: organic uptake by the millions of people underserved by current mental health solutions. This lack of reach might reflect myriad barriers, including low public awareness of SSIs and the pathways to accessing them. Massive reach is not the only pathway to meaningful impact, but it is a particularly important one for digital, self-guided SSIs given their small average effects²⁰.

SSIs' appeal is an essential component of their path to impact. While this study's acceptability measures (Credibility/Expectancy Questionnaire (CEQ), 'aha!' moments and star ratings) are a useful signal as to the kinds of SSIs that users resonate with; real-world data on appeal are much more valuable. For example, the motivational video central to Moral Elevation achieved virality, garnering over 120 million views on YouTube since 2014 (<https://www.youtube.com/watch?v=uaWA2GbcnJU>). While Moral Elevation was not the most efficacious SSI in our megastudy, its small benefits at its massive scale may mean it has the greatest real-world impact of any SSI.

Another implementation consideration is that the SSIs we tested might have different average effects when used in real-world settings⁴⁷. Engagement with digital mental health interventions tends to be poorer outside of clinical trial contexts⁴⁸, which could reduce effectiveness. Yet, as noted, real-world help-seekers might benefit more from SSIs than our study participants because they could be more motivated to get help. In addition, if SSIs achieve wide reach, their downstream effects may increase their public health impact—boosts in an SSI user's mood could spread to those around them, and SSIs' skills and concepts could diffuse into popular culture.

Large-scale implementation also requires careful attention to potential risks. Expanding the populations an SSI reaches might reveal unwanted spillover effects, as in the possible drop in long-term readiness to change we observed in this study⁴⁹. Similarly, reaching broader populations could increase the odds that SSIs' messaging is misinterpreted by some, especially because self-guided SSIs lack monitoring and opportunities for clarification. As an extreme example, some users might misinterpret an SSI that encourages 'facing challenging situations' as suggesting they should pursue a dangerous activity. While it may be reasonable to infer, based on SSIs' light-touch nature, that most people are at low risk of substantial downsides, it remains important to intentionally track adverse events using harmonized assessments that go beyond average change in mental health outcomes⁵⁰.

Leveraging short-term gains. Hopelessness and a lack of agency often prevent people experiencing depression from seeking care or taking other positive actions^{51,52}. Our results at immediate post-test make it clear that many users finish SSIs with a sense that further improvement is possible, especially when SSIs provide actionable guidance. However, by 4-week follow-up, these effects were much smaller, and readiness to change was slightly lower among those assigned to an SSI than the control.

One way to boost SSIs' efficacy may be to harness the apparent window of heightened motivation just after completing an SSI to encourage longer-term engagement in behaviour change and provide opportunities for service uptake. For example, SSIs could be used as entry points to a stepped-care model, in which people who complete an SSI are referred to longer-term digital health tools^{53,54}. Other work could help users find the SSI that best suits them, either through machine learning-based treatment-matching models or by allowing users to choose from a variety of SSIs.

Conclusion

Our findings suggest that several kinds of SSI for depression can meaningfully improve mental health outcomes, at least in the immediate

term. Longer-term gains tend to be much smaller; at 4-week follow-up, only 2 of the 12 SSIs we tested significantly reduced depression (and none did so after correcting for multiple comparisons). Work evaluating SSIs outside of paid study contexts is needed to determine their real-world effectiveness.

Methods

The study procedures were approved by the Northwestern University Social and Behavioral Sciences Review Board (protocol STU00220193) and were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Informed consent was obtained from all research participants. Participants were compensated US\$4.25 for completing all study activities. We preregistered the procedures, hypotheses, protocol, analysis plan, measures and decision criteria on 8 March 2025 before collecting data (refer to [ClinicalTrials.gov](https://clinicaltrials.gov) ID: [NCT06856668](https://clinicaltrials.gov/ct2/show/study/NCT06856668)).

Crowdsourcing interventions

Overview. We advertised an open call for SSIs for depression and received 66 submissions. The 50 submitting teams were based in 5 continents and involved 112 contributors, including students, popular YouTubers and representatives of leading mental health apps. Next, a team of digital mental health researchers and people with lived experience of depression chose a group of SSIs reflecting a range of promising approaches for further testing.

Crowdsourcing intervention submissions

We identified SSIs for the megastudy via crowdsourcing, with an open call for submissions from March to May 2024. Our crowdsourcing process largely followed the example set by a previous crowdsourced megastudy: the Strengthening Democracy Challenge³¹.

We created a website (<https://sites.northwestern.edu/10minutechallenge/>), handbook (<https://osf.io/agvh6/files/nkqzm>) and video (https://www.youtube.com/watch?v=Y9cSv_wenX4) explaining the project and how to submit. Next, members of our study team posted about the megastudy on social media and online communities on Reddit and Discord, aiming to reach as many people as possible with a broad range of backgrounds and perspectives. We also invited several content creators (for example, YouTubers) who we believed had relevant content to submit.

Anyone 13 years or older could submit to the megastudy. One could submit either an SSI or a description of a SSI (provided our teams could feasibly create the SSI together within the next few months). Any kind of SSI was eligible for review if it was under 10 min, not hateful or disturbing, online, scalable, self-guided, costless and comprehensible to English speakers. The submission portal asked about the submitting team's background, the SSI's theory of change and how the SSI might be disseminated; all of these questions are listed at <https://osf.io/8gys7>. Submitters were not compensated. However, we stated that, if an SSI was selected for testing, all members of the submitting team would be offered authorship on this paper.

Choosing submitted interventions to include in the megastudy experiment

Before crowdsourcing SSIs, we decided to include one active comparator SSI in the experiment: an evidence-based SSI for depression against which to compare the submitted SSIs. For this purpose, we used the adult version of the Action Brings Change Project (referred to here as Behavioural Activation), an SSI originally developed for teens³⁴. Although this SSI had not shown efficacy in reducing depression in adult samples, based on its success in adolescent samples³⁴, as well as formative evaluations in adult samples (refer to ref. 26 and Study 2 in ref. 55), we decided it was the most promising available comparator.

