Comparative effectiveness of remote digital gamified and group CBT skills training interventions for anxiety and depression among college students

Bantjes, Jason; Hunt, Xanthe; Cuijpers, Pim; Kazdin, Alan E.; Kennedy, Chris J.; Luedtke, Alex; Malenica, Ivana; Petukhova, Maria; Sampson, Nancy; Zainal, Nur Hani; Davids, Charl; Dunn-Coetzee, Munita; Gerber, Rone; Stein, Dan J.; Kessler, Ronald C.

published in
Behaviour Research and Therapy
2024

DOI (link to publisher)
10.1016/j.brat.2024.104554

document version
Publisher's PDF, also known as Version of record

Behaviour Research and Therapy 178 (2024) 104554

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

Comparative effectiveness of remote digital gamified and group CBT skills training interventions for anxiety and depression among college students: Results of a three-arm randomised controlled trial

Jason Bantjes [a,b,*], Xanthe Hunt [c], Pim Cuijpers [d], Alan E. Kazdin [e], Chris J. Kennedy [f], Alex Luedtke [g,h], Ivana Malenica [i], Maria Petukhova [j], Nancy Sampson [j], Nur Hani Zainal [j], Charl Davids [k], Munita Dunn-Coetzee [l], Rone Gerber [m], Dan J. Stein [b,n], Ronald C. Kessler [j]

[a] Mental Health, Alcohol, Substance Use and Tobacco (MAST) Research Unit, South African Medical Research Council, Cape Town, South Africa
[b] Department of Psychiatry and Mental Health, University of Cape Town, South Africa
[c] Department of Global Health, Institute for Life Course Health Research, Stellenbosch University, South Africa
[d] Department of Clinical, Neuro-, and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands
[e] Department of Psychology, Yale University, New Haven, CT, USA
[f] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[g] Department of Statistics, University of Washington, Seattle, WA, USA
[h] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[i] Department of Statistics, Harvard University, Cambridge, MA, USA
[j] Department of Health Care Policy, Harvard Medical School, Boston, MA, USA
[k] Centre for Student Counselling and Development, Student Affairs, Stellenbosch University, South Africa
[l] Student Counselling and Development, University of the Free State, South Africa
[m] Student Counselling and Development, University of Western Cape, Cape Town, South Africa
[n] Department of Psychiatry, SAMRC Unit on Risk and Resilience in Mental Disorders, Stellenbosch University, South Africa

A R T I C L E I N F O

Keywords:
Digital interventions
Online cognitive behavioral therapy (CBT)
Randomised controlled trial
South Africa
SuperBetter
University students

A B S T R A C T

Digital interventions can enhance access to healthcare in under-resourced settings. However, guided digital interventions may be costly for low- and middle-income countries, despite their effectiveness. In this randomised control trial, we evaluated the effectiveness of two digital interventions designed to address this issue: (1) a Cognitive Behavioral Therapy Skills Training (CST) intervention that increased scalability by using remote online group administration; and (2) the SuperBetter gamified self-guided CBT skills training app, which uses other participants rather than paid staff as guides. The study was implemented among anxious and/or depressed South African undergraduates (n = 371) randomised with equal allocation to Remote Group CST, SuperBetter, or a MoodFlow mood monitoring control. Symptoms were assessed with the Generalized Anxiety Disorder-7 (GAD-7) and the Patient Health Questionnaire-9 (PHQ-9). Intention-to-treat analysis found effect sizes at the high end of prior digital intervention trials, including significantly higher adjusted risk differences (*ARD*; primary outcome) in joint anxiety/depression remission at 3-months and 6-months for Remote Group CST (*ARD* = 23.3–18.9%, p = 0.001–0.035) and SuperBetter (*ARD* = 12.7–22.2%, p = 0.047–0.006) than MoodFlow and mean combined PHQ-9/GAD-7 scores (secondary outcome) significantly lower for Remote Group CST and SuperBetter than MoodFlow. These results illustrate how innovative delivery methods can increase the scalability of standard one-on-one guided digital interventions.
Preregistration international standard randomised controlled trial number (isrtcn) submission #: 47,089,643.

* Corresponding author. South African Medical Research Council, Mental health, Alcohol, Substance use and Tobacco (MAST) Research Unit, PO Box 19070, Tygerberg, 7505, Cape Town, South Africa.
E-mail address: Jason.Bantjes@mrc.ac.za (J. Bantjes).

https://doi.org/10.1016/j.brat.2024.104554
Received 3 October 2023; Received in revised form 19 April 2024; Accepted 1 May 2024
Available online 3 May 2024

1. Background

Meta-analyses of randomized controlled trials (RCTs) show clearly that digital interventions can be as effective as face-to-face psychotherapy in treating symptoms of anxiety (Cuijpers, Noma, Karyotaki, Cipriani, & Furukawa, 2019; Esfandiari, Mazaheri, Akbari-Zardkhaneh, Sadeghi-Firoozabadi, & Cheraghi, 2021; Papola et al., 2023) and depression (Carlbring, Andersson, Cuijpers, Riper, & Hedman-Lagerlöf, 2018). This is of special importance for university students in South Africa, where anxiety and depression are common (Bantjes, Kessler, Lochner, et al., 2023) and access to conventional treatment is low (Bantjes, Kessler, Hunt, Stein, & Kessler, 2023). South African university students are comfortable with digital interventions and have high access to digital technologies. And digital interventions are as effective in low- and middle-income countries such as South Africa as in high-income countries (Kim et al., 2023). However, most evidence for digital intervention effectiveness is limited to guided digital interventions in which trained staff provide feedback to users (Karyotaki et al., 2018). Evidence is much weaker for the effectiveness of self-guided digital interventions; that is, interventions delivered without using paid staff to guide users through the interventions (Karyotaki et al., 2017).

Guided digital interventions, although less expensive than conventional psychotherapy, require resources that exceed those available in South Africa to hire, train, and supervise guides. To illustrate: A recent mental health needs assessment survey of students in South African public universities estimated that only about one-fifth of the undergraduate students with clinically significant mental health problems receive any treatment, leading to at least 300,000 students having unmet need for treatment of anxiety and depression (Bantjes, Kessler, Hunt, et al., 2023). Given that intervention guides spent an average of 20 min with each patient per week, roughly 70 full-time guides or supervisors would be needed to provide guided digital therapy to all these students. This level of staffing is unrealistic even in a relatively affluent middle-income country like South Africa.

At least two possibilities exist to address this problem. One is to develop group versions of guided digital interventions as an alternative to the existing one-on-one versions, thereby increasing the number of patients each intervention worker can help. A good example is the transdiagnostic Remote Group Cognitive Behavioral Therapy Skills Training (CST) intervention developed by our group for anxious and depressed university students (Bantjes et al., 2021) as part of the World Mental Health International College Student (WMH-ICS) initiative (Cuijpers, Auerbach, et al., 2019). Remote Group CST is delivered in real-time to online groups of 9–12 participants by master's-level therapists in a live remote online group format, each session lasting 50 min. The intervention integrates effective components from several previously reported in-person group CBT interventions for anxious/depressed university students (Hamamci, 2006; Peden, Hall, Rayens, & Beebe, 2000; Takagaki et al., 2016; Yang, Zhao, Chen, Zu, & Zhao, 2018; Zemestani, Davoodi, Honarmand, Zargar, & Ottaviani, 2016). Participants are given a workbook to take notes during the group meetings and to refer to between meetings. A previous open trial showed that this Remote Group CST had high adherence and acceptability among anxious and depressed South African university students and that students experienced significant symptom reduction over the course of the intervention (Bantjes et al., 2021).

The other possibility is to increase self-guided digital intervention effectiveness by addressing the problem of adherence (Mohr, Cuijpers, & Lehman, 2011), which is usually (Richards & Richardson, 2012), although not always (Li et al., 2022), much lower in self-guided than guided interventions and is thought to account for the low effectiveness of self-guided digital interventions given that adherence is associated equivalently with therapeutic gains in self-guided and guided interventions (Gan, McGillivray, Han, Christensen, & Torok, 2021). Reviews of the literature suggest that self-guided digital intervention adherence among young people can be increased by using game-like features and leveraging peer feedback to encourage adherence in the same way as with paid guides (Ryan, Bergin, & Wells, 2018; Saleem et al., 2021). A good example of such an approach is the transdiagnostic self-guided gamified SuperBetter app (McGonigal, 2016). SuperBetter uses a myriad of CBT and positive psychology principles and components, but the main organizing components are behavioral activation with goal setting, cognitive restructuring and mindfulness training, engagement in pleasant and valued activities, exposure therapy, per support enhancement, psychological distancing, and tailored CBT action planning, and delivered using peer support to enhancement intervention engagement. SuperBetter has been evaluated in two randomized controlled trials, both of which found it to reduce anxiety and depression significantly more than a control condition (Marshall, Dunstan, & Bartik, 2021; Roepke et al., 2015). In addition, meta-analyses found that SuperBetter had one of the largest effect sizes of any online or app-based self-guided intervention evaluated in controlled treatment trials for anxiety (Firth, Torous, Nicholas, Carney, Rosenbaum, et al., 2017; Linardon, Cuijpers, Carlbring, Messer, & Fuller-Tyszkiewicz, 2019) or depression (Firth, Torous, Nicholas, Carney, Pratap, et al., 2017; Linardon et al., 2019; Serrano-Ripoll, Zamanillo-Campos, Fiol-DeRoque, Castro, & Ricci-Cabello, 2022; Weisel et al., 2019).

