MISHA – A Chatbot-delivered Stress Management Coaching for Students: Pilot Randomized Controlled Trial

Sandra Ulrich, Natascha Lienhard, Hansjörg Künzli, Tobias Kowatsch

Submitted to: JMIR mHealth and uHealth
on: December 04, 2023

Table of Contents

MISHA – A Chatbot-delivered Stress Management Coaching for Students: Pilot Randomized Controlled Trial

Sandra Ulrich[1*] MSc; Natascha Lienhard[2*] MSc; Hansjörg Künzli[2] LicPhil; Tobias Kowatsch[3, 4, 5] PhD

[1]School of Applied Psychology ZHAW Zurich University of Applied Sciences Zürich CH
[2]School of Applied Psychology ZHAW Zurich University of Applied Sciences Zurich CH
[3]Institute for Implementation Science in Health Care University of Zurich, Zürich CH
[4]School of Medicine University of St. Gallen St.Gallen CH
[*]these authors contributed equally

Corresponding Author:
Sandra Ulrich MSc
School of Applied Psychology
ZHAW Zurich University of Applied Sciences
Pfingstweidstrasse 96
2
Zürich
CH

Abstract

Background: Globally, students face increasing mental health challenges, including elevated stress levels and declining well-being, leading to academic performance issues and mental health disorders. However, due to stigma and symptom underestimation, students rarely seek effective stress management solutions. Conversational agents (CAs) in the health sector have shown promise in reducing stress, depression, and anxiety. Nevertheless, research on their effectiveness for stressed students remains limited.

Objective: This study aimed to develop a CA-delivered stress-management coaching intervention for students called MISHA, and to evaluate the effectiveness, engagement, and acceptance.

Methods: In an unblinded randomized controlled trial, Swiss students experiencing stress were recruited online. By a 1.1 randomization, participants (N=140) were allocated to the intervention or wait-list control group. Treatment effectiveness on changes on primary outcome perceived stress, and secondary outcomes depression, anxiety, psychosomatic symptoms, and active coping, were evaluated using analyses of variance (ANOVA) and general estimating equations (GEE).

Results: The per protocol (PP) analysis revealed evidence for improvement of stress, depression, and somatic symptoms with effects ranging from medium to medium-large (d=0.54 to d=0.67), while anxiety, and active coping did not change significantly (d=0.37; d=0.39). In the intention to treat (ITT) analysis, we found evidence for reduced stress (?=–0.13, 95% CI –0.20 to –0.05, P<.001), depressive symptoms (?=–0.23, 95% CI –0.38 to –0.08, P=.003), and psychosomatic symptoms (?=–0.16, 95% CI = –0.27 to –0.06, P=.003), while anxiety, and active comping did change. Overall, 60% of participants in the intervention group completed the coaching by completing the outro. The particularly appreciated the quality, quantity, credibility, and visual representation of information. While individual customization was rated lowest, the target group fitting was perceived high.

Conclusions: Findings indicate that MISHA is feasible, acceptable and effective in reducing perceived stress among students in Switzerland. Future research is needed for different populations, for example in students with high stress levels or compared to active controls. Clinical Trial: German Clinical Trials Register DRKS 00030004; Swiss Ethics BASEC-Nr. Req-2020-01038

(JMIR Preprints 04/12/2023:54945)
DOI: https://doi.org/10.2196/preprints.54945

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

MISHA – A Chatbot-delivered Stress Management Coaching for Students: Pilot Randomized Controlled Trial

Authors

Sandra Ulrich, MSc[1#], Natascha Lienhard, MSc[1#], Hansjörg Künzli, lic phil[1], Tobias Kowatsch, PhD[234]

[1] ZHAW School of Applied Psychology, Section Diagnostics and Counseling, Psychological Institute, Zurich, Switzerland

[2] Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

[3] School of Medicine, University of St.Gallen, St.Gallen, Switzerland

[4] Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

# NL and SU contributed equally to this manuscript and share first authorship.

Corresponding author:

Sandra Ulrich

School of Applied Psychology

ZHAW Zurich University of Applied Sciences

Pfingstweidstrasse 96

CH-8005 Zurich, Switzerland

Tel: +41 58 934 84 51

Email: sandra.ulrich@zhaw.ch

Abstract

Background: Globally, students face increasing mental health challenges, including elevated stress levels and declining well-being, leading to academic performance issues and mental health disorders. However, due to stigma and symptom underestimation, students rarely seek effective stress management solutions. Conversational agents (CAs) in the health sector have shown promise in reducing stress, depression, and anxiety. Nevertheless, research on their effectiveness for stressed students remains limited.

Objective: This study aimed to develop a CA-delivered stress-management coaching intervention for students called MISHA, and to evaluate its effectiveness, engagement, and acceptance.

Methods: In an unblinded randomized controlled trial, Swiss students experiencing stress were recruited online. Using a 1.1 randomization, participants (N=140) were allocated to either the intervention or wait-list control group. Treatment effectiveness on changes in primary outcome perceived stress and secondary outcomes depression, anxiety, psychosomatic symptoms, and active coping, were self-assessed and evaluated using analyses of variance (ANOVA) for repeated measure and general estimating equations (GEE).

Results: The per protocol (PP) analysis revealed evidence for improvement of stress, depression, and somatic symptoms with effects ranging from medium to medium-large (d=–0.36 to d=–0.60), while anxiety, and active coping did not change significantly (*d*=–0.29; *d*=0.13). In the intention to treat (ITT) analysis revealed similar results, we found evidence for reduced stress (*β*=–0.13, 95% CI –0.20 to –0.05, *P*<.001*)*, depressive symptoms (*β*=–0.23, 95% CI –0.38 to –0.08, *P*=.003), and psychosomatic symptoms (*β*=–0.16, 95% CI = –0.27 to –0.06, *P*=.003), while anxiety, and active coping did change. Overall, 60% of participants in the intervention group completed the coaching by completing the post-intervention survey. They particularly appreciated the quality, quantity, credibility, and visual representation of information. While individual customization was rated lowest, the target group fitting was perceived as high.

Conclusions: Findings indicate that MISHA is feasible, acceptable, and effective in reducing perceived stress among students in Switzerland. Future research is needed with different populations, for example, in students with high stress levels or compared to active controls.

Trial Registration: German Clinical Trials Register DRKS 00030004; Swiss Ethics BASEC-Nr. Req-2020-01038

Keywords: conversational agent, mobile health, smartphone, stress-management, lifestyle, behavior-change, coaching

Introduction

Stress is rapidly becoming a major issue affecting adults in developed countries, especially during periods of uncertainty and worry. Chronic stress is closely related to mental illnesses such as anxiety disorders and

depression, leading to various symptoms such as sleep disturbances, pain, dizziness, cardiovascular and digestive problems, as well as fatigue [1,2]. Younger individuals, particularly students [3–7], are experiencing a decline in mental health on a global scale [8,9]. Studies indicate that approximately 11 % of students experience impairments such as anxiety, depression, exhaustion and burn-out-like symptoms [1,10]. Furthermore, a high level of stress can have a negative impact on academic performance, resulting in changes of study direction, prolonged studies and even dropout [11,12].

Students encounter distinct challenges during their academic journey, including the need to assimilate a substantial amount of content, effectively manage their time, cope with performance expectations, and handle exam pressure [13]. Additionally, the developmentally sensitive period associated with this age group, combined with the academic environment, can contribute to increased levels of stress [6]. Furthermore, compared to previous generations, today's students appear to exhibit lower stress tolerance and inadequate stress coping mechanisms, which further exacerbate the situation [1,14,15]. In fact, a recent study [16] reported that stress levels among students have increased by nearly 40% due to the impact of the COVID-19 pandemic.

To prevent students from experiencing chronic stress and its long-term effects, the implementation of appropriate prevention programs is crucial. These programs aim to promote students’ self- and stress-management skills, including learning and time management techniques, to effectively cope with stress despite high-performance requirements and to counteract increasing stress levels in the target group [10,17]. Studies have demonstrated the positive impact of interventions such as behavioral therapy-based approaches, relaxation and mindfulness exercises, psychoeducation, and time and study management strategies in reducing stress among students [10,18,19]. Typically, evidence-based stress management programs combine psychoeducational sessions with relaxation exercises [20–22]. Importantly, stress-management programs should be specifically tailored to the needs of students. By considering the target group’s real-life context, these programs facilitate the transfer of acquired skills into everyday life [23].

Despite the importance of stress-management programs for students, successful uptake remains challenging [24]. Unfortunately, individuals experiencing stress often do not make use of stress-management techniques for several reasons. These include the fear of being stigmatized [25], underestimation of the impact of stress, limited availability of therapy options, and high cost, particularly for young people in education [26,27].