We received 66 SSI submissions and planned to include 10 for testing. To decide which of the submitted SSIs to include, we used a

Table 2 | Experimental conditions in a megastudy (N=7,505) of adult online workers experiencing depressive symptoms

Condition	Description	Prior testing
5 Habits to Beat Depression	A clinical psychologist and science communicator describes five 1-min habits to beat depression.	None
AI-Assisted Personalized Narrative	Users describe a negative thought they struggle with, and a large language model system generates a story in which someone overcomes that thought. Users are prompted to reflect on their feelings and the extent to which they believe each thought.	RCT ⁷⁵
Dialectical Behavioural Therapy Skills	A series of videos that present a subset of skills from dialectical behavioural therapy, aiming to regulate and manage negative emotions.	RCT of longer format ⁵³
Interactive Cognitive Reappraisal	An interactive and aesthetically pleasing text-based SSI in which participants learn the value of reframing negative thoughts, then reframe a thought of their own.	RCT of longer format ⁴¹
Inner Child Healing Walk	An interactive animated journey to reconnect and heal one's inner child through breathing exercises, positive affirmations and supportive animal companions.	None
Mindful Acceptance	A clinical psychologist explains ways to use mindful acceptance to deal with difficult thoughts and feelings.	RCT of longer format, unpublished (https://drive.google.com/file/d/1mwVle20JMO10abDiRWUK_Wovphurx_iA/view)
Mindful Attention Skills	A series of videos drawn from an intervention called Finding Focus explaining how to identify counterproductive evaluations and replace them with new evaluations that feel both true and helpful.	RCT of longer format ⁴²
Moral Elevation	Users learn about moral elevation, then watch a touching video (a Thai life insurance advertisement with over 120 million views as of July 2025) where a man does good deeds for people in his community. Finally, they plan an action to help others in their own life.	RCT of longer format ⁷⁶
Personalized Intervention Recommender	An interactive SSI that identifies the depression symptom one would most like to improve and offers a personalized recommendation for support options (for example, a self-guided mental health app or a professional therapist), matching one's preferences and capacity.	None
Reframing Negative Thoughts	An audio-only programme in which users are guided to reflect on how they might challenge negative automatic thoughts and beliefs they struggle with.	RCT of longer format ⁷⁷
Savouring Strategies	An interactive SSI targeting mood. It introduces strategies to savour positive memories, the present moment and future events.	None
Behavioural Activation (Active Comparator SSI)	An interactive behavioural activation SSI where users create an 'action plan' of positive social and personal actions that can improve their well-being. This SSI has demonstrated efficacy and reflects current 'best practices' in SSI design.	RCTs ³⁴
All About Trout (Attention Control)	An educational programme with information and interactive exercises about trout fish. Intended to require some attention and effort but minimally influence depressive symptoms and mood.	None

Browser versions of each intervention are available at <https://osf.io/agvh6/files/nqdtc>, with the exception of Inner Child Healing Walk as it could not be made publicly available. Downloadable materials are available at <https://osf.io/agvh6/files/btc7q>.

six-stage evaluation process involving teams of reviewers with varied areas of expertise, as well as 335 pilot testers (refer to <https://osf.io/z43nq> for information on this evaluation process). We engaged a variety of reviewers in the selection process to address potential personal and disciplinary biases. Broadly, our selection process balanced two motivations: to identify SSIs with the greatest potential for real-world effectiveness, and to include a diverse range of SSI designs. Thus, we had to consider trade-offs such as evidence versus novelty, as well as efficacy versus broad appeal.

After selecting nine SSIs, the editorial board decided that the submitted SSIs did not reflect a sufficiently broad range of intervention approaches. As a result, the board decided to invite an additional SSI submission from a team known to have created evidence-based dialectical behavioural therapy-centred digital content. That team submitted an SSI (a subset of Dialectical Behavioural Therapy Skills), and the editorial board decided to include it for testing. Four months later, the editorial board decided to include a 12th SSI (Reframing Negative Thoughts) from Calm Health as thanks for making a monetary donation to JLS's laboratory that enabled data collection for this study. The Calm Health team had no input in analysis or paper writing beyond authorship (C.J.M. and K.S.H.). None of the SSIs tested came

from core authorship team members' own teams. However, several of the included SSIs' submitting teams had collaborated with members of the core authorship team in the past.

Interventions. The 12 SSIs varied in format (for example, video versus text), intervention approach (for example, cognitive behavioural therapy versus mindfulness practice), and their creators' backgrounds (for example, clinical researchers versus popular online content creators). The control condition aimed to require as much attention as the other SSIs without influencing depressive symptoms or mood (a 3-min video, multiple-choice questions, reading passage and writing exercise, all about trout). Refer to Table 2 for information on conditions.

After collecting the 12 SSIs, the research team identified a set of SSI design characteristics that they suspected might predict differences in SSI effectiveness and engagement, informed by previous research on digital interventions⁵⁶. The characteristics were 'Video-based', 'Engaging', 'Interactive', 'Personalized', 'Effortful', 'Refers to outside mental health resources', 'Contains a meditation exercise', 'Contains features of cognitive reframing', 'Contains features of behavioural activation', 'Contains features of mindfulness', 'Created by a team led by clinicians or clinically trained mental health researchers' and 'Provides

Table 3 | All primary, secondary and screening outcome variables collected during a megastudy (N=7,505) comparing 12 SSIs for depression, by timepoint administered

Measure	Example item	Screen	Baseline	Post-test	4-week follow-up
PHQ-9 (primary)	How often have you experienced little interest or pleasure in doing things?	X	X		X
FATS (secondary)	How often did you change your thinking to be more realistic and helpful?		X		X
Agency: Pathways Subscale of the State Hope Scale (secondary)	There are lots of ways around any problem that I am facing now.		X	X	X
BHS-4 (secondary)	I feel that the future is hopeless and that things cannot improve.		X	X	X
Readiness to Change (Secondary)	How confident are you about making changes toward overcoming depression?		X	X	X
DCES—3-item version (secondary)	If I work hard, I can have a positive impact on my problems with depression.		X	X	X
CEQ (secondary)	How confident would you be in recommending this program to a friend who experiences similar problems?			X	

a take-home reminder'. Three raters (A.E.S., E.K.S. and B.T.K.) rated each SSI on various characteristics, then met to come to an agreement on these ratings. These ratings are provided at <https://osf.io/qbpbk6>.

Among the 11 submitting teams whose SSIs were tested, 9 included at least 1 member who was a digital mental health researcher, 8 included a clinical psychologist, 4 included a researcher from a field that was not psychology, 3 included members representing digital mental health companies and 3 included undergraduate students.

Megastudy experiment

Overview. Participants were randomized to 1 of 13 conditions to evaluate SSIs' efficacy in improving depression and related outcomes over 4 weeks.

Trial design. We used an experimental design with randomized, parallel assignment and a superiority framework. All conditions were allocated the same number of participants except for the passive control, which was allocated three times as many. We did not change the trial design after it commenced.