Here we present the results of an RCT designed to evaluate the aggregate comparative effectiveness of Remote Group CST and SuperBetter relative to an active control group among anxious and/or depressed university students in South Africa. None of the authors has any financial interest in any of these interventions.

2. Methods

2.1. Sample

We preregistered this RCT before recruitment (International Standard Randomised Controlled Trial Number (ISRCTN) Submission #47089643). The RCT was carried out in three university campuses in South Africa between May 1, 2022 and May 31, 2023. Recruitment occurred via an email invitation sent to a random sample of students. The email explained that the university's student mental health clinic was participating in a research study in which students experiencing anxiety and/or depression would be provided with one of several proven digital interventions to determine which one is most effective for which students. Interested students were directed to complete a baseline self-report assessment via a link embedded in the email. The email also included contact details in case students had questions before deciding whether to complete the assessment. The assessment included standard anxiety (Generalized Anxiety Disorder-7 [GAD-7]; Spitzer, Kroenke, Williams, & Löwe, 2006) and depression (Patient Health Questionnaire-9 [PHQ-9]; Kroenke, Spitzer, & Williams, 2001) screening scales in addition to a wide range of questions designed to predict which intervention would be best for which students (Supplementary Table S1). Inclusion criteria were at least moderate GAD-7 and/or PHQ-9 ($\geq$10) scores, consent to be randomized, internet access, and provision of a valid email address and telephone number. A total of $n = 371$ students met these inclusion criteria and were randomised with equal allocation across the three intervention arms with blocking by the cross-classification of baseline severity of anxiety (severe, moderate, mild/minimal) and depression (severe/moderately severe, moderate, mild/minimal), noting that at least one of these scores needed to be at least moderate to qualify for randomization. Randomization was carried out by a computer using balanced starting values and systematic selection within stratification cells. No incentives were provided to enrol in the study or participate in the intervention.

2.2. Recruitment procedures

The CONSORT recruitment flow diagram is depicted in Supplementary Fig. S1, and the CONSORT checklist (Schulz KF et al., 2010) is

in Supplementary Checklist S1. After completing the online assessment, eligible students randomized to Remote Group CST received a follow-up email directing them to a webpage with information about the group. The webpage noted that the group had a 10-week rolling format (i.e., content repeated in a rolling fashion every 10 weeks, allowing participants to enter at any time and stay for the next 10 weeks to cover all the content), provided a brief overview of each facilitator (name, biographic description, and picture), and asked the student to fill out a booking form to sign up for the group and facilitator of their choice. Once the form was filled out, a 15-min one-on-one online onboarding meeting was scheduled with the facilitator to orient the student to the group and answer any remaining questions before giving the student details about how to join the group.

A different follow-up email was sent to students randomized to the other two intervention arms. This email described the intervention, provided instructions on how to install the app along with details about how to contact a support person to assist with technical problems, and instructed the student to sign up for a brief (15-min) online onboarding call with a project facilitator before beginning the intervention.

A single reminder email was sent to students who did not sign up for their onboarding session. An additional email with instructions for rebooking was sent to students who signed up but failed to attend their onboarding session. Follow-up online self-report assessments were conducted 3 months and 6 months after onboarding. No attempt was made to administer follow-up assessments to randomized students who were not onboarded.

2.3. Interventions

As noted above, the two active interventions were Remote Group CST (Bantjes et al., 2021) and SuperBetter (McGonigal, 2016). The control group was mood monitoring implemented in the MoodFlow app (Moodflow, 2023). These interventions differ in critical respects (Table 1). Remote Group CST is a 10-week transdiagnostic CBT skills training intervention delivered synchronously online via video conferencing platform by trained master's-level psychologists. The focus is on five core CBT themes: Addressing Core Beliefs, Behavioral Activation, Cognitive Restructuring, Goal Setting and Problem-Solving, and Relaxation Training and Exposure Therapy. We offered two sessions per theme in sequence in recognition of students' diverse commitments and anticipating partial attendance. A full description of the group intervention is included as supplementary material. No homework assignments or pre-workshop reading was required although participants were encouraged to refer to the workbook as needed and practice the skills they learnt between sessions. All sessions were recorded with the permission of participants and spot reviewed by supervisors to guarantee fidelity. Only supervisors had access to these recordings. Weekly group supervision sessions with the group therapists reviewed recordings and reinforced initial trainings to prevent drift. SuperBetter is also a transdiagnostic CBT skills training intervention presented in an app that users are instructed to use for at least 10 min per day, 5 days per week for at least 10 weeks. The main components are behavioral activation with goal setting, cognitive restructuring and mindfulness training, engagement in pleasant and valued activities, and exposure therapy, all using a peer support game format to enhance engagement. MoodFlow, finally, is a widely used mood monitoring app that was selected as the control condition because mood monitoring is consistently found to have positive but modest and temporary aggregate effects on anxiety and depression with no evidence of adverse effects (Dubad et al., 2018, 2021; Webb, Hirshberg, Davidson, & Goldberg, 2022). MoodFlow is grounded in established mood management principles, including features for activity tracking, mood tracking, sleep tracking, journaling, and data visualization, but without any of the psychoeducational content found in Remote Group CST or SuperBetter. To maintain comparability across interventions, we instructed students randomized to the MoodFlow arm to use the intervention at least 10 min per day, 5 days per week for at least 10 weeks. More detailed descriptions of the three interventions are presented in the Methodology Supplement.

Table 1
Core content of the interventions.

Remote Group CST	SuperBetter	MoodFlow
Cognitive triangle (the link between thoughts, feelings, and actions). Identifying unhelpful automatic thoughts and core beliefs. Behaviour activation and aligning behaviour with values. Cognitive restructuring Goal setting and problem-solving Identifying social and emotional support systems. Relaxation training and exposure.	Behavioral activation with goal setting. Cognitive restructuring and mindfulness training. Engaging in pleasant activities. Engaging in valued activities. Exposure therapy. Psychological distancing.	Activity tracking. Mood tracking. Sleep tracker. Journaling. Data visualization.

Note: Abbreviations: Remote Group CST=Remote Group Cognitive Behavioral Therapy Skills Training. A full description of the content of each intervention can be found in Supplementary Tab les S7, S9 and S10.

2.4. Measures

Anxiety and depression. As noted above, symptoms of anxiety were assessed with the 7-question GAD-7 (Spitzer et al., 2006) and symptoms of depression with the 9-question PHQ-9 (Kroenke et al., 2001). These self-report scales are widely used in clinical trials and have excellent psychometric properties (Kroenke, Baye, & Lourens, 2019). The GAD-7 and PHQ-9 have high convergent validity (Beard & Björgvinsson, 2014; Rahman, Dhira, Sarker, & Mehareen, 2022), adequate discriminant validity (Johnson, Ulvenes, Øktedalen, & Hoffart, 2019; Titov et al., 2011), and strong sensitivity to change in clinical trials (Titov et al., 2011; Toussaint et al., 2020). Consistent with other WMH-ICS transdiagnostic trials (Benjet, Albor, et al., 2023; Wang et al., 2023), the primary outcome was remission on both the GAD-7 and PHQ-9 (i.e., scores of 0–4 on both the 0–21 GAD-7 and on the 0–27 PHQ-9 at follow-up). Cronbach's α was 0.86-0.94-0.94 for the GAD-7 and 0.79–0.94-0.93 for the PHQ-9 at baseline, 3-month, and 6-month follow-up, respectively. Secondary outcomes included reduction on the dimensional 0–48 summation of GAD-7 and PHQ-9 scores known as the Patient Health Questionnaire Anxiety and Depression Scale (PHQ-ADS; Kroenke et al., 2019), with Cronbach's α of 0.83-0.95-0.95 at baseline, 3-month, and 6-month follow-up, respectively, and subgroup reductions in GAD-7 scores in the subsample of participants who had baseline GAD-7 scores of 10+ as well as subgroup reductions in PHQ-9 scores in the subsample of participants who had baseline PHQ-9 scores of 10+.