Low-threshold, mobile health (mHealth) interventions such as smartphone apps could potentially bridge this gap. One meta-analysis [28] highlighted the advantages of apps, including location- and time-independence, reduced stigmatization and low costs [29]. Initial evidence suggests that smartphone apps can effectively reduce perceived stress, distress, depression, anxiety, and improve quality of life, psychological health, well-being, and self-regulation among student populations [30–32]. However, reported disadvantages of digital interventions, such as low adherence, legal concerns, lack of therapist relationship, and arbitrary scheduling, may diminish their effectiveness [29,33].

Conversational agents (CAs), commonly known as chatbots, are designed to simulate human-like conversations, and are increasingly utilized in clinical and non-clinical settings [34–36]. Initial findings demonstrate the feasibility, acceptance, and effectiveness of CAs in various health domains [37,38], including promoting physical activity [39], managing pain [40], reducing substance abuse [41,42], improving depression, distress and stress [43], enhancing general wellness and pain [44], and facilitating self-adherence and psychoeducation [38]. Although large language model-based CAs have recently gained increasing attention [45], they are still subject to basic research in computer science because of several severe shortcomings, such as hallucinations and non-conscious bias, among others [46]. LLM CAs are, therefore, not yet appropriate for safe and ethical delivery of several-week health interventions [47]. We therefore, decided for an established, save and transparent approach to CAs and employed a rule-based CA [39,40,48–51].

Studies investigating the effectiveness of stress management interventions delivered by a CA specifically tailored to the needs of students are still lacking. While recent studies have explored interventions such as Stressbot and CA Atena, their focus has been limited to short-term outcomes or specific topics. For instance, while Stressbot aimed to reinforce coping self-efficacy, its intervention period was only seven days [52]. Similarly, CA Atena's positive psychology and cognitive-behavioral approaches with a tailored focus on the unique needs of the COVID-19 pandemic rather than life context of students led to inconclusive outcomes regarding anxiety and stress reduction [53]. Furthermore, a previous study evaluating an AI-based chatbot that provided self-help interventions for students to reduce depression lacked detailed descriptions of evidence-based intervention designs, leaving uncertainty about the elements implemented [54]. However, evidence-based design is vital in developing CA-based coaching intervention programs [34] and stress-management interventions for specific groups such as students [23]. To our knowledge, there is no study describing the development and evaluation of the effectiveness of a CA-delivered stress-management coaching program lasting several weeks and adapted to the specific context of students in their everyday lives.

Consequently, we have developed an evidence-based, scalable, and CA-delivered stress-management coaching intervention for students called MISHA. It combines the following components: (a) providing psychoeducation about stress, mindfulness, and relaxation, (b) fostering participant motivation for self-reflection on stress and stress reactions, and (c) guiding participants in the regular practice of mindfulness and relaxation techniques. This comprehensive approach addresses key aspects of stress management, including knowledge acquisition, self-reflection, and practical application of mindfulness and relaxation techniques [19,55]. By focusing on these evidence-based intervention components, MISHA aims to empower students with effective tools and strategies to reduce stress and its long-term effects.

The goal of this pilot study was twofold: (1) to develop a scalable, evidence-based coaching intervention specifically designed for students and delivered via a CA, and (2) to assess the coaching intervention's effectiveness, engagement, and acceptance.

Methods

Intervention

App Development

MISHA was developed in collaboration with the ETH Zurich using the open-source software platform MobileCoach [56], designed for rule-based digital health interventions [48,57–59]. MISHA features a chat-based interface with multimedia elements and regular notifications to engage users. The app includes a chat channel, an audio library with relaxation exercises, psychoeducational illustrations, and FAQs (Figure 1). Communications takes place via predefined but dynamic answer options or by providing free-text input. Study participants were provided with access to a beta version of the MISHA app for Android devices through Firebase [60] and for iOS devices through TestFlight [61].

Figure 1. Screenshots of the MISHA app (coach selection, chat interface, reminder, audio library).

Coaching Concept of MISHA

The intervention concept for MISHA draws inspiration from an effective face-to-face prevention program [62], adapting its content and topics to suit a CA-delivered approach. MISHA's chat messages and notifications are aligned with the Health Action Process Approach (HAPA) model, emphasizing both motivational and volitional processes in behavior change [63].

MISHA integrates evidence-based strategies from CBT, mindfulness, and psychoeducation to provide information about stressors and coping techniques [55,64]. The stress management program includes fundamental elements derived from CBT, such as cognitive restructuring, identification, evaluation, and modification of maladaptive thought patterns [65]. In addition, techniques such as behavioral activation and activity monitoring from CBT were applied to directly support the participants in their desired goals in a collaborative approach. For further details on cognitive-behavioral techniques and session elements, please

refer to Appendix 1. The overall aim is to empower participants to reflect on their daily stressors and effectively manage their stress with new coping techniques.

Coaching content

MISHA offers a consecutive 12-session coaching program based on Gert Kaluza's stress management manual [20]. Sessions cover psychoeducation on stress, relaxation techniques, and student-specific topics such as exam anxiety. Topics are personalized, for example, setting SMART goals, individual, appointments with the CA, or selecting a CA. Participants can schedule sessions every 2 to 4 days, completing the program in 24 to 54 days (see Appendix 1 for an overview of sessions and detailed description of the content). Throughout the coaching, participants receive personalized feedback on the progression of the coaching, motivational reminders, as well as reminders in case of inactivity (please refer to Appendix 3 for detailed information on reminders). Personalization on an individual level is essential in promoting trust, engagement, adherence, and effectiveness to digital health interventions [66,67].

Study Design and Procedure

We conducted an unblinded, two-armed randomized pilot study in a population of university students in Switzerland. Study participants were allocated either to a 4 to 7 week coaching intervention or to a 40 day wait-list control group. The Cantonal Ethics Committee of Zurich (KEK-ZH, BASEC-Nr. Req-2020-01038) reviewed the research project and confirmed that the study did not fall within the scope of the Human Research Act. This research project was registered at the WHO-accredited German Clinical Trials Register (DRKS00030004). The trial was conducted following CONSORT-EHEALTH (Consolidated Standards of Reporting Trials-eHealth) guidelines. No significant content changes were made to the coaching intervention during the study period.

After downloading the MISHA app, participants were greeted and provided with information about the study procedure and coaching program. They were explicitly informed that the app does not serve as a substitute for psychotherapy and were given guidance on where to seek further help if needed. Study information was displayed within the app. To proceed, participants had to provide electronic informed consent by confirming that they had read and understood the study information. Subsequently, inclusion criteria were checked, and participants were directed to the baseline self-assessment pre-intervention (T1) using the app’s in-built LimeSurvey platform (LimeSurvey Project, Hamburg, Germany). The MobileCoach software automatically randomized participants into either the intervention or the wait-list control group by a 1:1 allocation using random numbers (0 to 1), with numbers below 0.5 assigned to the intervention group. Participants from the intervention group started the coaching program immediately. Upon program completion by a) working through all the modules, or, alternatively, b) after 54 intervention days, participants were directed to the post-intervention survey (T2) before moving to the final goodbye session. During the intervention, further self-reported outcomes were gathered (eg, usage-date, goal-achievement).

Participants from the wait-list control group received short weekly chat messages from MISHA, informing

them about the remaining duration of their wait and encouraging them to continue their participation in the study. After 40 days of waiting, they were presented with the post-intervention survey (T2) and given the opportunity to participate in the coaching program.

There was no human involvement throughout the study; however, participants had the option to contact the study team via email if they encountered technical issues or encountered problems with app download. Participants who completed the post-intervention survey had the opportunity to win a voucher worth CHF 200.00. Additionally, students of applied psychology at Zurich University of Applied Sciences ZHAW had the opportunity to earn 5 test person hours.

Recruitment

From October 6th to the end of October 2021, flyers were distributed via email to students at the University of Zurich UZH, the Zurich University of Teacher Education, FHNW School of Education, the University of Teacher Education in Special Needs Zurich, as well as the Zurich University of Applied Sciences ZHAW. Additionally, the flyer was posted on Facebook and LinkedIn. The app could be downloaded via flyer by following a weblink. Eligibility was determined within the MISHA app by self-report and included the following: (a) a minimum age of 18 years, (b) possession of and basic knowledge in the use of a smartphone, (c) sufficient knowledge of the German language, and (d) being a student at a Swiss university, university of applied sciences, university of teacher education or college of higher education.

Outcomes

Primary Outcome

To measure the effectiveness of the program, we assessed perceived stress pre- (T1) and post-intervention (T2) using the German version of the Perceived Stress Scale (PSS-10), a self-report questionnaire consisting of 10 items [68]. Participants rated their responses on a scale ranging from 0 (never) to 5 (very often).

Secondary Outcomes

We measured secondary outcomes, including depression, anxiety, somatic symptoms, and active coping pre- and post-intervention by self-report. See Appendix 3 for all outcomes and timepoints.