Participants. Participants were recruited from the online participant recruitment platforms CloudResearch Connect and Prolific^{57,58}. Participants were invited to participate in study sessions and compensated through their online recruitment platform. All other study activities took place through the online survey and experience management platform Qualtrics, via users' personal computers, tablets or smartphones⁵⁹. Participant randomization was automated through Qualtrics as well, using its 'evenly present elements' feature to allocate a roughly even number of participants to each condition. Participants and researchers did not interact during the study (except through automated reminder emails the researchers sent participants and a few cases in which participants asynchronously asked questions through the research collection platforms' direct message tool and B.K. replied via direct message). Researchers and participants were blind to condition assignment throughout the study.

Inclusion criteria. All inclusion criteria were preregistered before data collection. Participants needed to be located in the USA, at least 18 years old and able to read and write fluently in English. They also had to score at least 10 on the PHQ-9 screen during the screen survey, suggesting moderate depression⁶⁰. We included all participants who were randomized to a condition (and thus completed baseline measures) in the analyses (intention to treat). If a participant began a survey multiple times, we included only data from the first session in the analyses.

Procedure. First, participants completed a screen survey where they responded to the PHQ-9 and encountered a 'honeypot' item checking for bots (that is, an element of the survey that humans

should not usually find; <https://osf.io/kba7v>). Participants who failed the honeypot were automatically prevented from continuing in the study. Eligible participants were then asked to consent to participate in the study. After consenting, participants immediately began the baseline well-being questionnaire. Next, they were randomly allocated to one of the 12 SSIs or a control condition. After completing the treatment, they were asked to respond to some of the baseline measures again, as well as measures of satisfaction with the SSI. Four weeks later, participants were invited to a follow-up survey with the baseline well-being measures again and several SSI satisfaction measures.

Participants were compensated after completing each study session: US\$0.25 for screening (1 min), US\$3.00 for the first session (18 min) and US\$1.00 for the 4-week follow-up (5 min). Participants who were randomized to a condition were invited to complete the follow-up. After each study session, participants were offered a list of online mental health resources (refer to <https://osf.io/wz5u6>). At the end of the follow-up survey, participants were given access to online versions of all of the SSIs in the study (available at <https://osf.io/nqgte>).

Measures. We evaluated outcomes using self-report measures. Refer to Table 3 for measure collection timepoints, and <https://osf.io/76cfy> for all items. For each participant, at each measurement timepoint, we randomized the order in which measures were presented. All measures used in preregistered analyses are quantitative.

To measure depressive symptoms, we used the PHQ-9⁶⁰. In this scale, participants rate how often they have been bothered by nine concerns over the past 2 weeks on a scale from 0 ('Not at all') to 3 ('Nearly every day'). The total score ranges from 0 to 27, with higher scores indicating more severe depression. The PHQ-9 had a Cronbach's α of 0.75 at baseline, suggesting it was an internally consistent measure of depressive symptoms in this sample and session.

We measured participants' positive actions and thoughts (for example, engaging in social behaviours and challenging their negative thoughts) using the Frequency of Actions and Thoughts Scale (FATS)⁶¹. The scale includes 12 items about one's experiences over the past week. Items are rated from 0 ('Not at all') to 4 ('Every day'), with total scores ranging from 0 to 48. The FATS had $\alpha = 0.76$ at baseline.

To measure agency, we used the Pathways Subscale of the State Hope Scale, a three-item self-report measure of one's perceived ability to generate plans and work towards goals⁶². The scale asks about respondents' beliefs 'here and now'. Items are rated from 1 ('Definitely false') to 8 ('Definitely true'), with total scores ranging from 3 to 24. The measure had $\alpha = 0.85$ at baseline.

To measure hopelessness, we used the four-item Beck Hopelessness Scale (BHS-4)⁶³. The scale asks whether various statements are typical, with no specific time frame. Items are rated from 1 ('Rarely typical') to 3 ('Very typical'), with total scores ranging from 4 to 12. The BHS-4 had $\alpha = 0.87$ at baseline.

We evaluated readiness to make changes towards overcoming depression using two multiple-choice questions rated from 1 ('Not at all') to 4 ('Extremely'): 'How important is making changes toward overcoming depression to you right now?' and 'How confident are you about making changes toward overcoming depression?'. We summed these items into a readiness to change score ranging from 2 to 8. Readiness to change had $\alpha = 0.46$ at baseline.

We measured how much participants expected they would be able to change their feelings of depression using the three items from the Depression Change Expectancies Scale (DCES) with the highest item-total correlations in the original study on the scale⁶⁴. Items were rated from 1 ('Strongly disagree') to 5 ('Strongly agree'), with a total score ranging from 3 to 15. The DCES had $\alpha = 0.75$ at baseline.

We evaluated how acceptable participants found their intervention with a few measures, as there is little agreement on how to measure acceptability in digital mental health interventions⁶⁵. First, the CEQ, with scores ranging from 4 to 56 (ref. 66). Some of the items are rated 1 ('Not at all') to 9 ('Very much'), and others are rated 0 (0%) to 10 (100%). The CEQ had $\alpha = 0.92$ at baseline. We also measured intervention acceptability by asking if one experienced an 'aha!' moment during the intervention (no/yes); the significance of that 'aha!' moment, 1 (not at all significant) to 5 (very significant); and the overall quality of the intervention, 1 (worst quality) to 5 (best quality).

We collected the following demographic information about participants: age, gender, race, ethnicity, education, household income, relationship/marital status, political party, employment status, disability or chronic health condition, socioeconomic status (MacArthur Social Ladder, ordinal 1–10)⁶⁷, self-identification as having depression (yes/no), number of years living with depression (continuous) and having accessed various forms of mental health support (yes/no for each form of support).

All variables mentioned in this Article were self-reported by participants. Some categorical variables' levels were determined by the researchers (for example, disability or chronic health condition) and others by CloudResearch Connect (for example, gender and race). We did not systematically measure harms to participants.

Primary hypothesis. For each SSI, we hypothesized that participants assigned to the SSI would report a different extent of change in depressive symptoms between baseline and 4-week follow-up than participants assigned to the control condition.