Baseline covariates. A review of prior research investigating predictors of differential response to psychological treatments of anxiety (Arch & Ayers, 2013; Johnson & Hoffart, 2019; Schneider, Arch, & Wolitzky-Taylor, 2015) and depression (Cohen & DeRubeis, 2018; Cuijpers et al., 2022; Maj et al., 2020) identified 10 broad domains of potentially significant predictors: socio-demographics, university characteristics and academic experiences, anxiety and depression characteristics, comorbid mental disorders, mental health treatment (including treatment expectations), physical health, exposure to recent and lifetime stressors (including COVID-19 stressors), social networks and supports, personality (including temperament and indicators of psychological resilience), and views about internet-based mental health interventions. The baseline self-report assessment of these constructs was developed by the WMH-ICS collaborators for use in trials that, like this one, are part of the WMH-ICS initiative (e.g., Benjet, Zainal, et al., 2023; Wang et al., 2023). A total of 364 questions (279 required questions and 85 branch

questions asked as follow-ups after positive responses to required questions) were included in this assessment (Supplementary Table S1). Completion time was 45–60 min.

2.5. Data analysis

We evaluated intervention effects from an intention-to-treat (ITT) perspective (Gupta, 2011) adjusting for loss to follow-up with a doubly robust method that used both propensity score modeling and outcome modeling (i.e., modeling of the sort used to impute missing values with multiple imputation) based on the random forests classifier in the *ranger R* package (Wright, Wager, & Probst, 2021). The two types of adjusted estimates were combined using the Targeted Minimum Loss-based Estimation (TMLE) method (Rose, van der Laan, & J, 2018) in the *tmle3 R* package (Coyle, 2021). See details in the Methodology Supplement. In the analyses of remission, we report adjusted risk differences (*ARD*) separately for 3-month and 6-month follow-ups. In the analyses of dimensional outcome scores, we report adjusted mean differences (*AMD*) for the same follow-up periods. Statistical significance was evaluated using 0.05-level two-sided tests. Allocation concealment was applied to the statistician.

2.6. Ethical considerations

Ethical clearance for the study was obtained from the Health Sciences Research ethics Committee at Stellenbosch University and institutional permission was obtained from all participating universities. Students who were assessed as being at risk of suicide (indicated by suicidal ideation and plan with some intention to act on these thoughts) were not excluded from the interventions but were contacted to provide them with additional information about the availability of counselling and crisis services on their campus.

3. Results

3.1. Power analysis

Consistent with prior guided digital CBT-based intervention findings (Andersson, Carlbring, & Rozental, 2019; Chauvin, Hoyt, & Otto, 2022) we assumed conservatively that ARD would be approximately 10% based on remission rates of 35% for the active interventions and 25% for the control group. A sample size of approximately n = 130 per arm would be required to have 0.8 power to detect such effects using 0.05-level two-sided tests (Champely et al., 2020). We assumed that the joint remission rates would be equivalent for Remote Group CST and

Table 2
Selected baseline characteristics of participants.

	Total		Remote Group CST		SuperBetter		MoofFlow	
	n	(%)	n	(%)	n	(%)	n	(%)
I. Gender								
Male	67	(18.1)	25	(20.7)	20	(15.9)	22	(17.7)
Female	304	(81.9)	96	(79.3)	106	(84.1)	102	(82.3)
II. Gender orientation								
Heterosexual	262	(70.6)	92	(76.0)	88	(69.8)	82	(66.2)
Gay, lesbian, or bisexual	49	(13.2)	11	(9.1)	16	(12.7)	16	(12.7)
Other	60	(16.2)	18	(14.9)	22	(17.5)	22	(17.7)
III. Age								
18-19	109	(29.4)	24	(19.8)	48	(38.1)	37	(29.8)
20	69	(18.6)	30	(24.8)	20	(15.9)	19	(15.3)
21-22	125	(33.7)	46	(38.0)	38	(30.2)	41	(33.1)
23-36	68	(18.3)	21	(17.4)	20	(15.9)	27	(21.8)
IV. Population group								
Black	163	(43.9)	52	(43.0)	56	(44.4)	55	(44.4)
White	83	(22.4)	24	(19.8)	31	(24.6)	28	(22.6)
Other race	125	(33.7)	45	(37.2)	39	(31.0)	41	(33.1)
V. Parents education								
Less than secondary graduate	31	(8.4)	6	(5.0)	12	(9.5)	13	(10.5)
Secondary graduate	106	(28.6)	35	(28.9)	32	(25.4)	39	(31.5)
Some postsecondary	113	(30.5)	31	(25.6)	43	(34.1)	39	(31.5)
University graduate	121	(32.6)	49	(40.5)	39	(31.0)	33	(26.6)
VI. Severity of anxiety (GAD-7)[a]								
Mild or none	59	(15.9)	18	(14.9)	20	(15.9)	21	(16.9)
Moderate	127	(34.2)	42	(34.7)	43	(34.1)	42	(33.9)
Severe	185	(49.9)	61	(50.4)	63	(50.0)	61	(49.2)
VII. Severity of depression (PHQ-9)[b]								
Mild or none	60	(16.2)	20	(16.5)	22	(17.5)	18	(14.5)
Moderate (including moderately severe)	211	(56.9)	70	(57.9)	73	(57.9)	68	(54.8)
Severe	100	(27.0)	31	(25.6)	31	(24.6)	38	(30.6)
VIII. Comorbidity								
Moderate on one and mild-none on the other	92	(24.8)	30	(24.8)	32	(25.4)	30	(24.2)
Moderate on both	67	(18.1)	22	(18.2)	22	(17.5)	23	(18.5)
Severe[c] on one and mild-none on the other	27	(7.3)	8	(6.6)	10	(7.9)	9	(7.3)
Severe[c] on one and moderate on the other	112	(30.2)	38	(31.5)	40	(31.7)	34	(27.4)
Severe[c] on both	73	(19.7)	23	(19.0)	22	(17.5)	28	(22.6)
(*n*)	(371)		(121)		(126)		(124)	

Note. Abbreviations: Remote Group CST=Remote Group Cognitive Behavioral Therapy Skills Training; *n*=number of participants defined by the row headings who were in the total randomized baseline sample or in intervention arms; % = the percent of all randomized baseline respondents in the column who were in the subgroups defined by the row headings; GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9.

[a] Severe refers to GAD-7 = 15+, moderate to GAD-7 = 10–14, mild or none to GAD-7 = 0–9.

[b] Severe refers to PHQ-9 = 20+, moderately severe to PHQ-9 = 15–19, moderate to PHQ-9 = 10–14, mild or none to PHQ-9 = 0–9.

[c] Including either severe or moderately severe PHQ-9.

SuperBetter, leading us to target a sample of 130 participants per arm to estimate average treatment effects. We did not attempt to power the study to detect noninferiority of SuperBetter to Remote Group CST.

3.2. Sample distribution

Table 2 presents descriptive characteristics of the students who completed the baseline assessment and were randomized. 81.9% were female, 70.6% heterosexual, 48.0% 18–20 years old, 18.3% 23+ years old (the oldest was 36), 43.9% Black (using the self-identifying categories employed in the South African population census), 22.4% White, 33.7% some other population group, and 67.4% first-generation university students. 84.1% of participants met baseline criteria for clinically significant anxiety (GAD-7≥10), including 49.9% severe (GAD-7≥15) and 34.2% moderate (GAD-7 = 10–14). 83.9% of participants met baseline criteria for clinically significant depression (PHQ-9≥10), including 27.0% severe (PHQ-9≥20) and 56.9% moderate or moderately severe (PHQ-9 = 10–19). 68.0% of participants met baseline criteria for both clinically significant anxiety and clinically significant depression. Distributions of these attributes were very similar across intervention arms.

3.3. Intervention uptake and loss to follow-up

70.4% (n = 261/371) of randomized students were onboarded and initiated the intervention (Supplementary Fig. 1), including 71.1% (n = 86/121) of those randomized to Remote Group CST, 69.0% (n = 87/126) to SuperBetter, and 71.0% (n = 88/124) to MoodFlow. Meta-data on Remote Group CST adherence found that onboarded participants attended an average (median) of 4 group sessions (2–7 interquartile

Table 3
Patterns and selected baseline predictors of completing the 3-month and 6-month follow-up assessments among the 371 participants who completed the baseline assessment.