Depression, Anxiety, Somatic Symptoms: We used the PHQ-SADS [69] to detect depressive symptoms, anxiety, and somatic symptoms, which consists of the Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7), and the Patient Health Questionnaire (PHQ-15). The PHQ-9 is a 9-item questionnaire assessing depressive symptoms [70]. Participants rate the frequency of each symptom over the past two weeks, ranging from 0 (not at all) to 3 (nearly every day). The GAD-7 is a 7-item questionnaire that measures anxiety symptoms [71]. Participants rate the frequency of each symptom over the past two weeks, ranging from 0 (not at all) to 3 (nearly every day). The PHQ-15 is a 15-item questionnaire measuring psychosomatic symptoms. [72]. Participants rate the severity of each symptom over the previous 4 weeks, ranging from 0 (not bothered at all) to 2 (bothered a lot). For this study, items 14 (trouble sleeping) and 15

(low energy or tiredness) were collected in the PHQ-9 (similar in both questionaries) but had to be converted according to the manual [73]. By combining these individual components, the PHQ-SADS provides a comprehensive assessment of depressive symptoms, anxiety, and somatic symptoms.

Active Coping: According to the HAPA model [74], we evaluated participants' engagement in stress management activities by asking them to rate how often they had actively taken steps to reduce stress in the past 5 days. The question was assessed on a rating scale ranging from 1 (never) to 4 (regularly). This allowed us to understand the participants' level of proactive involvement in managing their stress.

Predictor

Self-Efficacy-Expectancy: Various health behavior change models, including the HAPA model [74], consider self-efficacy expectancy to be a key aspect of health behavior change. However, research findings on the impact on stress interventions are mixed [75–77]. To address this, we assessed self-efficacy expectancy using the General Self-Efficacy Scale (GSES) [78]. Pre-intervention, participants rated their agreement with statements on their ability to handle tasks effectively on a 4-point Likert scale ranging from 1 (not at all true) to 4 (exactly true). The total score of the GSES ranges from 10 to 40, with higher scores indicating higher self-efficacy.

Exploratory

Working Alliance: To assess the interaction between participants and MISHA, we employed the German version of the Working Alliance Inventory-Short Revised (WAI-SR) [79] post-intervention. This self-report questionnaire comprises 12 items that capture the quality of the therapeutic relationship and collaboration between participants and the CA via three dimensions: goal, task and bond. Responses were rated with an adapted scale from 1 (I do not agree at all) to 6 (I completely agree) post-intervention.

Subjective stress expertise and goal achievement: Throughout the coaching period, we assessed participants' goal achievement three times (session 1, 6, 11) using a scale of 1 to 10, where 1 referred to the goal as clearly not achieved and 10 referred to the goal as fully achieved. We further measured participants' stress expertise three times (session 2, 5, 13) using a similar scale, ranging from 1 (no idea how stress manifests itself in me) to 10 (I know exactly how I react when under stress).

Engagement and Acceptance

The extent to which a participant has to engage with the intervention to derive the maximum benefits is termed intended use [80]. For MISHA, we defined intended use for participants as completing the post-intervention assessment, regardless of completing all sessions. This definition was based on the fact that participants may have varied goals and desired outcomes, leading to differences in their use of MISHA's features, including frequency and duration [81,82]. It also implies that participants do not necessarily need to interact with all available intervention components. Furthermore, participants might discontinue using the intervention upon achieving their personal goals, indicating that non-usage is not due to loss of interest

[83,84]. Additionally, we ground this approach on the self-determination theory, where autonomy by providing choice is essential [85].

To assess participants' engagement in the coaching program, we analyzed usage data from the intervention group by calculating the ratio of replied conversational turns based on the number of messages sent by MISHA in relation to messages replied by participants. We also tracked the number of sessions completed by participants, and the number of reminders sent to participants in cases of inactivity (ie, if participants stopped interacting during a session). Additionally, we tracked the number of minutes of audio files played by participants throughout the intervention.

We evaluated the feasibility and acceptance of MISHA using the Mobile App Rating Scale (uMARS) [86] post-intervention. The uMARS is a validated questionnaire that assesses the dimensions engagement, functionality, aesthetics, information, perceived quality, and perceived impact. All subscales use a 5-point Likert scale ranging from 1 to 5, where higher scores indicate a more favorable judgment. In this study, a total of 19 items were translated from English to German to assess engagement (entertainment, interest, customization, interactivity, and target group of the app), information (quality of information, quantity of information, visual information, credibility of source), perceived quality (recommendation, usage, payment, overall rating), perceived impact (awareness, knowledge, attitudes, behavior change, seeking help, and intention to change). In addition to the uMars, participants had the opportunity to provide feedback in free text prompted by the following questions: "What did you like most about the MISHA app?" and "What would you improve in the MISHA app?".

Sample Size Calculation

The sample size was estimated for a Generalized Estimating Equation (GEE) based on a repeated-measures (within-between interaction) analysis of variance (ANOVA). A small to medium time-by-group interaction effect size (f=.15) for the primary outcome perceived stress due to prior results [87] was expected. The G*Power analysis [88] revealed that a sample size of 90 participants would be sufficient with a power of .80 and a correlation of 0.5 between measurements. Due to the high percentage of dropouts observed in earlier studies, the target sample was increased to 180 participants [89].

Data Analysis

Descriptive statistics, independent t-tests, and chi-square tests were conducted to analyze baseline differences in demographics and outcomes between the intervention and control group.

In our analysis, we examined the effectiveness of the intervention by assessing changes in the primary outcome perceived stress scores over time within each group (intervention and control) and comparing these changes between groups. We first conducted a per-protocol analysis (PP), including only participants who completed both surveys. This was done by using a repeated measure analysis of variance (rmANOVA) with perceived stress as the dependent variable, time as within-subject factor, and group as between-group factor.

Secondary outcomes depression, anxiety, psychosomatic symptoms, and active coping were analyzed accordingly.

In compliance with CONSORT guidelines, we also conducted an intention-to-treat analysis (ITT) wherein all randomized participants were included, regardless of their adherence to the coaching intervention. This analysis was performed using general estimating equation (GEE). In module 1, we conducted an unadjusted evaluation with time (T1 and T2), group (intervention and control), and treatment (group by time interaction) as independent variables, with perceived stress as dependent variable. The incorporation of time allows the examination of the dependent variable stress over different time points, group allows for comparison of stress between groups, and the interaction between group and time allows for an examination of whether the changes in outcomes over time differ between the intervention and control groups. In Model 2, we did an adjusted analysis with the inclusion of the covariate general self-efficacy for the primary outcome perceived stress. The same independent variables were considered as in Model 1. Secondary outcomes were evaluated accordingly. A log link function, gamma distribution, and unstructured covariance structure were applied. This modeling approach provided the best fit with the outcomes and allowed us to avoid restrictions on the covariance structure. To reduce the impact of influential observations and outlier effects, we employed a robust estimator which is consistent with standard procedures when using GEE.

Using GEE [90] offered several advantages. Firstly, it allowed us to consider the correlations between the measurement times in longitudinal data, which is important for analyzing repeated measures. Additionally, GEE allowed us to include incomplete data sets by using an estimating equation to handle missing data. GEEs use all available data and estimate missing outcome values under the assumption of Missing Completely at Random (MCAR). To assess the assumption of MCAR, we conducted Little’s MCAR Test. Calculation of between-group effect sizes (Cohen's d) were based on the pooled standard deviation and labeled as small (d=0.2), medium (d=0.5) and large (d=0.8). Further, we explored the potential relation of working alliance and perceived impact on treatment outcomes using a correlation. All statistical analyses were performed using SPSS (version 28). We applied qualitative content analysis [91,92], using thematic maps [93] to answer the open-ended questions.

Results

Demographics and Baseline Scores

In total, 230 individuals downloaded the app. Of these, 148 (64.3%) were assessed for eligibility and completed the baseline survey. Before randomization, 8 (3.5%) participants discontinued using the app and 140 (60.9%) were randomized into intervention (n=70) and wait-list control (n=70) group. The full participant flow is depicted in Figure 2.