Primary analysis. We used a mixed-effects linear regression model to test if depressive symptoms changed to a different extent from baseline to 4-week follow-up in each of the 12 SSIs compared with the control condition. Timepoint is a two-level factor (baseline and week 4 follow-up). The model included a participant identifier as a random intercept. The dependent variable (Y_{ij}) represents the depression score for participant i at measurement timepoint j . The intercept (β_0) represents the expected depression score at baseline in the control condition. The coefficient β_1 represents the main effect of time, capturing the average change in depression scores from baseline to 4-week follow-up in the control condition. The coefficient β_2 represents the main effect of condition, capturing the average difference between the control and intervention conditions at baseline. The coefficient β_3 represents the interaction effect of time and condition: the differential change over time between the intervention and control conditions, or whether depression scores changed differently from baseline to follow-up in an intervention condition compared to the control. The term u_i represents the participant-level random intercept, accounting for individual differences in baseline depression levels, and ϵ_{ij} represents the within-person residual error, which is the variability in depression scores not explained by the fixed effects or random intercept:

$$Y_{ij} = \beta_0 + \beta_1 \times \text{timepoint}_{ij} + \beta_2 \times \text{condition}_{ij} + \beta_3 \times (\text{timepoint}_{ij} \times \text{condition}_{ij}) + u_i + \epsilon_{ij}.$$

Secondary hypotheses and analyses, sensitivity analyses and robustness checks. We preregistered six secondary hypotheses examining SSIs' effects on secondary outcomes, comparisons between SSIs and moderators of treatment effects. We also preregistered three sensitivity analyses: examining whether adding covariates to the main model changed the findings, assessing differential attrition and evaluating the effects of multiple comparisons. We also conducted several additional analyses. These analyses are presented at <https://osf.io/agvh6/files/r5kyb>.

Inference criteria. We used P value as our criterion for statistical significance, preregistering that we would report effects of $P \leq 0.05$ as significant, $0.05 < P \leq 0.1$ as marginally significant and $P > 0.1$ as non-significant. However, for simplicity's sake, we decided to eliminate the 'marginally significant' category in this Article and simply report $P < 0.05$ as significant and all else as non-significant.

We used two-sided tests for all analyses. Unless otherwise indicated, results that mention differences (for example, greater or weaker) reflect statistically significant differences. To obtain Cohen's d estimates from a binary predictor in a mixed-effects model, we divided the predictor's regression coefficient by its standard deviation at baseline, computed as the outcome's between-person variance (random intercepts) plus its within-person (residual) variance⁶⁸.

We ran mixed-effects regression models using the lme4 and lmerTest packages in R version 4.4.2^{69–71}. We visualized analyses using the sjPlot 2.8.15 and ggPlot2 3.5.1 R packages^{72,73}.

Target sample size calculation. We sought to recruit 500 participants per intervention condition and 1,500 participants for the control condition, because tripling the control sample size increased the power for analyses comparing the intervention conditions with the control. Thus, we planned to recruit 7,500 participants in all. We did not conduct interim analyses and did not preset stopping guidelines.

We preregistered this sample size to detect change from baseline to 4-week follow-up between each intervention and the control condition with an effect size of $d = 0.17$ or larger (that is, roughly a 0.78-point difference on the PHQ-9 score) with 90% power and an alpha of 0.05. This sample size also enabled the detection of differences in change from baseline to follow-up between each intervention condition of $d = 0.19$ or larger (that is, roughly a 0.88-point difference on the PHQ-9 score). Based on estimates from previous research, power analyses assumed a test-retest correlation of $r = 0.68$ for PHQ-9 score over 4 weeks⁷⁴, and that 20% of participants who completed the baseline would not complete the 4-week follow-up²⁶. Although more statistical power would have been preferable (financial constraints prevented a larger sample), the study was appropriately powered to detect some reliable effect size estimates from previous work, such as the $d = 0.18$ effect on 3-month depressive symptoms versus an attention control found in ref. 34.

Data quality. Data quality is an important concern in online data collection with anonymous participants. CloudResearch Connect and Prolific both use a range of strategies to ensure participants are providing reliable data, paying attention and not using assistive software (for example, using generative artificial intelligence tools to generate responses to questions), but these strategies are imperfect. In addition, although beginning the study multiple times was not permitted, some participants may have done so without being detected; for example, they may have created multiple accounts on one platform or completed the study on both CloudResearch Connect and Prolific.

We preregistered the exclusion criterion 'if a participant begins a survey multiple times, we will only keep data from the first time'. Responses from 299 out of 7,416 IP addresses reached randomization twice, and 10 did so three times. In total, 271 of these IP addresses were used on both CloudResearch Connect and Prolific. In addition, the

median duration between the first and second time the same IP address completed a survey was 3.91 weeks (interquartile range 3.13–4.29), so many of these participants completed two SSIs before completing the 4-week follow-up. In total, 228 (74%) of these duplicate IP addresses also had the same age, gender and race, probably reflecting one participant completing the survey multiple times. To deal with these duplicates, we dropped all but the first baseline, post-test and 4-week follow-up response per ‘likely duplicated participant’ (defined as an IP, age, gender and race combination that appeared at least twice). We also added two non-preregistered robustness checks: one that reran the analyses with the complete sample (that is, not accounting for duplicate participants) and another that removed all responses from participants with a duplicated IP address.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The deidentified data from this study are publicly available via Open Science Framework at <https://osf.io/agvh6> (Files > Data for Analyses).

Code availability

The analysis scripts, interventions and other study materials are publicly available via Open Science Framework at <https://osf.io/agvh6>.