	3-month completion rate and heterogeneity						6-month completion rate and heterogeneity					
	Completion				Heterogeneity		Completion				Heterogeneity	
	n	(%)	χ^{2a}	df	χ^{2b}	df	n	(%)	χ^{2a}	df	χ^{2b}	df
Total	233	(62.8)					154	(41.5)				
I. Intervention arm			0.9	2					1.4	2		
Remote Group CST	79	(65.3)					55	(45.4)				
SuperBettter	75	(59.5)					48	(38.1)				
MoodFlow	79	(63.7)					51	(41.1)				
II. Gender			0.7	1	1.0	2			1.8	1	5.1	2
Male	39	(58.2)					23	(34.3)				
Female	194	(63.8)					131	(43.1)				
III. Gender orientation			2.3	2	3.9	4			3.1	2	6.1	4
Heterosexual	167	(63.7)					116	(44.3)				
Gay, lesbian, or bisexual	26	(53.1)					18	(36.7)				
Other	40	(66.7)					20	(33.3)				
IV. Age			4.5	3	12.0	6			5.0	3	9.2	6
18-19	68	(62.4)					43	(39.4)				
20	50	(72.5)					36	(52.2)				
21-22	77	(61.6)					52	(41.6)				
23-36	38	(55.9)					23	(33.8)				
V. Population group			3.7	2	5.6	4			0.2	2	8.3	4
Black	109	(66.9)					68	(41.7)				
White	54	(65.1)					36	(43.4)				
Other race	70	(56.0)					50	(40.0)				
VI. Parents education			4.2	3	10.4	6			1.6	3	16.3	6
Less than secondary graduate	20	(64.5)					13	(41.9)				
Secondary graduate	59	(55.7)					40	(37.7)				
Some postsecondary	78	(69.0)					52	(46.0)				
University graduate	76	(62.8)					49	(40.5)				
VII. Severity of baseline anxiety (GAD-7)[a]			0.3	2	2.2	4			0.5	2	5.9	4
Mild or none	36	(61.0)					22	(37.3)				
Moderate	82	(64.6)					54	(42.5)				
Severe	115	(62.2)					78	(42.2)				
VIII. Severity of baseline depression (PHQ-9)[b]			0.6	2	2.4	4			0.2	2	2.6	4
Mild or none	39	(65.0)					24	(40.0)				
Moderate (including moderately severe)	129	(61.1)					88	(41.7)				
Severe	65	(65.0)					42	(42.0)				
IX. Baseline comorbidity			0.2	4	4.3	8			0.6	4	6.0	8
Moderate on one and mild-none on the other	57	(62.0)					36	(39.1)				
Moderate on both	42	(62.7)					29	(43.3)				
Severe[c] on one and mild-none on the other	18	(66.7)					10	(37.0)				
Severe[c] on one and moderate on the other	70	(62.5)					48	(42.9)				
Severe[c] on both	46	(63.0)					31	(42.5)				

Note. Abbreviations: *n*=number of baseline participants who completed the 3-month follow-up assessment; % = the proportion of baseline respondents represented by *n*; *df* = degrees of freedom; Remote Group CST=Remote Group Cognitive Behavioral Therapy Skills Training; GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9.
Severe refers to PHQ-9 = 20+, moderately severe to PHQ-9 = 15–19, moderate to PHQ-9 = 10–14, mild or none to PHQ-9 = 0–9.
Including either severe or moderately severe PHQ-9.

[a] Tests of the significance of the associations between baseline predictors and completing the 3-month follow-up assessment in the total sample. None of these tests is significant at the .05 level.

[b] Tests of the significance of variation in the association of the baseline predictor with completing the 3-month follow-up assessment across intervention arms. All such tests were nonsignificant. None of these tests is significant at the .05 level.

[c] Severe refers to GAD-7 = 15+, moderate to GAD-7 = 10–14, mild or none to GAD-7 = 0–9.

range). Attendance at group sessions was affected by interruptions in the electricity supply, because of ongoing problems with loadshedding in South Africa. Adherence data were not available for SuperBetter or MoodFlow. Very high proportions of onboarded participants completed the 3-month (91.9% [n = 79/86] Remote Group CST, 86.2% [n = 75/87] SuperBetter, 89.8% [n = 79/88] MoodFlow) and 6-month (64.0% [n = 55/86] Remote Group CST, 55.2% [n = 48/87] SuperBetter, 58.0% [n = 51/88] MoodFlow) follow-up assessments. As randomization occurred prior to onboarding, though, all randomized students were retained in the analysis, among whom 62.8% (n = 233/371) completed the 3-month assessment and 41.5% (n = 154/371) the 6-month assessment (Table 3). Neither of these follow-up rates differed significantly by; (1) treatment arm; (2) participant gender, sexual orientation, age, race, parent education, disorder severity, or disorder comorbidity; or (3) the interaction between treatment arm and these other baseline variables.

3.4. Models to adjust for nonrandom loss to follow-up

As noted above in the section on data analysis, we adjusted for loss to follow-up using a doubly robust estimation method that combined propensity score models with outcome models (i.e., models of the sort used to assign missing values with multiple imputation). All baseline covariates (Supplementary Table S1) were considered as potential predictors in both sets of models. Six propensity score loss to follow-up models were estimated (3 treatment arms x 2 follow-up periods), each predicting whether the participant completed the follow-up assessment at time *t*. Propensity score weights of $1/p_{it}$ based on these models (where p_{it} was the predicted probability of participant *i* completing the follow-up at time *t*) were then used to weight the data for each participant who completed the follow-up. Ten-fold cross-validated (10 F-CV) area under the receiver operating characteristic curve (AU-ROC) for the 3-month models was between 0.38 (SE = 0.18) for MoodFlow and 0.55 (SE = 0.19) for Remote Group CST and for the 6-month models was between 0.40 (SE = 0.18) for MoodFlow and 0.47 (SE = 0.19) for SuperBetter (Supplementary Table S2). None of these values was significantly higher than expected by chance, indicating that loss to follow-up was random with respect to the wide range of covariates included in the baseline assessment.

Twelve outcome models were estimated (3 treatment arms x 2 follow-up periods x 2 outcomes) in the subsample of participants who completed follow-up assessments. The coefficients in those models were used to impute outcome scores for each baseline respondent who did not complete the follow-up assessment. 10 F-CV AU-ROC for the joint symptom remission models was between 0.53 (SE = 0.19) for MoodFlow and 0.62 (SE = 0.29) for Remote Group CST at 3 months and between 0.58 (SE = 0.23) for MoodFlow and 0.73 (SE = 0.32) for Remote Group CST at 6 months (Supplementary Table S2), none of which was significantly higher than chance. 10 F-CV adjusted R^2 for the models that predicted PHQ-ADS scores were between 0.08 (SE = 0.05) for Remote Group CST and 0.27 (SE = 0.08) for SuperBetter at 3 months and between −0.01 (SE = undefined) for Remote Group CST and 0.31 (SE = 0.10) for MoodFlow at 6 months (Supplementary Table S2), none of which was significantly different from chance.

3.5. Aggregate comparative treatment effects

ARD: The joint GAD-7 and PHQ-9 remission rate at 3-month follow-up based on analyses design to adjust for loss to follow-up was 29.6% in the total sample (Table 4) and varied significantly across intervention arms (18.3–41.6%; $\chi^2_2 = 11.5$, p=0.009) due to participants in the Remote Group CST and SuperBetter arms both having significantly higher remission rates than participants in the MoodFlow arm: *ARD* = 23.3% for Remote Group CST vs. MoodFlow ($\chi^2_1 = 10.9$, p=0.001) and *ARD* = 12.7% for SuperBetter vs. MoodFlow ($\chi^2_1 = 3.9$, p=0.047). Number needed to treat (*NNT*) = 4–8 was in the small (SuperBetter) to medium (Remote Group CST) effect size range based on conventional guidelines (Kraemer, Frank, & Kupfer, 2011). The remission rate was non-significantly higher for Remote Group CST than SuperBetter (*ARD* = 10.6%; $\chi^2_1 = 2.0$, p=0.15). The estimated 6-month joint remission rate, in comparison, was 33.8% in the total sample and varied significantly across arms (20.1–42.3%; $\chi^2_2 = 8.8$, p=0.032) due to participants in both the Remote Group CST and SuperBetter arms having significantly higher remission rates than those in the MoodFlow arm: *ARD* = 18.9% for Remote Group CST vs. MoodFlow ($\chi^2_1 = 4.4$, p=0.035) and *ARD* = 22.2% for SuperBetter vs. MoodFlow ($\chi^2_1 = 7.5$, p=0.006). *NNT* = 5 was in the intermediate effect size range for both comparisons (Kraemer et al., 2011). The remission rate was non-significantly higher for SuperBetter than Remote Group CST (*ARD* = 3.3%; $\chi^2_1 = 0.1$, $p = 0.73$).