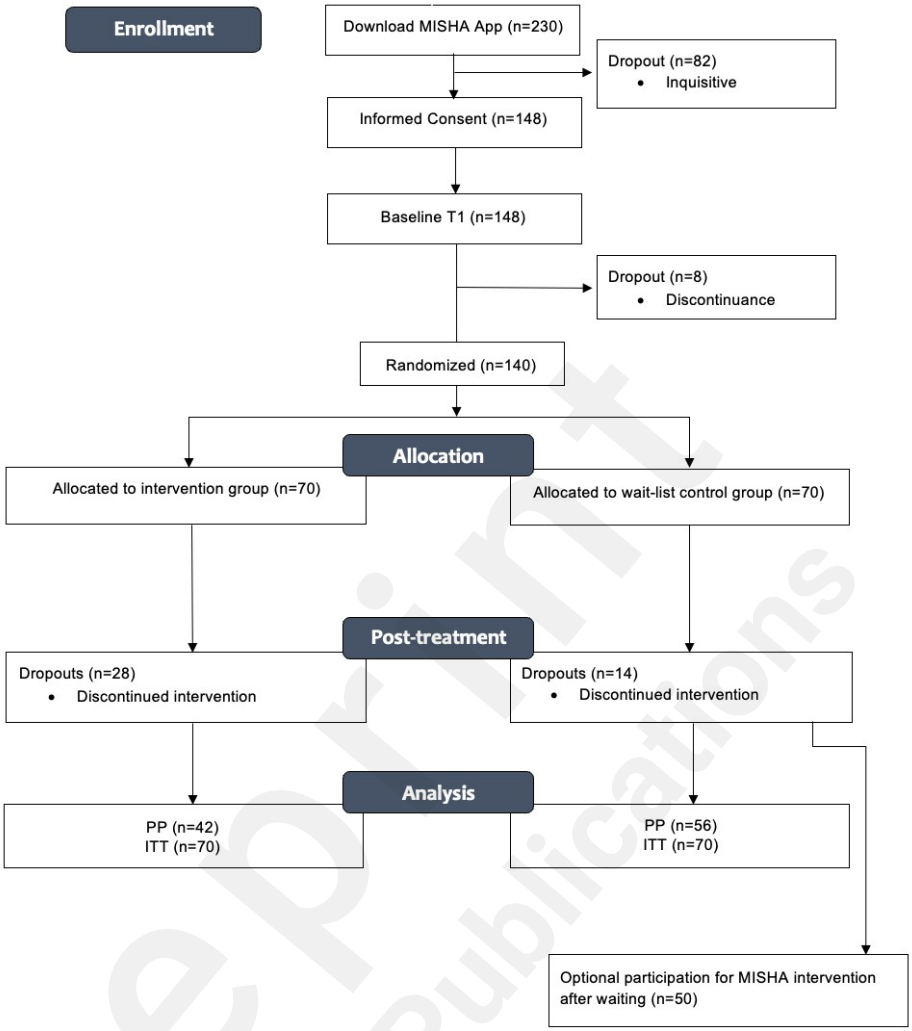


Figure 2. Study flow chart.

Participants had a mean age of 26.71 years (SD 6.29), while 23.6% (33/140) identified as male, 103/140 (73.6%) as female, 3/140 (2.1%) as non-binary, 1 person declined to provide information about their gender (see Table 1). Regarding relationship status, 83/140 (59.3%) participants reported being married or in a relationship, while 57/140 (40.7%) were single. Regarding educational background, the majority of participants (90/140, 64.3%) had an apprenticeship/vocational or high school diploma. A significant proportion (37/140, 26.4%) had a university degree at Bachelor level or higher vocational education or training, while 12/140 (8.6%) had other qualifications. In terms of their field of study, most participants (131, 93.6%) were studying at a university of applied sciences or university, while 7 (5.0%) were studying at other institutions. Among the participants, 124 (88.6%) had a degree in (applied) psychology, social sciences (6, 4.4%), or other fields (5.1%). There were no differences between groups for any of the outcomes at baseline.

Table 1. Sample description at baseline.

Outcome	Control (n=70)	Intervention (n=70)	*P* value[a]
Age (years), mean (SD)	26.21 (5.56)	27.22 (6.96)	**.75**
Gender, n (%)			**.78**

Male	17 (24.3)		
Female	52 (74.3)	51 (49.5)	
Non-binary	1 (1.4)	2 (66.7)	
Highest Education, n (%)			**.78**
Apprenticeship/vocational or high school diploma	47 (67.1)	43 (61.4)	
Higher vocational education and training HFH	6 (8.6)	7 (10.0)	
Degree at BSc level University/PH/FH	17 (24.3)	20 (28.6)	
Relationship status, n (%)			**.86**
Single	29 (41.4)	28 (40.0)	
Married or in relationship	41 (58.6)	42 (60.0)	
Study institute, n (%)			**.39**
University of Applied Science FH	67 (95.7)	64 (91.4)	
University/Swiss Federal Institute of Technology ETH	3 (4.3)	4 (5.7)	
University of Education PH	0 (0.0)	1 (1.4)	
Others	0 (0.0)	1 (1.4)	
Study subject, n (%)			**.33**
Applied Psychology	63 (92.6)	61 (88.4)	
Social Work	0 (0.0)	2 (2.9)	
Information (Technology)	1 (1.5)	0 (0.0)	
Economics/Business	1 (1.5)	1 (1.4)	
Pedagogy	0 (0.0)	1 (1.4)	
Natural/Earth Sciences	0 (0.0)	1 (1.4)	
Social sciences	3 (4.4)	3 (4.3)	
Outcomes, mean (SD)			
Perceived stress (PSS-10)	28.79 (5.27)	28.40 (5.45)	.67
Depression (PHQ-9)	8.16 (4.57)	7.83 (4.16)	.66
Anxiety (GAD-7)	6.84 (4.05)	6.69 (3.77)	.81
Psychosomatic symptoms (PHQ-15)	9.26 (4.09)	8.87 (4.39)	.59
Self-Efficacy (GSES)	29.09 (3.36)	29.21 (2.86)	.81
Active coping (HAPA)	2.43 (.79)	2.29 (.85)	.31

[a]Baseline group comparison between intervention vs. wait-list control group with *t* test or chi-square.

Effectiveness

To evaluate the effectiveness of the intervention and to take missing data into account, a PP analysis of the time by group interaction was carried out followed by an ITT analysis. For the PP analysis (Table 2), we found evidence for a treatment effect (group by time interaction) from pre- to post-intervention between the intervention and control group for stress (*P*=.001; *d*=–0.60), depressive- (*P*=.003; *d*=–0.50), and psychosomatic symptoms (*P*=.010; *d*=–0.36), but not for anxiety and active coping behavior.

Table 2. Pre- and post-intervention means, results of the per-protocol (PP) repeated measure ANOVA analysis, between group effect sizes (Cohen *d*).

	Pre-intervention	Post-intervention	Between-group effects (intervention vs. wait-list control group post-intervention)

Measure	M (SD)	M (SD)	Cohen *d*[a]	95% CI[b]	Partial η^2	ANOVA *F* test (*df*)	*P* value
Primary outcome							
Perceived Stress (PSS-10)							
Intervention (n=42)	28.41 (5.53)	24.24 (5.93)	–0.60	–1.01 to –0.19	0.10	10.69 (1, 96)	.001
Control (n=56)	28.36 (4.93)	27.61 (5.38)					
Secondary outcomes							
Depression (PHQ-9)							
Intervention (n=42)	7.90 (4.24)	5.95 (3.45)	–0.50	–0.91 to –0.10	0.09	9.29 (1, 96)	.003
Control (n=56)	7.86 (4.13)	7.86 (4.02)					
Anxiety (GAD-7)							
Intervention (n=42)	6.52 (3.69)	5.62 (3.22)	–0.29	–0.69 to 0.11	0.03	3.18 (1, 96)	.08
Control (n=56)	6.41 (3.32)	6.59 (3.47)					
Somatic symptoms (PHQ-15)							
Intervention (n=42)	9.19 (4.81)	7.50 (3.78)	–0.36	–0.76 to –0.04	0.07	6.92 (1, 96)	.01
Control (n=56)	9.07 (3.89)	9.00 (4.43)					
Active Coping (HAPA)							
Intervention (n=42)	2.21 0.87)	2.67 (0.75)	0.13	–0.27 to 0.53	0.04	3.60 (1, 96)	.06
Control (n=56)	2.45 (0.81)	2.57 (0.78)					

[a]Cohen *d* based on means and the pooled standard deviation of the PP analysis
[b]95% CI of Cohen *d* (between groups, post-intervention)

In the ITT analysis for the unadjusted model (Model 1), we found evidence for a treatment effect (group by time interaction) from pre- to post-intervention between the intervention and control group for stress (*P*<.001), depressive symptoms (*P*=.003), and psychosomatic symptoms (*P*=.003). No treatment effect was found for anxiety *P*=.13), and active coping (*P*=.09).

After adjusting for the covariate self-efficacy expectancy (Model 2), we found evidence for treatment effects similar to Model 1, please refer to Table 3. Further, there was evidence for an effect of self-efficacy expectancy on perceived stress (*P* <.001), depression (*P*<.001), anxiety (*P*<.001), and psychosomatic symptoms (*P*<.001), but not on active coping.

Table 3. Results of the outcome intention to treat (ITT) analysis (model 1), including self-efficacy as covariate (model 2), using generalized estimating equations.