References

- Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D., & Saxena, S. Depression: A global public health concern. *World Health Organization* <https://doi.org/10.1037/e517532013-004> (2012).
- World mental health today: latest data. *World Health Organization* <https://www.who.int/publications/i/item/9789240113817> (2025).
- Cuijpers, P. Targets and outcomes of psychotherapies for mental disorders: an overview. *World Psychiatry* **18**, 276–285 (2019).
- Kazdin, A. E. & Blase, S. L. Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspect. Psychol. Sci.* **6**, 21–37 (2011).
- Coombs, N. C., Meriwether, W. E., Caringi, J. & Newcomer, S. R. Barriers to healthcare access among U.S. adults with mental health challenges: a population-based study. *SSM - Popul. Health* **15**, 100847 (2021).
- Ramos, G., Hernandez-Ramos, R., Taylor, M. & Schueller, S. M. 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> (2024).
- Jardine, J. et al. Between rhetoric and reality: real-world barriers to uptake and early engagement in digital mental health interventions. *ACM Trans. Comput. Hum. Interact.* **31**, 27:1–27:59 (2024).
- Krukowski, R. A. et al. Digital health interventions for all? Examining inclusivity across all stages of the digital health intervention research process. *Trials* **25**, 98 (2024).
- Yardley, L. et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am. J. Prev. Med.* **51**, 833–842 (2016).
- Schleider, J., Dobias, M. L., Sung, J. Y. & Mullarkey, M. C. Future directions in single-session youth mental health interventions. *J. Clin. Child Adolesc. Psychol.* **49**, 264–278 (2020).
- Schleider, J. et al. Single-session interventions for mental health problems and service engagement: umbrella review of systematic reviews and meta-analyses. *Annu. Rev. Clin. Psychol.* <https://doi.org/10.1146/annurev-clinpsy-081423-025033> (2025).
- Schleider, J. & Beidas, R. S. Harnessing the single-session intervention approach to promote scalable implementation of evidence-based practices in healthcare. *Front. Health Serv.* **2**, 997406 (2022).
- Dobias, M. L., Morris, R. R. & Schleider, J. Single-session interventions embedded within Tumblr: acceptability, feasibility, and utility study. *JMIR Form. Res.* **6**, e39004 (2022).
- Ghosh, A., McDanal, R. & Schleider, J. L. Digital single-session interventions for child and adolescent mental health: evidence and potential for dissemination across low- and middle-income countries. *Adv. Psychiatry Behav. Health* **3**, 129–138 (2023).
- Schleider, J. & Weisz, J. R. Can less be more? The promise (and perils) of single-session youth mental health interventions. *Behav. Ther.* **40**, 256–261 (2017).
- Ball, J., Thompson, Z., Meiser-Stedman, R. & Chiu, K. Self-Administered single session interventions for mental health in young people: a systematic review and meta-analysis. *SSRN* <https://doi.org/10.2139/ssrn.4856261> (2024).
- Lorenzo-Luaces, L. et al. A pragmatic randomized controlled trial of stepped care cognitive-behavioral therapy for internalizing distress in adults. *Cogn. Ther. Res.* **48**, 998–1013 (2024).
- Lorenzo-Luaces, L. & Howard, J. Efficacy of an unguided, digital single-session intervention for internalizing symptoms in web-based workers: randomized controlled trial. *J. Med. Internet Res.* **25**, e45411 (2023).
- Mullarkey, M. et al. Web-based single session intervention for perceived control over anxiety during COVID-19: randomized controlled trial. *JMIR Ment. Health* **9**, e33473 (2022).
- Dodge, K. A. et al. Population mental health science: guiding principles and initial agenda. *Am. Psychol.* **79**, 805–823 (2024).
- Prentice, D. A. & Miller, D. T. When small effects are impressive. *Psychol. Bull.* **112**, 160–164 (1992).
- Manvelian, A. et al. A practical guide to adapting evidence-based treatments into digital single-session interventions for youth. Preprint at https://doi.org/10.31234/osf.io/vjgxr_v1 (2025).
- Schleider, J., Abel, M. R. & Weisz, J. R. Do immediate gains predict long-term symptom change? Findings from a randomized trial of a single-session intervention for youth anxiety and depression. *Child Psychiatry Hum. Dev.* **50**, 868–881 (2019).
- Cohen, K., Dobias, M., Morris, R. & Schleider, J. Improving uptake of mental health crisis resources: randomized test of a single-session intervention embedded in social media. *J. Behav. Cogn. Ther.* **33**, 24–34 (2023).
- Osborn, T. L. et al. Single-session digital intervention for adolescent depression, anxiety, and well-being: Outcomes of a randomized controlled trial with Kenyan adolescents. *J. Consult. Clin. Psychol.* **88**, 657–668 (2020).
- Kaveladze, B., Ghosh, A., Funkhouser, C. J., Schueller, S. M. & Schleider, J. L. Longer single-session interventions may not be better: evidence from two randomized controlled trials with online workers facing mental-health struggles. *Clin. Psychol. Sci.* <https://doi.org/10.1177/21677026251358836> (2025).
- Yeager, D. S. & Dweck, C. S. Mindsets that promote resilience: when students believe that personal characteristics can be developed. *Educ. Psychol.* **47**, 302–314 (2012).
- Gualano, M. R. et al. The long-term effects of bibliotherapy in depression treatment: systematic review of randomized clinical trials. *Clin. Psychol. Rev.* **58**, 49–58 (2017).
- Schleider, J., Dobias, M., Sung, J., Mumper, E. & Mullarkey, M. C. Acceptability and utility of an open-access, online single-session intervention platform for adolescent mental health. *JMIR Ment. Health* **7**, e20513 (2020).
- Ryan, R. M. & Deci, E. L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **55**, 68–78 (2000).
- Voelkel, J. G. et al. Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity. *Science* **386**, eadh4764 (2024).