AMD: The estimated mean PHQ-ADS score at the 3-month follow-up was 18.9 (SD = 12.3) in the total sample (Table 5) and did not vary significantly across intervention arms (16.9 [SD = 12.8]-21.1 [12.3]); $\chi^2_2 = 5.0$, $p = 0.17$). However, participants in the Remote Group CST arm had a significantly lower mean follow-up score than participants in the MoodFlow arm and *AMD* = −4.2 ($\chi^2_1 = 4.7$, p=0.031). The 6-month estimated mean PHQ-ADS score, in comparison, was 16.7 (SD = 11.0) in the total sample, with significant variation across arms (14.5 [SD =

Table 4
3-month and 6-month joint remission rates of both GAD-7 and PHQ-9 across arms along with adjusted risk differences and number needed to treat.[a]

	Remission Rate[b]		Versus SuperBetter				Versus MoodFlow				χ^2_2[c]
	n	(%)	ARD	(SE)	χ^2_1	NNT	ARD	(SE)	χ^2_1	NNT	
I. 3-Month											
Remote Group CST	79	(41.6)	10.6	(7.5)	2.0	–	23.3*	(7.1)	10.9	4	
SuperBetter	75	(31.0)	–	–	–	–	12.7*	(6.4)	3.9	8	
MoodFlow	79	(18.3)	−12.7*	(6.4)	3.9	8	–	–	–	–	
Total	233	(29.6)									11.5*
II. 6-Month											
Remote Group CST	55	(39.0)	3.3	(9.5)	0.1	–	18.9*	(8.9)	4.4	5	
SuperBetter	48	(42.3)	–	–	–	–	22.2*	(8.1)	7.5	5	
MoodFlow	51	(20.1)	−22.2*	(8.1)	7.5	5	–	–	–	–	
Total	154	(33.8)									8.8*

Note. Abbreviations: GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9; Remote Group CST = Remote Group Cognitive Behavioral Therapy Skills Training; *n* = number of participants randomized to the different intervention arms that completed the 3-month or 6-month follow-up assessments; % = the remission rate associated with *N*; *ARD* = adjusted risk difference; *SE* = standard error of ARD; *NNT* = number needed to treat.
*Significant difference in *ARD* between the arm in the row heading and the arm in the column heading at the .05-level, two-sided test.
[a] Adjusted for differences across universities and for informative loss to follow-up using targeted minimum loss-based estimation. See the Methodology Supplement for details.
[b] Joint remission was defined as scores of 0–4 on both the GAD-7 and the PHQ-9.
[c] This is a 2-degree of freedom test for the significance of differences across all three arms.

Table 5
Baseline and follow-up mean PHQ-ADS scores across arms along with adjusted mean differences.[a]

	n	Baseline		Follow-up		Versus SuperBetter				Versus MoodFlow				χ^{2b}_2
		M	(SD)	M	(SD)	AMD	(SE)	χ^2_1	d	AMD	(SE)	χ^2_1	d	
I. 3-Month														
Remote Group CST	79	29.5	(22.2)	16.9	(12.8)	−1.3	(1.9)	0.5	0.1	−4.2*	(1.9)	4.7	0.3	
SuperBetter	75	29.7	(20.5)	18.2	(11.6)	–	–	–	–	−2.9	(1.9)	2.4	0.2	
MoodFlow	79	29.1	(21.5)	21.1	(12.3)	2.9	(1.9)	2.4	0.2	–	–	–	–	
Total	233	29.4	(21.4)	18.9	(12.3)									5.0
II. 6-Month														
Remote Group CST	55	29.2	(24.6)	14.5	(11.6)	−1.2	(2.1)	0.3	0.1	−5.6*	(2.1)	7.0	0.5	
SuperBetter	48	30.1	(24.2)	15.7	(10.4)	–	–	–	–	−4.5*	(2.1)	4.7	0.4	
MoodFlow	51	29.8	(23.1)	20.1	(10.6)	4.5*	(2.1)	4.7	0.4	–	–	–	–	
Total	154	29.4	(24.0)	16.7	(11.0)									8.0*

Note. Abbreviations: PHQ-ADS = the 0–48 Patient Health Questionnaire Anxiety and Depression Scale; *n* = number of participants randomized to the different intervention arms that completed the 3-month or 6-month follow-up assessments; *M* = mean PHQ-ADS score; *SD* = standard deviation of PHQ-ADS Score; *AMD* = adjusted mean difference; *SE* = standard error of *AMD*; *d* = Cohen's *d* effect size; Remote Group CST = Remote Group Cognitive Behavioral Therapy Skills Training. *Significant difference between the arm.

[a] Adjusted for differences across universities and for informative loss to follow-up using targeted minimum loss-based estimation. See the Methodology Supplement for details.

[b] This is a 2-degree of freedom test for the significance of differences across all three arms.

11.6]-20.1 [10.6]; $\chi^2_2 = 8.0$, p=0.045) due to participants in both the Remote Group CST and SuperBetter arms having significantly lower scores than those in the MoodFlow arm, with *AMD* = −5.6 for Remote Group CST vs. MoodFlow ($\chi^2_1 = 7.0$, p=0.008) and *AMD* = −4.5 for SuperBetter vs. MoodFlow ($\chi^2_1 = 4.7$, $p = 0.031$). These effect sizes were in the small range based on conventional guidelines (Cohen, 1988). The mean was non-significantly lower with Remote Group CST than SuperBetter, with *AMD* = −1.2 ($\chi^2_1 = 0.3$, $p = 0.59$).

3.6. Subgroup comparative treatment effects

ARD: The estimated GAD-7 and PHQ-9 remission rates at 3-month follow-up among participants with baseline GAD = 7 = 10+ and PHQ-9 = 10+ were 34.5% and 45.7%, respectively (Supplementary Tables S3 and S4). These remission rates did not vary significantly across intervention arms in the GAD-7 = 10+ subsample (25.3–45.8%, $\chi^2_2 = 5.9$, p = 0.05) but did in the PHQ-9-10+ subsample (34.2–54.8%, $\chi^2_2 = 6.1$, p = 0.047). For both GAD-7 and PHQ-9, the highest remission rate was in the Remote Group CST arm (45.8–54.8%) followed by a non-significantly lower remission rate in the SuperBetter arm (32.1–48.5%, χ^2_1 = 0.1–0.5, p = 0.71–0.50), and a lowest remission rate in the MoodFlow arm (25.3–34.2%) that was significantly lower for both GAD-7 and PHQ-7 than in the Remote Group CST arm (χ^2_1 = 5.8–5.6, p = 0.01–0.018, NNT = 5) but not significantly lower than in the SuperBetter arm (χ^2_1 = 0.4–3.0, p = 0.53–0.08, NNT = 15–7).

The situation for 6-month remission was somewhat different. Estimated GAD-7 and PHQ-9 remission rates were 36.4% and 52.8%, respectively, in the total subsamples (Supplementary Tables S3 and S4). The GAD-7 remission rates did not vary significantly across arms (27.6–40.9% $\chi^2_2 = 2.4$, p = 0.12) but the PHQ-9 remission rates did vary significantly across arms (37.7–66.1%, χ^2_2-8.7, p = 0.003). In the case of both outcomes, the highest remission rate was in the SuperBetter arm (40.9–66.1%) followed by a non-significantly lower remission rate in the Remote Group CST arm (40.2–56.3%, $\chi^2_1 = 0.0$–0.8, p = 0.91–0.36), and the lowest remission rate in the MoodFlow arm (27.6–37.7%). The significant variation across arms for PHQ-9 was due to a significantly higher ARD associated with SuperBetter than MoodFlow (*ARD* = 28.4, $\chi^2_1 = 8.5$, p = 0.004, NNT = 5). None of the other *ARD*s was significant for either outcome ($\chi^2_1 = 0.0$–3.0, p = 0.97–0.08).

AMD: The estimated 3-month follow-up mean GAD-7 and PHQ-9 scores in the subsamples of participants with baseline GAD = 7 = 10+ and PHQ-9 = 10+ were 9.9 (SD = 6.4) and 10.2 (SD = 7.8), respectively (Supplementary Tables S5 and S6). Neither of these varied significantly across intervention arms ($\chi^2_2 = 2.4$–4.2, p = 0.33–0.12). The situation was different, though, for estimated 6-month means, which varied significantly across arms for both GAD-7 ($\chi^2_2 = 6.3$, p = 0.041) and PHQ-9 ($\chi^2_2 = 8.4$, p = 0.015). In the case of GAD-7, this was due to a significantly lower *AMD* for Remote Group CST than MoodFlow (*AMD* = −3.0, $\chi^2_1 = 6.3$, p = 0.010), whereas in the case of PHQ-9 it was due to a significantly lower *AMD* for SuperBetter than MoodFlow (*AMD* = −4.1, $\chi^2_1 = 8.0$, p = 0.003). None of the other *AMD*s was significant for either outcome ($\chi^2_1 = 0.1$–3.6, p = 0.70–0.06).