Outcome	Model	Model

	1[a] Estimate β	SE	*P* value	95% CI	2[b] Estimate β	SE	*P* value	95% CI
Perceived Stress (PSS-10)								
Intercept	3.36	N/A	N/A	N/A	4.18	N/A	N/A	N/A
Time[c]	–0.03	0.02	.17	–0.08 to 0.05	–0.04	0.02	.12	–0.08 to 0.01
Group[d]	–0.13	0.03	.69	–0.05 to 0.08	–0.01	0.03	.75	–0.07 to 0.05
Treatment[e]	–0.13	0.04	<.001	–0.20 to –0.05	–0.12	0.04	.001	–0.19 to –0.04
Self-efficacy					–0.03	0.01	<.001	–0.04 to –0.02
Depression (PHQ-9)								
Intercept	2.22	N/A	N/A	N/A	3.98	N/A	N/A	N/A
Time	–0.01	0.05	.83	–0.11 to –0.09	–0.20	0.05	.69	–0.12 to 0.08
Group	–0.04	0.08	.65	–0.20 to 0.12	–0.01	0.08	.87	–0.16 to 0.14
Treatment	–0.23	0.08	.003	–0.38 to –0.08	–0.21	0.07	.006	–0.35 to –0.06
Self-efficacy					–0.06	0.01	<.001	–0.08 to –0.04
Anxiety (GAD-7)								
Intercept	2.06	N/A	N/A	N/A	3.71	N/A	N/A	N/A
Time	–.00	0.06	.99	–0.12 to 0.12	–0.00	0.06	.94	–0.12 to 0.11
Group	–.02	0.08	.81	–0.18 to 0.14	–0.01	0.08	.91	–0.17 to 0.14
Treatment	–.14	0.09	.13	–0.31 to 0.04	–0.11	0.09	.22	–0.28 to 0.06
Psychosomatic symptoms (PHQ-15)								
Intercept	2.33	N/A	N/A	N/A	3.90	N/A	N/A	N/A
Time	–0.01	0.04	.77	–0.08 to 0.61	–0.01	0.04	.78	–0.08 to 0.06
Group	–0.04	0.07	.60	–0.18 to 0.11	–0.03	0.07	.68	–0.17 to 0.11
Treatment	–0.16	0.06	.003	–0.27 to –0.06	–0.15	0.06	.007	–0.26 to –0.04
Self-efficacy					–0.06	0.01	<.001	–0.08 to –0.03
Active Coping (HAPA)								
Intercept	0.89	N/A	N/A	N/A	0.69	N/A	N/A	N/A
Time	0.05	0.05	.23	–0.03 to 0.14	0.05	0.05	.24	–0.04 to 0.14
Group	–0.06	0.06	.28	–0.17 to 0.05	–0.06	0.06	.26	–0.17 to 0.05
Treatment	0.11	0.07	.09	–0.02 to 0.25	0.12	0.07	.09	–0.02 to 0.25
Self-efficacy					0.01	0.01	.39	–0.01 to 0.02

[a]Model 1 = Unadjusted model (without covariate).
[b]Model 2= Adjusted model for general self-efficacy expectance (GSES), mean =29.18.
[c]Time effect represents the rate of improvement for both intervention and wait-list control groups.
[d]Group effect represents intervention or wait-list control group.
[e]Treatment effect is represented by group and time interaction.

Exploratory

In terms of working alliance, participants in the intervention group reported a mean working alliance score of

4.23 (SD 0.89) post-intervention. When exploring the potential influence of the working alliance on changes in outcomes from pre- to post-intervention, we did not find evidence for correlations on any of the outcomes (Pearson correlation *r* ranging from –.021 to .223). The participants rated their subjective stress expertise and goal achievement throughout the coaching program (3 times). For goal achievement, we observed a significant increase from the first to the third 6 measurement (with a large effect (*d*=–1.07,). Refer to table 4 for further details.

Table 4. Means for subscales bond, task and goal of working alliance, results of a paired *t* test for stress expertise and goal achievement.

		Start of the intervention, mean (SD)	End of the intervention, mean (SD)	*t* test	*P* value[b]	Cohen *d* (95% CI)
WAI-SR[a]						
	Total	N/A	4.23 (0.89)	N/A	N/A	N/A
	Bond	N/A	4.20 (1.01)	N/A	N/A	N/A
	Task	N/A	4.18 (0.82)	N/A	N/A	N/A
	Goal	N/A	4.30 (0.84)	N/A	N/A	N/A
Stress (n=45)	expertise	7.51 (1.47)	7.64 (1.60)	.47	.64	–0.07 (–0.36 to 0.22)
Goal (n=24)	achievement	3.88 (2.54)	6.71 (2.14)	–5.24	<.001	–1.07 (–1.57 to –0.56)

[a]Working Alliance Inventory (WAI-SR) with Likert scale from 1 to 7.

[b]Within group comparison: Start of the intervention versus end of intervention.

Engagement and Acceptance

In the intervention group, 60% (42/70) finished the coaching program by completing the post-intervention survey (completers) and used the intervention as intended. Although the Little's test indicated that missing were completely at random (MCAR) for perceived stress (χ^2=0.53; *P*=.47), depression (χ^2=0.23; *P*=.63), anxiety (χ^2=2.02; *P*=.16), psychosomatic symptoms (χ^2=0.6; *P*=.80), and active coping (χ^2=0.05; *P*=.82), we conducted a dropout analysis due to the potential risk of differential attrition, particularly with significantly higher dropouts observed in the intervention group [94]. The analysis revealed no significant differences in outcomes (eg, stress, depression) or demographics (ie gender, age) between completers and dropouts.

Overall, 45% of the completers (19/42) worked through all 13 sessions, played 86.52 (SD 120.54) minutes of relaxation audios, and received 115.88 (SD 5.06) reminders, please see Table 4 for further information. On average, MISHA sent 400 messages (SD 205.61) and participants answered a mean of 297 messages (mean 296.54, SD 169.80), resulting in an average engagement ratio of 74.3%.

Table 5. Indicators of engagement: Intended use, session completion, relaxation and reminders.

Indicators of engagement	n	Relaxation applied, mean	Reminders received, mean (SD)[a]

		(SD)	
Completers (intended use[b)]	42	86.52 (120.54)	15.88 (5.06)
Completed all sessions	19	97.11 (71.43)	14.89 (5.75)
Stopped interacting after session 12	3	299.33 (377.33)	17.67 (1.15)
Stopped interacting after session 11	3	29.00 (26.91)	17.67 (0.58)
Stopped interacting after session 10	1	12.00 (0.0)	27.00 (0.0)
Stopped interacting after session 9	2	46.00 (26.87)	21.00 (1.41)
Stopped interacting after session 8	4	53.50 (42.45)	16.00 (3.92)
Stopped interacting after session 7	3	89.33 (82.25)	18.00 (1.73)
Stopped interacting after session 6	3	56.00 (73.53)	17.00 (1.0)
Stopped interacting after session 5	1	26.00 (0.0)	14.00 (0.0)
Stopped interacting after session 4	1	16.00 (0.0)	9.00 (0.0)
Stopped interacting after session 3	2	4.00 (4.24)	8.50 (2.12)
Stopped interacting after session 2	0	n/a	n/a
Stopped interacting after session 1	0	n/a	n/a

[a]Reminders in case of inactivity during sessions.

[b]Intended use is defined by completing the post-intervention survey, regardless of number of sessions that were completed.

The participants in the intervention group rated the subscale information highest, with a mean of 4.26 (SD 0.46), followed by engagement (mean 3.42, SD 0.70), perceived impact (mean 3.35, SD 0.87) and subjective quality of the app (mean 2.99, SD 0.87). Regarding engagement, individual customization was rated lowest with a mean of 2.71 (SD 0.84), while the target group fit was perceived as high (mean 3.95, SD 0.90). The participants liked the visual information of the CA and rated it high regarding correctness, clarity, and logic (mean 4.45, SD 0.55). Only a few participants showed high willingness to pay for the app (mean 2.10, SD 0.91), or anticipated high future use (mean 2.98, SD 1.05). The recommendation of the app to others was good, with a mean of 3.43 (SD 1.19) within the subjective app quality scale.

Qualitative feedback

The participants in the intervention group had the opportunity to provide free-text responses regarding their positive feedback on the CA intervention (Figure 4) and suggestions for improvement (Figure 5). The number of responses is displayed within the circles.