32. Milkman, K. L. et al. Megastudies improve the impact of applied behavioural science. *Nature* **600**, 478–483 (2021).
33. Milkman, K. L. et al. A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proc. Natl Acad. Sci. USA* **119**, e2115126119 (2022).
34. Schleider, J. et al. A randomized trial of online single-session interventions for adolescent depression during COVID-19. *Nat. Hum. Behav.* **6**, 258–268 (2022).
35. Protzko, J. Invariance: what does measurement invariance allow us to claim?. *Educ. Psychol. Meas.* **85**, 458–482 (2025).
36. Hlynsson, J. I., Skúlason, S., Andersson, G. & Carlbring, P. Why are we still using the PHQ-9? A historical review and psychometric evaluation of measurement invariance. *Psychiatr. Q.* <https://doi.org/10.1007/s1126-025-10208-9> (2025).
37. Luong, R. & Flake, J. K. Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychol. Methods* **28**, 905–924 (2023).
38. Lee, M. R. & Shen, M. Winner's curse: bias estimation for total effects of features in online controlled experiments. In *Proc. 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 491–499 (Association for Computing Machinery, 2018).
39. Plessen, C. Y., Panagiotopoulou, O. M., Tong, L., Cuijpers, P. & Karyotaki, E. Digital mental health interventions for the treatment of depression: a multiverse meta-analysis. *J. Affect. Disord.* **369**, 1031–1044 (2025).
40. Steinberg, J. S. et al. Randomized trial testing a self-guided digital mental health intervention teaching calming skills for Ukrainian children. *NPJ Ment. Health Res.* **4**, 20 (2025).
41. Morris, R. R., Schueller, S. M. & Picard, R. W. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. *J. Med. Internet Res.* **17**, e72 (2015).
42. Mrazek, A. J. et al. Attention training improves the self-reported focus and emotional regulation of high school students. *Technol. Mind Behav.* **3**, 4 (2022).
43. Kumarasamy, V., Goodfellow, N., Ferron, E. M. & Wright, A. L. Evaluating the problem of fraudulent participants in health care research: multimethod pilot study. *JMIR Form. Res.* **8**, e51530 (2024).
44. Douglas, B. D., Ewell, P. J. & Brauer, M. Data quality in online human-subjects research: comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS ONE* **18**, e0279720 (2023).
45. Chapkovski, P., Peer, E. & Brandt, E. From WEIRD to GREAT? Exploring data quality of global platforms for online research. Preprint at <https://doi.org/10.31234/osf.io/xvfnq> (2024).
46. Protzko, J. The Online Volunteer Subject. Preprint at <https://doi.org/10.31234/osf.io/gh2ck> (2022).
47. Bryan, C. J., Tipton, E. & Yeager, D. S. Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nat. Hum. Behav.* **5**, 980–989 (2021).
48. Baumel, A., Muench, F., Edan, S. & Kane, J. M. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *J. Med. Internet Res.* **21**, e14567 (2019).
49. Nafziger, J. Spillover effects of nudges. *Econ. Lett.* **190**, 109086 (2020).
50. Arnautovska, U. & Milton, A. The era of digital mental health interventions: we know they can be effective but are they also safe?. *BJPsych Open* **11**, e89 (2025).
51. Garrey, S. K., Takemoto, E., Petrsoric, L. & Gargano, L. M. Self-Efficacy and mental health help-seeking behavior of World Trade Center Health Registry enrollees, 2015–2016. *Int. J. Environ. Res. Public Health* **19**, 7113 (2022).
52. McDermott, R. C. et al. Hope for help-seeking: a positive psychology perspective of psychological help-seeking intentions. *Couns. Psychol.* **45**, 237–265 (2017).
53. Rizvi, S. L. et al. Randomized clinical trial of a brief, scalable intervention for mental health sequelae in college students during the COVID-19 pandemic. *Behav. Res. Ther.* **149**, 104015 (2022).
54. Wakefield, S. et al. Improving Access to Psychological Therapies (IAPT) in the United Kingdom: a systematic review and meta-analysis of 10-years of practice-based evidence. *Br. J. Clin. Psychol.* **60**, e12259 (2021).
55. Kaveladze, B. T. *Using Brief Internet Interventions to Challenge Loneliness at Scale* (UC Irvine, 2023).
56. Groot, J. et al. The effectiveness of fully automated digital interventions in promoting mental well-being in the general population: systematic review and meta-analysis. *JMIR Ment. Health* **10**, e44658 (2023).
57. Hartman, R. et al. Introducing Connect by CloudResearch: advancing online participant recruitment in the digital age. Preprint at <https://doi.org/10.31234/osf.io/ksgyr> (2023).
58. Palan, S. & Schitter, C. Prolific.ac—a subject pool for online experiments. *J. Behav. Exp. Finance* **17**, 22–27 (2018).
59. Qualtrics (Qualtrics, 2005).
60. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
61. Terides, M. D. et al. The frequency of actions and thoughts scale: development and psychometric validation of a measure of adaptive behaviours and cognitions. *Cogn. Behav. Ther.* **45**, 196–216 (2016).
62. Snyder, C. R. et al. Development and validation of the State Hope Scale. *J. Pers. Soc. Psychol.* **70**, 321–335 (1996).
63. Perczel Forintos, D., Rózsa, S., Pilling, J. & Kopp, M. Proposal for a short version of the Beck Hopelessness Scale based on a national representative survey in Hungary. *Community Ment. Health J.* **49**, 822–830 (2013).
64. Eddington, K. M., Dozois, D. J. A. & Barks-Dermott, B. J. Evaluation of the internal consistency, factor structure, and validity of the Depression Change Expectancy Scale. *Assessment* **21**, 607–617 (2014).
65. Perski, O. & Short, C. E. Acceptability of digital health interventions: embracing the complexity. *Transl. Behav. Med.* **11**, 1473–1480 (2021).
66. Devilly, G. J. & Borkovec, T. D. Psychometric properties of the credibility/expectancy questionnaire. *J. Behav. Ther. Exp. Psychiatry* **31**, 73–86 (2000).
67. Adler, N. E., Epel, E. S., Castellazzo, G. & Ickovics, J. R. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychol.* **19**, 586–592 (2000).
68. Feingold, A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychol. Methods* **14**, 43–53 (2009).
69. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
70. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).
71. R Core Team. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2015).
72. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
73. Lüdtke, D. et al. sjPlot: Data visualization for statistics in social science. *Zenodo* <https://doi.org/10.5281/zenodo.2400856> (2018).
74. Sun, Y. et al. The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital. *BMC Psychiatry* **20**, 474 (2020).

75. Bhattacharjee, A. et al. Perfectly to a tee: understanding user perceptions of personalized LLM-enhanced narrative interventions. In *Proc. 2025 ACM Designing Interactive Systems Conference* 1387–1416 (ACM, 2025).
76. McGuire, A. P., Howard, B. A. N., Erickson, T. M. & Creech, S. K. Moral Elevation Online Intervention for Veterans Experiencing Distress Related to Posttraumatic Stress Disorder and Moral Injury (MOVED): pilot trial of a 4-week positive psychology web-based intervention. *JMIR Form. Res.* **7**, e39894 (2023).
77. Huberty, J. L., Espel-Huynh, H. M., Neher, T. L. & Puzia, M. E. Testing the pragmatic effectiveness of a consumer-based mindfulness mobile app in the workplace: randomized controlled trial. *JMIR MHealth UHealth* **10**, e38903 (2022).

Acknowledgements

We thank all members of the teams who submitted their interventions to this megastudy for participating in this team effort and trusting in our research team. N. B. Allen contributed to Personalized Intervention Recommender. P. Patrzek and S. Nakamura contributed to Inner Child Healing Walk. This work was funded by the National Institute of Mental Health (T32 MH115882; B.T.K.), a US\$10,000 donation to the Lab for Scalable Mental Health from Calm Health and a US\$2,500 CloudResearch Innovations in Online Research grant (to be spent only on data collection through CloudResearch Connect). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript, except that Calm Health's SSI was tested in the megastudy (refer to Methods) and CloudResearch Connect was used to recruit and compensate participants.