4. Discussion

This is both the first controlled trial to evaluate the effects of Remote Group CST compared to a control condition and the first to evaluate the comparative effectiveness of SuperBetter relative to an active comparator. The 39–42% of students randomized to the active intervention who experienced joint remission of GAD-7 and PHQ-9 were at the upper end of the pooled remission rates for guided digital interventions found in a meta-analysis of prior controlled trials for depression (Karyotaki et al., 2018) and in a separate meta-analysis of digital CBT interventions for anxiety, depression, and other common behavioral health problems (Andersson et al., 2019). The ITT effect sizes for joint remission were in the small to medium range using conventional definitions (Kraemer et al., 2011). However, as intervention uptake in this trial was quite low, the lower-bound Treatment-On-the-Treated (TOT) effect, which can be defined roughly as the ratio of ITT to the proportion of randomized students who were onboarded, would be in the range typically considered medium or large. This observation is consistent with meta-analyses of previous SuperBetter trials (Firth, Torous, Nicholas, Carney, Pratap et al., 2017; Firth, Torous, Nicholas, Carney, Rosenbaum, et al., 2017; Linardon et al., 2019; Serrano-Ripoll et al., 2022; Weisel et al., 2019).

These results compare favorably to the other large comparative effectiveness digital intervention of low- and middle-income university students with anxiety and/or depression, which was carried out in Latin America (Benjet, Albor, et al., 2023). *ARD* for 3-month (the only follow-up period reported for the other trial) joint remission with SuperBetter (*ARD* = 12.7%) and Remote Group CST (*ARD* = 23.3%) are higher than the ARD = 11.2% for guided digital CBT in the Latin American trial. The *ARD* in the Latin American trial for self-guided digital CBT versus control, in comparison, was nonsignificant (*ARD* = −1.9). The comparison between SuperBetter and the self-guided digital CBT in the Latin American trial is especially striking from a scalability perspective given that both self-guided interventions are delivered without any professional guidance. In addition, the time spent by Remote Group CST therapists (5 min per patient per week for a 50-min

group of 12 participants) is much less than the 20 min per patient per week spent by guided intervention coaches in the Latin American trial. It remains to be seen in larger trials, though, if Remote Group CST yields significantly better outcomes than SuperBetter in some subset of students and over longer periods of time. We found that the seeming aggregate advantage of Remote Group CST over SuperBetter disappeared at the 6-month follow-up. If this continues to be the case in larger trials and if no subgroups of participants are found for whom Remote Group CST is superior to SuperBetter, then SuperBetter would be the preferred intervention for all participants from a cost perspective.

Study strengths included the focus on underserved South African university students with clinically significant anxiety and/or depression, an assessment of two scalable digital interventions compared to an active control condition, the administration of a much more extensive baseline battery of potentially important prescriptive predictors than in prior CBT trials, and the use of a state-of-the-art analysis method to adjust for the effects of loss to follow-up. Notably, our Remote Group CST intervention demonstrated efficacy at the 3-month follow-up that matched that of SuperBetter—an application with significant commercial success and revenue in the millions. There were also several important limitations, including the use of self-report measures rather than clinician assessments of anxiety and depression, a sample size too low to evaluate noninferiority of SuperBetter to Remote Group CST, an initial intervention uptake rate much lower than in prior digital CBT anxiety and depression interventions (Lattie et al., 2019), failures to assess either intervention acceptability to students or engagement with SuperBetter or MoodFlow, and the inability to prevent participants from obtaining other types of treatment.

Two of these limitations were built into the design purposefully. (1) We chose the GAD-7 and PHQ-9 self-report scales to assess outcomes because we lacked the resources needed to base outcome assessments on clinical interviews and because the GAD-7 and PHQ-9 are well-validated and commonly used outcomes in clinical trials (McMillan, Gilbody, & Richards, 2010; Toussaint et al., 2020). (2) We powered the trial only to investigate aggregate effects because we wanted to establish the existence of statistically significant benefits of both SuperBetter and Remote Group CST in the population of interest before mounting a much larger trial to develop an individualized treatment rule (ITR) that will evaluate whether these two interventions differ significantly in comparative effectiveness across students. Because of this longer-term goal, it was important for us to carry out the current trial using a design that required participants to complete the lengthy baseline assessment that will be required to develop an ITR even though this assessment was not needed to study aggregate intervention effects.

The exceptionally low intervention uptake rates that occurred across arms, in comparison, were not anticipated. An obvious speculation is that this occurred because of the lengthy baseline assessment. However, this speculation is inconsistent with the fact that other WMH-ICS interventions that used the same baseline assessment battery did not experience a similar uptake problem (e.g., Benjet, Zainal, et al., 2023; Wang et al., 2023). Post hoc qualitative debriefing suggested that the problem instead occurred because we asked students to do too much beyond completing the long baseline assessment (i.e., going online and booking a slot in a group, downloading the app, reaching out to schedule an onboarding call with the study coordinator) before initiating personal contact with the project coordinators who initiated the interventions. Our onboarding protocol has subsequently been changed to address this problem by scheduling an evidence-based 50-min psychoeducational telephone engagement session (Bisby, Balakumar, Scott, Titov, & Dear, 2023) with each eligible student prior to randomization to increase intervention engagement.

The absence of follow-up data on acceptability to users or on self-reported engagement with SuperBetter or MoodFlow were oversights that made it impossible to study the effects of TOT, but this information will be added to outcome assessments in extensions of the trial. In addition, no objective information was available on engagement with these interventions because we did not have a collaboration with the owners of the SuperBetter or MoodFlow apps that would allow us to gain access to the meta-data for users.

Importantly, the sample was self-selected, as only students eager to obtain help were enrolled in the interventions. Different outcomes are likely to be observed with populations of students who are not seeking help. Finally, we were unable to control participant engagement with other treatments. We asked participants in the 3-month follow-up assessment if they received any other psychological counselling or medication for their anxiety or depression since randomization. 28.0–32.9% of participants across arms reported that they did ($\chi^2_2 = 0.5$, $p = 0.79$). As this was uncontrolled, we had no way to adjust for this in analysis. In theory, more detailed information about types of treatments received could be considered as part of a mediator analysis, but no such information was collected, raising uncertainties about the mechanisms through which the benefits of Remote Group CST and SuperBetter occur. This is a more general limitation in psychotherapy research, as information is seldom collected about other types of treatments received and such information, when available, is virtually used in analyses.

4.1. Future directions

As noted in the prior section, the current study was carried out as a first step in designing a much larger precision treatment trial that will attempt to develop an ITR for assigning the optimal digital intervention to each student seeking treatment for anxiety and/or depression. The goal in developing such an ITR will be to optimize aggregate remission for a fixed cost by assigning the lowest-cost intervention (in our case, self-guided SuperBetter) to all students for whom this intervention is more effective than a control condition but noninferior to more expensive interventions (in our case, Remote Group CST) and to assign the more expensive intervention only to the subset of students for whom it is significantly more effective than both the control condition and the lower-cost intervention. The remaining students, those whose outcomes under both the lower- and higher-cost interventions are estimated by the ITR to be no better than in a control condition, will be referred to more intensive interventions (in our case, in-person one-on-one treatment at the student health clinic).

It is noteworthy that a much larger recent WMH-ICS comparative effectiveness trial, one powered to develop an ITR and using a different set of self-guided and guided digital CBT-based interventions, was carried out in a similar sample of anxious and/or depressed university students in Latin America (Benjet, Zainal, et al., 2023). An ITR was developed in that study using the same battery of baseline covariates as in the current trial. Roughly 20% of participants had an estimated remission rate with a conventional self-guided digital intervention that was noninferior to the remission rate with the guided intervention but superior to the remission rate associated with treatment as usual. Another roughly 60% of participants had a significantly higher estimated remission rate with the guided intervention. The remaining roughly 20% were not helped by either intervention. Given the much stronger aggregate effect of SuperBetter in the current South African trial than of the conventional self-guided intervention in the Latin American trial, we anticipate that we will find a higher proportion of participants in the planned South African precision treatment trial who will be optimized by SuperBetter (i.e., noninferior to Remote Group CST and superior to MoodFlow) when we develop an ITR with the South African data. However, we were unable to evaluate this hypothesis in the current trial because the sample size requirements for estimating an ITR are much greater than for estimating main effects (Luedtke, Sadikova, & Kessler, 2019).

The rolling recruitment format of Remote Group CST (i.e., content repeated in a rolling fashion every 10 weeks, allowing participants to enter at any time and stay for the next 10 weeks to cover all the content) was necessary to make this intervention feasible. Questions can be raised, though, whether variation in intervention effects exist depending

on which of the five core intervention themes is delivered first and, if so, whether this order effect is constant across participants. We plan to investigate these questions once the sample size increases but it was not possible to do so in the current small sample.