Figure 3. Thematic map of positive participant feedback

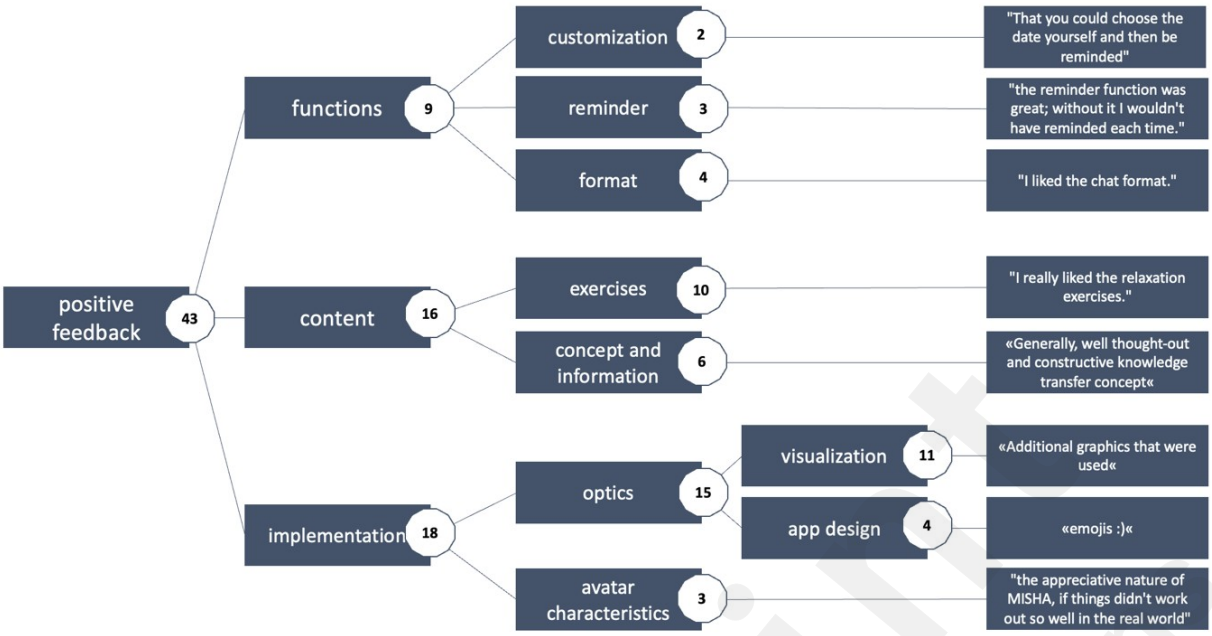
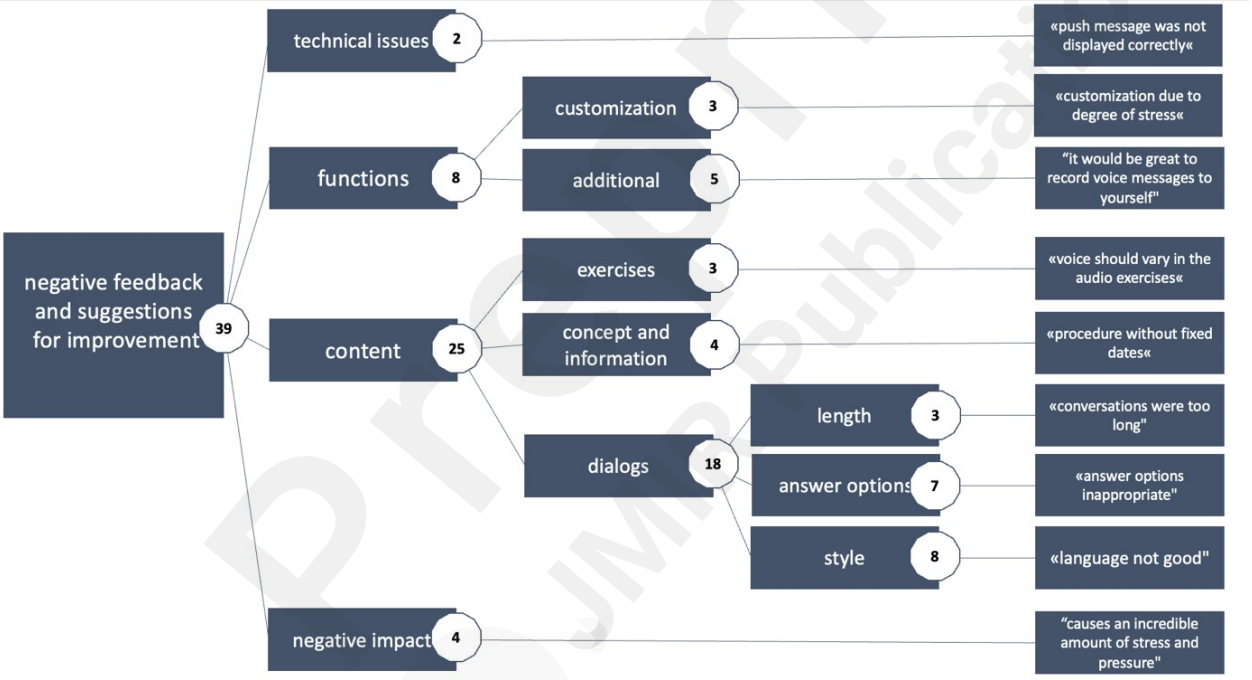


Figure 4. Thematic map from negative feedback of participants



Discussion

The aim of this study was to describe the development and evaluation of the effectiveness of the MISHA app, a rule-based CA-delivered stress-management coaching intervention specifically tailored to the living environment of students. We described the MISHA app' s evidence-based design and systematic evaluation. In both the PP and ITT analyses, we found evidence for decreased stress levels among participants in the intervention group compared to the control group, with a medium-large between-group effect (PP: d=–0.60). Additionally, we observed evidence for a reduction in depressive symptoms with a medium-large effect (d=–0.50), as well as in psychosomatic symptoms with a small to medium effect (d=–0.36), while anxiety and active coping did not change. In the ITT analyses, a weak relation was found between self-efficacy and

perceived stress, depression, anxiety, and psychosomatic symptoms, while the treatment effect persisted for stress, depression, and psychosomatic symptoms.

Our findings are consistent with other studies evaluating CA effectiveness in non-clinical populations. For instance, a study on CA Shim [95] among stressed young adults, despite small sample size, reported stress reduction and improved psychological well-being, mirroring our results. Another study [52] assessed Stressbot, a 7-day Messenger CA intervention aimed at enhancing coping self-efficacy among university students. Initial results showed reduced stress levels and improved self-efficacy post intervention. A large single-arm study evaluated Viki, an instant-messenger platform-based intervention [96], and found reduced stress and depressive symptoms. However, unlike out study, they reported a significant decrease in anxiety. In our study, the concurrent pandemic situation, or upcoming exams may have triggered increased uncertainties and fears. In a study involving CA Atena [53], the overall reduction in anxiety and stress levels may not have been substantial, but the intervention showed promise in supporting individuals with high stress levels during the Covid pandemic. Another study evaluated an AI driven CA with the aim to reduce depression in university students by reflecting their emotions, thoughts, and behavior [54]. Authors found reduced levels of depression and anxiety in the intervention group.

With its strong focus on goal setting, a crucial element in coaching [97], and being based on a behavior change model [74], the MISHA coaching intervention appears to effectively help students to manage their stress. Towards the end of the coaching program, participants significantly rated their goal achievement higher with a large effect (d=–1.07), indicating the intervention's effectiveness in this regard. However, some participants expressed a desire for customization options, particularly regarding stress levels.

Regarding evidence from mHealth interventions for students, one study [98] found positive effects on stress and overall well-being in a 30-day app-based intervention on stress management through mindfulness meditation among medical students. A systematic review confirmed that digital interventions for the enhancement of mental well-being among college students can be effective in improving depression, anxiety, and mental well-being [99].

Given the mixed findings regarding the impact of self-efficacy expectancy on stress interventions targeting students [75–77], we explored whether self-efficacy was related to perceived stress. We found only a weak relation, while the treatment effectiveness remained unchanged. Therefore, in this study, self-efficacy does not seem to have influenced the treatment's effectiveness in reducing stress.

In line with other studies [100–102], participants formed a working alliance with CA MISHA. Qualitative analyses revealed participants’ appreciation for MISHA’s supportive nature, especially during challenging moments. Most participants enjoyed interacting with MISHA, found the information provided appropriate, and expressed increased intention to change their behavior related to stress. Some desired for additional features (eg, voice recording), found answer options or language style to be inappropriate, and disliked the lengthy dialogs. The various exercises, reminders and visualizations were perceived as positive, and the

constructive knowledge transfer was appreciated. In summary, it appears that a CA could be a well-accepted medium for stress prevention measures among students.

Limitations

This study has several limitations. Firstly, despite statistically significant findings, it is essential to recognize that the absolute improvement in perceived stress, depressive, and psychosomatic symptoms was small. Yet, these improvements may still hold clinical relevance, and students experiencing even slight relief from perceived tress can benefit from CA-based coaching. Medium effect sizes indicate practical significance but may not always translate into substantial clinical change, and results should be interpreted with caution and in light of the context. Further, all participants were self-selected, which limits the generalizability of our findings and introduces the potential for self-selection bias. Participants' may have a particular interest in the subject and therefore cannot be considered a representative sample. It is important to note that their pre-existing characteristics may differ from those in the broader population, and caution should be used when generalizing these findings to a wider context. Further, this study is based on a convenience sample and should not be considered representative of all students. In particular, our sample, with the majority studying psychology (90.5%) and predominantly female participants (73.6%), does not accurately reflect the student population in Switzerland, which shows an approximately even gender distribution (53% female) [103]. Questions therefore remain regarding the accessibility of the intervention to individuals who may not have an interest in psychological content and whether men and women can be equally reached by a mindfulness-focused chatbot like MISHA.