Author contributions

B.T.K. was responsible for the study's conception and design, data acquisition and analysis and interpretation and drafting of the paper. J.G.V., M.N.S., J.P., L.L.L., S.M.S., M.K.N. and J.L.S. contributed to the study's design and interpretation, and made substantial revisions to the paper. M.H., A.E.S., E.K.S., Y.M.X., M.P.M. and J.P.Z. contributed to organizing the intervention crowdsourcing effort and selecting submitted interventions. S.I.A., A.B., I.G., R.H.-R., K.S.H., J.K.J., A.C.K., R.S.M.K., R.K., M.N.L., M.L., M.A.L., A.T.M., A.M.M., A.P.M., J.M., A.J.M., M.D.M., D.C.M., R.R.M., C.J.M., H.N., A.O.-U., J.R.P., D.S.R., S.L.R., M.W.S., S.E.S., M.E.T., A.R.T., C.V.T., A.N.T., J.J.W., K.E.W. and E.J.W. developed the interventions included for testing in the megastudy. All authors approved the submitted paper.

Competing interests

The authors declare the following competing interests. Virtually everyone who contributed an intervention that was tested—S.I.A., A.B., I.G., R.H.-R., K.S.H., J.K.J., A.C.K., R.S.M.K., R.K., M.N.L., M.L., M.A.L., A.T.M., A.M.M., A.P.M., J.M., A.J.M., M.D.M., D.C.M., R.R.M., C.J.M., H.N., A.O.-U., J.R.P., D.S.R., S.L.R., M.W.S., S.E.S., M.E.T., A.R.T., C.V.T., A.N.T. and J.J.W.—probably had a competing interest in that they wanted their intervention to perform well in the experiment and wanted to present the results in a positive light for their intervention. To account for these competing interests, the core authorship team did not accept

input from intervention contributors on the experimental design and primary and secondary preregistered outcomes. Submitters were permitted to comment on a draft of the final paper and suggest edits via a publicly viewable online document that all authors could see. The Lab for Scalable Mental Health received a US\$10,000 donation from Calm Health and used all of it to fund the megastudy experiment. Given this financial support and the possibility of future support, members of the Lab for Scalable Mental Health (B.T.K., M.H., A.E.S. and J.P.Z.) may have been motivated to present findings regarding Calm Health's SSI Reframing Negative Thoughts in a positive light. J.M. has accepted consulting fees from Boehringer Ingelheim and Shirley Ryan Ability Lab. M.K.N. receives publication royalties from Macmillan, Pearson and UpToDate. He has been a paid consultant in the past 3 years for Cambridge Health Alliance, and for legal cases regarding a death by suicide. He has stock options in Cerebral Inc. He is an unpaid scientific advisor for Empatica, Koko, TalkLife and the JED Foundation. S.L.R. earns ad revenue from YouTube. L.L.L. was supported by NIH grant R01MH135502. L.L.L. has received consulting fees from Syra Health, Inc., who had no involvement in the current project. S.M.S. has received consulting payments from Boehringer Ingelheim and Otsuka Pharmaceuticals and is a member of the Headspace Scientific Advisory Board, for which he receives compensation. J.L.S. has served on the Scientific Advisory Board for Walden Wise and the Clinical Advisory Board for Koko (unpaid) and receives book royalties from New Harbinger; Oxford University Press; and Little, Brown Book Group. She is cofounder and chief scientific advisor for Navi. No Navi products were used or are referenced in this paper. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-026-02415-6>.

Correspondence and requests for materials should be addressed to Benjamin T. Kaveladze.

Peer review information *Nature Human Behaviour* thanks Korena S. Klimczak and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2026

Benjamin T. Kaveladze¹✉, Jan G. Voelkel², Michael N. Stagnaro³, Mingjing Huang¹, Amanda E. Smock¹, Erin K. Sullivan⁴, Yao M. Xu⁵, Madison P. McCall⁶, Juan Pablo Zapata¹, Syed Ishtiaque Ahmed⁷, Ananya Bhattacharjee⁷, Iva Georgieva⁸, Rosa Hernandez-Ramos⁹, Kenneth S. Huber¹⁰, Julia K. Jennings¹¹, Arianna C. Kirk¹², Robert S. M. Knowles¹³, Rachel Kornfield¹, Monika Neff Lind⁹, Michelle Liu¹⁴, Michael A. Liut¹⁵, Alex T. Mariakakis⁷, Ali M. Mattu¹⁶, Adam P. McGuire¹⁷, Jonah Meyerhoff¹, Alissa J. Mrazek¹¹, Michael D. Mrazek¹¹, David C. Mohr¹, Robert R. Morris¹³, Christopher J. Mosunic¹⁰, Hana Nip⁹, Allen Olson-Urtecho¹⁸, Jessa R. Podell¹¹, Dustyn S. Ransom¹¹, Shireen L. Rizvi¹⁹,

**Matthew W. Southward²⁰, Sarah Elizabeth Stoeckl⁹, Madison E. Taylor⁹, Amy R. Texter²¹, Calvin V. Tower¹¹,
Angelique N. Trotter¹¹, Joseph J. Williams⁷, Katherine E. Wislocki⁹, Elijah J. Woodson¹¹, John Protzko⁴,
Lorenzo Lorenzo-Luaces²², Stephen M. Schueller⁹, Matthew K. Nock²³ & Jessica L. Schleider¹**

¹Northwestern University, Chicago, IL, USA. ²Cornell University, Ithaca, NY, USA. ³Yale University, New Haven, CT, USA. ⁴Central Connecticut State University, New Britain, CT, USA. ⁵Stanford University, Palo Alto, CA, USA. ⁶University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁷University of Toronto, Toronto, Ontario, Canada. ⁸Institute of Philosophy and Sociology, Bulgarian Academy of Sciences, Sofia, Bulgaria. ⁹University of California, Irvine, Irvine, CA, USA. ¹⁰Calm Health, Palo Alto, CA, USA. ¹¹University of Texas at Austin, Austin, TX, USA. ¹²Visionari Media, Santa Barbara, CA, USA. ¹³Koko, San Francisco, CA, USA. ¹⁴Loyola University Chicago, Chicago, IL, USA. ¹⁵University of Toronto Mississauga, Mississauga, Ontario, Canada. ¹⁶Mind Stuff LLC, Oakland, CA, USA. ¹⁷University of Texas at Tyler, Tyler, TX, USA. ¹⁸Studio Bahia, Houston, TX, USA. ¹⁹Albert Einstein College of Medicine, New York, NY, USA. ²⁰The Ohio State University, Columbus, OH, USA. ²¹University of Colorado Boulder, Boulder, CO, USA. ²²Indiana University Bloomington, Bloomington, IN, USA. ²³Harvard University, Cambridge, MA, USA. ✉e-mail: bkavelad@uci.edu