5. Conclusion

We found evidence that both the self-guided SuperBetter app and the remote group CBT Skills Training intervention evaluated in this trial outperformed a mood monitoring control condition over both 3-month and 6-month follow-up periods in treating university students with clinically significant anxiety and/or depression. Further research is needed to determine if an ITR can be developed to improve aggregate intervention effects by assigning different interventions to different participants. If not, then the highly scalable self-guided SuperBetter will be the preferred intervention for all students from a cost perspective. In either case, results of a precision treatment trial will be of great value in addressing the high level of unmet need for treatment of anxiety and depression found among South African university students.

Funding

This study was funded by the South African Medical Research Council (through the MCSP awarded to Jason Bantjes) and the South African National Research Foundation (grant number 142143 awarded to Jason Bantjes).

Role of the funder

The funder has no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

CRediT authorship contribution statement

Jason Bantjes: Writing – review & editing, Writing – original draft, Supervision, Investigation, Funding acquisition, Conceptualization. **Xanthe Hunt:** Writing – review & editing, Project administration, Investigation, Data curation. **Pim Cuijpers:** Writing – review & editing. **Alan E. Kazdin:** Writing – review & editing. **Chris J. Kennedy:** Writing – review & editing. **Alex Luedtke:** Writing – review & editing. **Ivana Malenica:** Writing – review & editing. **Maria Petukhova:** Writing – review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Nancy Sampson:** Writing – review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Nur Hani Zainal:** Writing – review & editing, Writing – original draft. **Charl Davids:** Writing – review & editing. **Munita Dunn-Coetzee:** Writing – review & editing. **Rone Gerber:** Writing – review & editing. **Dan J. Stein:** Writing – review & editing. **Ronald C. Kessler:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

In the past 3 years, Dr. Kessler was a consultant for Cambridge Health Alliance, Canandaigua VA Medical Center, Holmusk, Partners Healthcare, Inc., RallyPoint Networks, Inc., and Sage Therapeutics. He has stock options in Cerebral Inc., Mirah, PYM, Roga Sciences and Verisense Health. The other authors have no competing interests to declare. None of the authors has any financial interest in any of these interventions.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2024.104554.

References

Andersson, G., Carlbring, P., & Rozental, A. (2019). Response and remission rates in internet-based cognitive behavior therapy: An individual patient data meta-analysis. *Frontiers in Psychiatry, 10*, 749. https://doi.org/10.3389/fpsyt.2019.00749

Arch, J. J., & Ayers, C. R. (2013). Which treatment worked better for whom? Moderators of group cognitive behavioral therapy versus adapted mindfulness based stress reduction for anxiety disorders. *Behaviour Research and Therapy, 51*(8), 434–442. https://doi.org/10.1016/j.brat.2013.04.004

Bantjes, J., Kazdin, A. E., Cuijpers, P., Breet, E., Dunn-Coetzee, M., Davids, C., et al. (2021). A web-based group cognitive behavioral therapy intervention for symptoms of anxiety and depression among university students: Open-label, pragmatic trial. *JMIR Mental Health, 8*(5), Article e27400. https://doi.org/10.2196/27400

Bantjes, J., Kessler, M., Lochner, C., Breet, E., Bawa, A., Roos, J., et al. (2023). The mental health of university students in South Africa: Results of the national student survey. *Journal of Affective Disorders, 321*, 217–226. https://doi.org/10.1016/j.jad.2022.10.044

Bantjes, J., Kessler, M. J., Hunt, X., Stein, D. J., & Kessler, R. C. (2023). Treatment rates and barriers to mental health service utilisation among university students in South Africa. *International Journal of Mental Health Systems, 17*(1), 38. https://doi.org/10.1186/s13033-023-00605-7

Beard, C., & Björgvinsson, T. (2014). Beyond generalized anxiety disorder: Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *Journal of Anxiety Disorders, 28*(6), 547–552. https://doi.org/10.1016/j.janxdis.2014.06.002

Benjet, C., Albor, Y., Alvis-Barranco, L., Contreras-Ibáñez, C. C., Cuartas, G., Cudris-Torres, L., et al. (2023). Internet-delivered cognitive behavior therapy versus treatment as usual for anxiety and depression among Latin American university students: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 91*(12), 694–707. https://doi.org/10.1037/ccp0000846

Benjet, C., Zainal, N. H., Albor, Y., Alvis-Barranco, L., Carrasco-Tapias, N., Contreras-Ibáñez, C. C., et al. (2023). A precision treatment model for internet-delivered cognitive behavioral therapy for anxiety and depression among university students: A secondary analysis of a randomized clinical trial. *JAMA Psychiatry, 80*(8), 768–777. https://doi.org/10.1001/jamapsychiatry.2023.1675

Bisby, M. A., Balakumar, T., Scott, A. J., Titov, N., & Dear, B. F. (2023). An online therapist-guided ultra-brief treatment for depression and anxiety: A randomized controlled trial. *Psychological Medicine*, 1–12. https://doi.org/10.1017/s003329172300260x

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. *Cognitive Behaviour Therapy, 47*(1), 1–18. https://doi.org/10.1080/16506073.2017.1401115

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., et al. (2020). Basic functions for power analysis. Retrieved September 21 from https://cran.r-project.org/web/packages/pwr/pwr.pdf.

Chauvin, G. V., Hoyt, D. L., & Otto, M. W. (2022). Perspectives on the course of major depression: Trajectories of remission and relapse with CBT or no treatment. *Journal of Behavioral and Cognitive Therapy, 32*(3), 231–237. https://doi.org/10.1016/j.jbct.2022.04.003

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Academic press.

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology, 14*, 209–236. https://doi.org/10.1146/annurev-clinpsy-050817-084746

Coyle, J. R. (2021). *tmle3: The extensible TMLE framework*. Retrieved March 4 from https://tlverse.org/tmle3/.

Cuijpers, P., Auerbach, R. P., Benjet, C., Bruffaerts, R., Ebert, D., Karyotaki, E., et al. (2019). The World health organization World mental health international college student initiative: An overview. *International Journal of Methods in Psychiatric Research, 28*(2), Article e1761. https://doi.org/10.1002/mpr.1761

Cuijpers, P., Ciharova, M., Quero, S., Miguel, C., Driessen, E., Harrer, M., et al. (2022). The contribution of "individual participant data" meta-analyses of psychotherapies for depression to the development of personalized treatments: A systematic review. *Journal of Personalized Medicine, 12*(1). https://doi.org/10.3390/jpm12010093

Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: A network meta-analysis. *JAMA Psychiatry, 76*(7), 700–707. https://doi.org/10.1001/jamapsychiatry.2019.0268

Dubad, M., Elahi, F., & Marwaha, S. (2021). The clinical impacts of mobile mood-monitoring in young people with mental health problems: The MeMO study. *Frontiers in Psychiatry, 12*, 687270. https://doi.org/10.3389/fpsyt.2021.687270, 687270.

Dubad, M., Winsper, C., Meyer, C., Livanou, M., & Marwaha, S. (2018). A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. *Psychological Medicine, 48*(2), 208–228. https://doi.org/10.1017/s0033291717001659

Esfandiari, N., Mazaheri, M. A., Akbari-Zardkhaneh, S., Sadeghi-Firoozabadi, V., & Cheraghi, M. (2021). Internet-delivered versus face-to-face cognitive behavior