Second, regarding engagement, we have analyzed usage data from the intervention group, including completion rates, session completion, message response rate, reminders, and media player for relaxation). These objective measures offer valuable insights into participants' interactions with the coaching program and help ensure the robustness of our findings. However, it is difficult to measure how devoted participants were when using the app. To date, there is no consensus on measuring engagement in digital interventions [81]. According to Perski et al. (2017) [104], engagement can be defined as a multidimensional construct that can be measured using self-reported outcomes, usage data, or even psychophysical parameters. Future research should assess participants' time and motivation on offline engagement with exercises, and aspects of attention, interest, and affects should be considered. Furthermore, in-depth usage data should be gathered to assess the association between engagement, effectiveness, and optimal intended use.

Third, in this study, participants established a working alliance with the CA. However, it is important to acknowledge that CAs lack human-like empathy or emotions [105]. They may struggle to understand the nuances of human language and lack the emotional intelligence and personal experience of a human, even if they can express empathy-like utterances. A recent study demonstrated that human-AI collaboration outperformed humans, leading to a 19.6 % increase in empathy in peer-to peer text-based mental health support conversations [106]. While AI can mimic empathy and generate appropriate responses in text-based

conversations, it's important to remember that these are still artificial constructs.

Fourth, various technical limitations need to be listed. At the beginning of the intervention, there were technical difficulties related to the audio files of the relaxation exercises. Some exercises could not be played. Additionally, several participants indicated that the app was not updating properly; however, this issue was resolved within a few days. Furthermore, there was a 2-day interruption at the beginning because a technical adjustment had to be made to ensure that the system could recognize completed sessions. It remains unclear whether these technical issues led to more dropouts, frustration, or non-use of the exercises. Unfortunately, the recording of the minutes of listened audio files did not function flawlessly. While audio minutes were measured, they must be interpreted with great caution due to uncertainty in measurement. Additionally, if the display of push notifications on the mobile phone was not set as the default, some messages were displayed without text. The number of people for whom this was the case and whether it negatively affected adherence cannot be conclusively determined. Any reported bugs in MISHA were addressed by a member of the study team within a 24-hour timeframe. There were no reported instances of server downtime.

Fifth, it is important to recognize the potential for improvements to enhance interaction in MISHA. The nature of the current CA is rule-based: while allowing for evidence-based program development, the flexibility of interaction is limited by pre-defined answer options. While participants appreciated various aspects such as visualization, reminders, or exercises, personalized input via text input was missing and some answer options were perceived as inappropriate. AI-based technology such as large language models (LMMs) or natural language processing (NLP) could be taken into consideration to improve text processing in MISHA. NLP and LMM enable the CA to interpret user inputs more dynamically with increased natural interactions [107,108]. AI-based CAs are increasingly applied in health care to provide education and disease-management. The literature on AI-based CAs indicates high overall performance and satisfactory user experience, high engagement, and positive health-related outcomes [109]. However, to date, CA intervention in the field of mental health are almost entirely rule-based [110]. Ethical considerations concerning AI technology should be addressed to mitigate potential misjudgments and risks. Research highlights the critical issue of inadequate transparency in data input and algorithms, undermining the reliability and validity of results [109,111]. Currently, both rule-based and LLM-based CAs are suitable to administer script-based interventions such as CBT elements like psychoeducation, goal-setting, or reflective tasks. While in future, LLM-based interventions may be able to deliver more complex interventions in the field of psychology, it is crucial to consider potential risk and limitations of implementing these technologies [112].

Sixth, it is important to acknowledge the possibility of a digital placebo effect [113]. In an unblinded trial, participants might attribute their improvements to the mere use of an mHealth intervention, rather than its specific components. Expectations and engagement could introduce positive bias into the outcomes. Future research should carefully plan control conditions, which might include active control groups or sham interventions [139].

Conclusions

This paper outlines the evidence-based development of MISHA, a scalable coaching intervention specifically designed for students in their everyday life. The results of this study confirmed that a CA-based coaching can be successfully delivered and is effective in reducing stress in students. It could not be confirmed that self-efficacy is related to the treatment effect. The establishment of a strong working alliance between participants and the CA, along with their perceived goal-achievement, further reinforces the potential effectiveness of this intervention. Future research should analyze effectiveness over time, use active control groups, and enhance user interaction. In sum, providing psychoeducation about stress, mindfulness, and relaxation seems to empower students with effective tools and strategies to reduce stress.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

TK is affiliated with the Centre for Digital Health Interventions (CDHI), a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich, the Department of Management, Technology, and Economics at ETH Zurich, and the Institute of Technology Management and School of Medicine at the University of St. Gallen. CDHI is funded in part by CSS, a Swiss health insurer, Mavie Next, an Austrian health insurer, and MTIP, a Swiss digital health investor. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies was involved in this research.

Abbreviations

CA: Conversational Agent
CBT: Cognitive Behavior Therapy
DRKS: German Clinical Trial Register
GEE: Generalized Estimating Equations
HAPA: Health Action Process Model
LMMs: Large Language Models
mHealth: mobile health
NLP: Natural Language Processing
ITT: Intention-to-treat
PP: Per-protocol

Multimedia Appendix

Multimedia Appendix 1: Overview coaching intervention
Multimedia Appendix 2: Reminder Escalation
Multimedia Appendix 3: Outcome and timepoints

Literature

1. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, Sammut S. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. Journal of Affective Disorders 2015 Mar;173:90–96. doi: 10.1016/j.jad.2014.10.054
2. Löwe, B, Spitzer, R. L., Zipfel, S., Herzog, W. PHQ-D Gesundheitsfragebogen für Patienten. Manual. Komplettversion und Kurzform. 2002. Available from: https://www.klinikum.uni-heidelberg.de/fileadmin/Psychosomatische_Klinik/download/PHQ_Kurzanleitung1.pdf
3. American Psychological Association. Stress in America: On second COVID-19 anniversary, money, inflation, war pile on to nation stuck in survival mode. Available from: https://www.apa.org/news/press/releases/stress/2022/march-2022-survival-mode [accessed Aug 8, 2022]
4. Bundesamt für Statistik. Gesundheit der Studierenden an den Schweizer Hochschulen. Themenbericht der Erhebung 2016 zur sozialen und wirtschaftlichen Lage der Studierenden. 2018. Available from: https://www.bfs.admin.ch/de/home/statistiken/bildung-wissenschaft/personen-ausbildung/soziale-wirtschaftliche-lage-studierenden.assetdetail.6526111.html [accessed Dec 17, 2023]
5. Hofmann F-H, Sperth M, Holm-Hadulla RM. Psychische Belastungen und Probleme Studierender: Entwicklungen, Beratungs- und Behandlungsmöglichkeiten. Psychotherapeut 2017 Sep;62(5):395–402. doi: 10.1007/s00278-017-0224-6
6. Pogrebtsova E, Craig J, Chris A, O'Shea D, González-Morales MG. Exploring daily affective changes in university students with a mindful positive reappraisal intervention: A daily diary randomized controlled trial. Stress and Health 2018 Feb;34(1):46–58. doi: 10.1002/smi.2759
7. Shaffique S, Farooq SS, Anwar H, Asif HM, Akram M, Jung SK. Meta-analysis of Prevalence of Depression, Anxiety and Stress Among University Students. RADS J Biol Res Appl Sci 2020 Sep 21;11(1):1–6. doi: 10.37962/jbas.v11i1.308
8. American Psychiatric Association, editor. Diagnostic and statistical manual of mental disorders: DSM-5. 5. Aufl. Washington, D.C: American Psychiatric Association; 2013. ISBN:978-0-89042-554-1
9. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, Demyttenaere K, Ebert DD, Green JG, Hasking P, Murray E, Nock MK, Pinder-Amaker S, Sampson NA, Stein DJ, Vilagut G, Zaslavsky AM, Kessler RC, WHO WMH-ICS Collaborators. WHO world mental health surveys international college student project: Prevalence and distribution of mental disorders. Journal of Abnormal Psychology 2018 Oct;127(7):623–638. doi: 10.1037/abn0000362
10. Gusy B, Lesener T, Wolter C. Burnout bei Studierenden. PiD - Psychotherapie im Dialog 2018 Sep;19(03):90–94. doi: 10.1055/a-0556-2588
11. Eicher V, Staerklé C, Clémence A. I want to quit education: A longitudinal study of stress and optimism as predictors of school dropout intention. Journal of Adolescence 2014 Oct;37(7):1021–1030. doi: 10.1016/j.adolescence.2014.07.007
12. Middendorff E, Apolinarski B, Becker K, Bornkessel P, Brandt T, Heissenberg, S, Naumann, H, Poskowsky, J. Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2016. Bundesministerium für Bildung und Forschung BMBF; 2016. Available from: https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/4/31338_21_Sozialerhebung_2016_Zusammenfassung.pdf?__blob=publicationFile&v=3
13. Kumaraswamy N. Academic stress, anxiety and depression among college students - a brief review.

International Review of social Sciences and Humanities 2013;1(5):135–143.

14. Bland HW, Melton BF, Welle P, Bigham L. Stress tolerance: New challenges for millennial college students. College Student Jouranl 2012;46(2):362–375.