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|---|
| Data collection | We used Qualtrics, CloudResearch Connect, and Prolific for data collection (March to June 2025). |
| Data analysis | We ran mixed-effects regression models using the lme4 and lmerTest packages in R version 4.4.2 (Bates et al., 2015; Kuznetsova et al., 2017; R Core Team, 2015). We visualized analyses using the sjPlot 2.8.15 and ggPlot2 3.5.1 R packages (Lüdtke et al., 2021; Wickham et al., 2019). The analysis script is available at https://osf.io/agvh6/files/42xqa |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

A de-identified version of the data is publicly available at <https://osf.io/agvh6>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We did not pre-register or report sex- and gender-based analyses because we wanted to focus this manuscript on the main analyses comparing outcomes across interventions in the full sample. Besides aiming to recruit a roughly even number of men and women, gender was not a major consideration in our study design.
Reporting on race, ethnicity, or other socially relevant groupings	We collected the following socially relevant categorization variables to characterize the sample and to use as demographic covariates in sensitivity analyses: age, gender, race, ethnicity, education, household income, relationship/marital status, political party, employment status, disability or chronic health condition, socioeconomic status (MacArthur Social Ladder), self-identification as having depression, number of years living with depression, and having accessed various forms of mental health support. All variables were self-reported by participants. Some categorical variables' levels were determined by the researchers based on categories used in previous studies (e.g., disability or chronic health condition), and other variables' levels were determined by CloudResearch Connect because the platform collected that data for each participant prior to the study (e.g., gender and race). Confounding variables were not a major concern for this study as it was a randomized controlled trial; however, our sensitivity analyses included covariates that may help to address concerns around confounding.
Population characteristics	See above / below
Recruitment	<p>Participants were American adults experiencing elevated depressive symptoms. They were recruited from the online participant recruitment platforms CloudResearch Connect and Prolific. This choice of sample introduced some biases. First, participants were paid online workers, who may differ from the broader populations SSIs aim to serve. For example, many participants were likely not actively seeking mental health support when they found the study, so we did not reach them at moments of need when SSIs might be more helpful. Overall, caution is needed when generalizing our results beyond the population sampled: nonprobability opt-in panels of adult online workers experiencing elevated (mostly moderately elevated) depressive symptoms.</p> <p>Participants were compensated after completing each study session: \$0.25 for screening (1 min), \$3.00 for the first session (18 min), and \$1.00 for the four-week follow-up (5 min).</p>
Ethics oversight	The procedures were approved by the Northwestern University Social and Behavioral Sciences Review Board (Protocol STU00220193). The studies were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Informed consent was obtained from all research participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This was a randomized controlled trial collecting mostly quantitative data. It used an experimental design with parallel assignment and a superiority framework.
Research sample	<p>The sample was composed of American online workers 18 years or older with moderate or more severe depressive symptoms (measured using the PHQ-9 at screening), recruited online. The sample's mean age was 36.6 years (SD = 11.8), 59% were women and 36% were men, 74% were White and 14% were Black, and 60% had received some form of mental health support in the past month (full demographic information are available at https://osf.io/jzbnq). The sample was not representative. We chose this sample because we were interested in evaluating single-session interventions' efficacy among a general sample of adult online workers. A more nationally representative sample was not feasible due to financial constraints.</p> <p>As shown in the CONSORT diagram (https://osf.io/agvh6/files/b7c86), 83 individuals declined participation (i.e., did not complete the consent form), and 1507 participants did not complete the four-week follow-up.</p>
Sampling strategy	<p>We recruited the sample through convenience sampling -- posting study advertisements on CloudResearch Connect and Prolific. We pre-registered and determined sample size before data collection, using an online tool (https://powerimmjs.rpsychologist.com).</p> <p>We aimed to recruit 500 participants per intervention condition and 1,500 for the control condition (because tripling the sample size in the control improved the power we could generate from the same number of participants in analyses comparing the intervention conditions to the control). Thus, we planned to recruit 7,500 participants in all.</p>

We pre-registered this sample size to detect change from baseline to four-week follow-up between each intervention and the control condition with an effect size of $d = 0.17$ or larger (i.e., roughly a 0.78-point difference on the PHQ-9 score) with 90% power and $\alpha = 0.05$. This sample size also enabled the detection of differences in change from baseline to follow-up between each intervention condition of $d = 0.19$ or larger (i.e., roughly a 0.88-point difference on the PHQ-9 score). Based on estimates from previous research, our power analysis assumed a test-retest correlation of $r = 0.68$ for PHQ-9 score over four weeks (Sun et al., 2020), and that 20% of participants who completed the baseline would not complete the four-week follow-up (Kaveladze et al., 2025).

Data collection

All data were collected through the online survey and experience management platform Qualtrics, via users' personal computers, tablets, or smartphones (Qualtrics, 2005). All study procedures took place online, with no synchronous interactions between participants and researchers.

Timing

Data collection occurred from March to June 2025.

Data exclusions

Data exclusions are detailed in the CONSORT diagram (<https://osf.io/b7c86>). We pre-registered the exclusion criterion, "if a participant begins a survey multiple times, we will only keep data from the first time". Responses from 299 out of 7,416 IP addresses reached randomization twice and 10 did so three times. 271 of these IP addresses were used on both CloudResearch Connect and Prolific. In addition, the median duration between the first and second time the same IP address completed a survey was 3.91 weeks (IQR 3.13 - 4.29), so many of these participants completed two SSIs before completing the four-week follow-up. 228 (74%) of these duplicate IP addresses also had the same age, gender, and race, likely reflecting one participant completing the survey multiple times. To deal with these duplicates, we dropped all but the first baseline, post-test and four-week follow-up response per likely duplicated participant (defined as an IP, age, gender, and race combination that appeared at least twice). We also added two non-pre-registered robustness checks: one that re-ran the analyses with the complete sample (i.e., not accounting for duplicate participants) and another that removed all responses from participants with a duplicated IP address.

As shown in the CONSORT diagram (<https://osf.io/agvh6/files/b7c86>), 83 individuals did not decline participation (i.e., did not complete the consent form), and 1507 participants did not complete the four-week follow-up.

Non-participation

Participant dropout is detailed in the CONSORT diagram (<https://osf.io/b7c86>)

Randomization

Participant randomization was automated through Qualtrics' randomization function.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern
- Plants

Methods

- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Study protocol

Data collection

All data were collected through the online survey and experience management platform Qualtrics, via users' personal computers, tablets, or smartphones (Qualtrics, 2005). All study procedures took place online, with no synchronous interactions between participants and researchers.

Outcomes

We pre-registered all outcomes prior to data collection. All outcomes were self-reported through the online survey platform Qualtrics.

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>