therapy for anxiety disorders: Systematic review and meta-analysis. *International Journal of Preventive Medicine, 12*, 153. https://doi.org/10.4103/ijpvm.ijpvm_208_21
Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., et al. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: A meta-analysis of randomized controlled trials. *World Psychiatry, 16*(3), 287–298. https://doi.org/10.1002/wps.20472
Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *Journal of Affective Disorders, 218*, 15–22. https://doi.org/10.1016/j.jad.2017.04.046
Gan, D. Z. Q., McGillivray, L., Han, J., Christensen, H., & Torok, M. (2021). Effect of engagement with digital interventions on mental health outcomes: A systematic review and meta-analysis. *Frontiers in Digital Health, 3*, Article 764079. https://doi.org/10.3389/fdgth.2021.764079
Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research, 2*(3), 109–112. https://doi.org/10.4103/2229-3485.83221
Hamamci, Z. (2006). Integrating psychodrama and cognitive behavioral therapy to treat moderate depression. *The Arts in Psychotherapy, 33*(3), 199–207. https://doi.org/10.1016/j.aip.2006.02.001
Johnson, S. U., & Hoffart, A. (2019). Moderators and predictors of outcome in metacognitive and cognitive behavioural therapy for co-morbid anxiety disorders. *Clinical Psychology & Psychotherapy, 26*(4), 399–408. https://doi.org/10.1002/cpp.2361
Johnson, S. U., Ulvenes, P. G., Øktedalen, T., & Hoffart, A. (2019). Psychometric properties of the general anxiety disorder 7-item (GAD-7) scale in a heterogeneous psychiatric sample. *Frontiers in Psychology, 10*, 1713. https://doi.org/10.3389/fpsyg.2019.01713
Karyotaki, E., Ebert, D. D., Donkin, L., Riper, H., Twisk, J., Burger, S., et al. (2018). Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. *Clinical Psychology Review, 63*, 80–92. https://doi.org/10.1016/j.cpr.2018.06.007
Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., et al. (2017). Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: A meta-analysis of individual participant data. *JAMA Psychiatry, 74*(4), 351–359. https://doi.org/10.1001/jamapsychiatry.2017.0044
Kim, J., Aryee, L. M. D., Bang, H., Prajogo, S., Choi, Y. K., Hoch, J. S., et al. (2023). Effectiveness of digital mental health tools to reduce depressive and anxiety symptoms in low- and middle-income countries: Systematic review and meta-analysis. *JMIR Mental Health, 10*, Article e43066. https://doi.org/10.2196/43066
Kraemer, H. C., Frank, E., & Kupfer, D. J. (2011). How to assess the clinical impact of treatments on patients, rather than the statistical impact of treatments on measures. *International Journal of Methods in Psychiatric Research, 20*(2), 63–72. https://doi.org/10.1002/mpr.340
Kroenke, K., Baye, F., & Lourens, S. G. (2019). Comparative validity and responsiveness of PHQ-ADS and other composite anxiety-depression measures. *Journal of Affective Disorders, 246*, 437–443. https://doi.org/10.1016/j.jad.2018.12.098
Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
Lattie, E. G., Adkins, E. C., Winquist, N., Stiles-Shields, C., Wafford, Q. E., & Graham, A. K. (2019). Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: Systematic review. *Journal of Medical Internet Research, 21*(7), Article e12869. https://doi.org/10.2196/12869
Li, S. H., Achilles, M. R., Werner-Seidler, A., Beames, J. R., Subotic-Kerry, M., & O'Dea, B. (2022). Appropriate use and operationalization of adherence to digital cognitive behavioral therapy for depression and anxiety in youth: Systematic review. *JMIR Mental Health, 9*(8), Article e37640. https://doi.org/10.2196/37640
Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. *World Psychiatry, 18*(3), 325–336. https://doi.org/10.1002/wps.20673
Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science, 7*(3), 445–461. https://doi.org/10.1177/2167702618815466
Maj, M., Stein, D. J., Parker, G., Zimmerman, M., Fava, G. A., De Hert, M., et al. (2020). The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry, 19*(3), 269–293. https://doi.org/10.1002/wps.20771
Marshall, J. M., Dunstan, D. A., & Bartik, W. (2021). Smartphone psychological therapy during COVID-19: A study on the effectiveness of five popular mental health apps for anxiety and depression. *Frontiers in Psychology, 12*, 775775. https://doi.org/10.3389/fpsyg.2021.775775, 775775.
McGonigal, J. (2016). *SuperBetter: The power of living gamefully*. Penguin Books.
McMillan, D., Gilbody, S., & Richards, D. (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders, 127*(1–3), 122–129. https://doi.org/10.1016/j.jad.2010.04.030
Mohr, D. C., Cuijpers, P., & Lehman, K. (2011). Supportive accountability: A model for providing human support to enhance adherence to eHealth interventions. *Journal of Medical Internet Research, 13*(1), e30. https://doi.org/10.2196/jmir.1602
Moodflow. (2023). Moodflow: Journal, diary & mood tracker. Retrieved Aug 12 from https://www.moodflow.co/.
Papola, D., Ostuzzi, G., Tedeschi, F., Gastaldon, C., Purgato, M., Del Giovane, C., et al. (2023). CBT treatment delivery formats for panic disorder: A systematic review and network meta-analysis of randomised controlled trials. *Psychological Medicine, 53*(3), 614–624. https://doi.org/10.1017/s0033291722003683
Peden, A. R., Hall, L. A., Rayens, M. K., & Beebe, L. L. (2000). Reducing negative thinking and depressive symptoms in college women. *Journal of Nursing Scholarship, 32*(2), 145–151. https://doi.org/10.1111/j.1547-5069.2000.00145.x
Rahman, M. A., Dhira, T. A., Sarker, A. R., & Mehareen, J. (2022). Validity and reliability of the Patient Health Questionnaire scale (PHQ-9) among university students of Bangladesh. *PLoS One, 17*(6), Article e0269634. https://doi.org/10.1371/journal.pone.0269634
Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review, 32*(4), 329–342. https://doi.org/10.1016/j.cpr.2012.02.004
Roepke, A. M., Jaffee, S. R., Riffle, O. M., McGonigal, J., Broome, R., & Maxwell, B. (2015). Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. *Games for Health Journal, 4*(3), 235–246. https://doi.org/10.1089/g4h.2014.0046
Rose, S., van der Laan, & J, M. (2018). Ltmle. In M. J. van der Laan, & S. Rose (Eds.), *Targeted learning in data science: Causal inference for complex longitudinal studies* (pp. 35–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-65304-4_4.
Ryan, C., Bergin, M., & Wells, J. S. (2018). Theoretical perspectives of adherence to web-based interventions: A scoping review. *International Journal of Behavioral Medicine, 25*(1), 17–29. https://doi.org/10.1007/s12529-017-9678-8
Saleem, M., Kühne, L., De Santis, K. K., Christianson, L., Brand, T., & Busse, H. (2021). Understanding engagement strategies in digital interventions for mental health promotion: Scoping review. *JMIR Mental Health, 8*(12), Article e30000. https://doi.org/10.2196/30000
Schneider, R. L., Arch, J. J., & Wolitzky-Taylor, K. B. (2015). The state of personalized treatment for anxiety disorders: A systematic review of treatment moderators. *Clinical Psychology Review, 38*, 39–54. https://doi.org/10.1016/j.cpr.2015.02.004
Schulz, K. F., Altman, D. G., Moher, D., & the Consort Group. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*(18). https://doi.org/10.1186/1741-7015-8-18
Serrano-Ripoll, M. J., Zamanillo-Campos, R., Fiol-DeRoque, M. A., Castro, A., & Ricci-Cabello, I. (2022). Impact of smartphone app–based psychological interventions for reducing depressive symptoms in people with depression: Systematic literature review and meta-analysis of randomized controlled trials. *JMIR Mhealth Uhealth, 10*(1), Article e29621. https://doi.org/10.2196/29621
Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092
Takagaki, K., Okamoto, Y., Jinnin, R., Mori, A., Nishiyama, Y., Yamamura, T., et al. (2016). Behavioral activation for late adolescents with subthreshold depression: A randomized controlled trial. *European Child & Adolescent Psychiatry, 25*(11), 1171–1182. https://doi.org/10.1007/s00787-016-0842-5
Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., & Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive Behaviour Therapy, 40*(2), 126–136. https://doi.org/10.1080/16506073.2010.550059
Toussaint, A., Hüsing, P., Gumz, A., Wingenfeld, K., Härter, M., Schramm, E., et al. (2020). Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). *Journal of Affective Disorders, 265*, 395–401. https://doi.org/10.1016/j.jad.2020.01.032
Wang, A. Y., Vereschagin, M., Richardson, C. G., Xie, H., Hudec, K. L., Munthali, R. J., et al. (2023). Evaluating the effectiveness of a codeveloped e-mental health intervention for university students: Protocol for a randomized controlled trial. *JMIR Research Protocols, 12*, Article e49364. https://doi.org/10.2196/49364
Webb, C. A., Hirshberg, M. J., Davidson, R. J., & Goldberg, S. B. (2022). Personalized prediction of response to smartphone-delivered meditation training: Randomized controlled trial. *Journal of Medical Internet Research, 24*(11), Article e41566. https://doi.org/10.2196/41566
Weisel, K. K., Fuhrmann, L. M., Berking, M., Baumeister, H., Cuijpers, P., & Ebert, D. D. (2019). Standalone smartphone apps for mental health—a systematic review and meta-analysis. *NPJ Digital Medicine, 2*(1), 118. https://doi.org/10.1038/s41746-019-0188-8
Wright, M. N., Wager, S., & Probst, P. (2021). A fast implementation of random forests. Retrieved Sep 20 from https://cran.r-project.org/web/packages/ranger/ranger.pdf.
Yang, X., Zhao, J., Chen, Y., Zu, S., & Zhao, J. (2018). Comprehensive self-control training benefits depressed college students: A six-month randomized controlled intervention trial. *Journal of Affective Disorders, 226*, 251–260. https://doi.org/10.1016/j.jad.2017.10.014
Zemestani, M., Davoodi, I., Honarmand, M. M., Zargar, Y., & Ottaviani, C. (2016). Comparative effects of group metacognitive therapy versus behavioural activation in moderately depressed students. *Journal of Mental Health, 25*(6), 479–485. https://doi.org/10.3109/09638237.2015.1057326