15. Mackenzie S, Wiegel JR, Mundt M, Brown D, Saewyc E, Heiligenstein E, Harahan B, Fleming M. Depression and suicide ideation among students accessing campus health care. The American Journal of Orthopsychiatry 2011;(81):101–107. doi: https://doi.org/10.1111/j.1939-0025.2010.01077.x

16. Ehrentreich S, Metzner L, Deraneck S, Blavutskaya Z, Tschupke S, Hassler M. Einflüsse der Coronapandemie auf gesundheitsbezogene Verhaltensweisen und Belastungen von Studierenden. Prävention und Gesundheitsförderung 2021 Aug 23; doi: 10.1007/s11553-021-00893-2

17. Ackermann E, Schumann W. Die Uni ist kein Ponyhof: Zur psychosozialen Situation von Studierenden. Praev Gesundheitsf 2010 Aug;5(3):231–237. doi: 10.1007/s11553-010-0234-5

18. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, Wong CC. Efficacy and moderation of mobile app–based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: Randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60. doi: 10.2196/mental.8597

19. Yusufov M, Nicoloro-SantaBarbara J, Grey NE, Moyer A, Lobel M. Meta-analytic evaluation of stress reduction interventions for undergraduate and graduate students. International Journal of Stress Management 2019 May;26(2):132–145. doi: 10.1037/str0000099

20. Kaluza G. Stressbewältigung: Trainingsmanual zur psychologischen Gesundheitsförderung. Berlin, Heidelberg: Springer; 2018. doi: 10.1007/978-3-662-55638-2ISBN:978-3-662-55637-5

21. Meichenbaum DH. Intervention bei Stress. Anwendung und Wirkung des Stressimpfungstrainings. 2. Aufl. Bern: Huber; 2003.

22. Reschke K, Schröder H. Optimistisch den Stress meistern: ein Progamm für Gesundheitsförderung, Therapie und Rehabilitation. 2., überarb. und erw. Aufl. Tübingen: Deutsche Gesellschaft für Verhaltenstherapie; 2010. ISBN:978-3-87159-316-1

23. Kaluza G. Psychologische Gesundheitsförderung und Prävention im Erwachsenenalter: Eine Sammlung empirisch evaluierter Interventionsprogramme. Zeitschrift für Gesundheitspsychologie 2006 Oct;14(4):171–196. doi: 10.1026/0943-8149.14.4.171

24. Marsh CN, Wilcoxon SA. Underutilization of Mental Health Services Among College Students: An Examination of System-Related Barriers. Journal of College Student Psychotherapy 2015 Jul 3;29(3):227–243. doi: 10.1080/87568225.2015.1045783

25. Song X, Anderson T, Himawan L, McClintock A, Jiang Y, McCarrick S. An Investigation of a Cultural Help-Seeking Model for Professional Psychological Services With U.S. and Chinese Samples. Journal of Cross-Cultural Psychology 2019 Oct;50(9):1027–1049. doi: 10.1177/0022022119878506

26. Figueroa CA, Aguilera A. The Need for a Mental Health Technology Revolution in the COVID-19 Pandemic. Front Psychiatry 2020 Jun 3;11:523. doi: 10.3389/fpsyt.2020.00523

27. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. Journal of Adolescent Health 2010 Jan;46(1):3–10. doi: 10.1016/j.jadohealth.2009.08.008

28. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health—a systematic review and meta-analysis. npj Digit Med 2019 Dec;2(1):1–10. doi: 10.1038/s41746-019-0188-8

29. Laux G. Online-/Internet-Programme zur Psychotherapie bei Depression - eine Zwischenbilanz.

Journal für Neurologie, Neurochirurgie und Psychiatrie 2017;18(1):16–24.

30. Yang E, Schamber E, Meyer RML, Gold JI. Happier healers: Randomized controlled trial of mobile mindfulness for stress management. The Journal of Alternative and Complementary Medicine 2018 May;24(5):505–513. doi: 10.1089/acm.2015.0301

31. Sun S, Lin D, Goldberg S, Shen Z, Chen P, Qiao S, Brewer J, Loucks E, Operario D. A mindfulness-based mobile health (mHealth) intervention among psychologically distressed university students in quarantine during the COVID-19 pandemic: A randomized controlled trial. Journal of Counseling Psychology 2022 Mar;69(2):157–171. doi: 10.1037/cou0000568

32. Schulte-Frankenfeld PM, Trautwein F. App-based mindfulness meditation reduces perceived stress and improves self-regulation in working university students: A randomised controlled trial. Applied Psych Health & Well 2022 Nov;14(4):1151–1171. doi: 10.1111/aphw.12328

33. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. Journal of Consulting and Clinical Psychology 2020 Jan;88(1):1–13. doi: 10.1037/ccp0000459

34. Dhinagaran DA, Martinengo L, Ho M-HR, Joty S, Kowatsch T, Atun R, Tudor Car L. Designing, Developing, Evaluating, and Implementing a Smartphone-Delivered, Rule-Based Conversational Agent (DISCOVER): Development of a Conceptual Framework. JMIR Mhealth Uhealth 2022 Oct 4;10(10):e38740. doi: 10.2196/38740

35. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y-L, Atun R. Conversational agents in health care: Scoping review and conceptual analysis. J Med Internet Res 2020 Aug 7;22(8):e17158. doi: 10.2196/17158

36. He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, Hou X. Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. Journal of Medical Internet Research 2023 Apr 28;25(1):e43862. doi: 10.2196/43862

37. Bird T, Mansell W, Wright J, Gaffney H, Tai S. Manage your life online: A web-based randomized controlled trial evaluating the effectiveness of a problem-solving intervention in a student sample. Behav Cogn Psychother 2018 Sep;46(5):570–582. doi: 10.1017/S1352465817000820

38. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456–464. doi: 10.1177/0706743719828977

39. Kramer J-N, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, Fleisch E, Kowatsch T. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Annals of Behavioral Medicine 2020 Jun 12;54(7):518–528. doi: 10.1093/abm/kaaa002

40. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): Pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 3;8(4):e15806. doi: 10.2196/15806

41. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, Robinson A. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. J Med Internet Res 2021 Mar 23;23(3):e24850. doi: 10.2196/24850

42. Haug S, Paz Castro R, Scholz U, Kowatsch T, Schaub MP, Radtke T. Assessment of the efficacy of a mobile phone–delivered just-in-time planning intervention to reduce alcohol use in adolescents: Randomized controlled crossover trial. JMIR Mhealth Uhealth 2020 May 26;8(5):e16937. doi: 10.2196/16937

43. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):1–17. doi: 10.2196/16021

44. Ma T, Sharifi H, Chattopadhyay D. Virtual Humans in Health-Related Interventions: A Meta-Analysis. Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems Glasgow Scotland Uk: ACM; 2019. p. 1–6. doi: 10.1145/3290607.3312853

45. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med Nature Publishing Group; 2023 Jun 30;1–3. doi: 10.1038/s41591-023-02412-6

46. Hastings J. Preventing harm from non-conscious bias in medical generative AI. The Lancet Digital Health Elsevier; 2024 Jan 1;6(1):e2–e3. PMID:38123253

47. Goldberg CB, Adams L, Blumenthal D, Brennan PF, Brown N, Butte AJ, Cheatham M, deBronkart D, Dixon J, Drazen J, Evans BJ, Hoffman SM, Holmes C, Lee P, Manrai AK, Omenn GS, Perlin JB, Ramoni R, Sapiro G, Sarkar R, Sood H, Vayena E, Kohane IS. To do no harm — and the most good — with AI in health care. Nat Med Nature Publishing Group; 2024 Mar;30(3):623–627. doi: 10.1038/s41591-024-02853-7

48. Ollier J, Neff S, Dworschak C, Sejdiji A, Santhanam P, Keller R, Xiao G, Asisof A, Rüegger D, Bérubé C, Hilfiker Tomas L, Neff J, Yao J, Alattas A, Varela-Mato V, Pitkethly A, Vara MD, Herrero R, Baños RM, Parada C, Agatheswaran RS, Villalobos V, Keller OC, Chan WS, Mishra V, Jacobson N, Stanger C, He X, von Wyl V, Weidt S, Haug S, Schaub M, Kleim B, Barth J, Witt C, Scholz U, Fleisch E, Wangenheim F von, Car LT, Müller-Riemenschneider F, Hauser-Ulrich S, Asomoza AN, Salamanca-Sanabria A, Mair JL, Kowatsch T. Elena+ Care for COVID-19, a pandemic lifestyle care intervention: Intervention design and study protocol. Front Public Health 2021 Oct 21;9:1–17. doi: 10.3389/fpubh.2021.625640

49. Stieger M, Flückiger C, Rüegger D, Kowatsch T, Roberts BW, Allemand M. Changing personality traits with the help of a digital personality change intervention. Proc Natl Acad Sci U S A 2021 Feb 23;118(8):e2017548118. PMID:33558417

50. Castro O, Mair JL, Salamanca-Sanabria A, Alattas A, Keller R, Zheng S, Jabir A, Lin X, Frese BF, Lim CS, Santhanam P, van Dam RM, Car J, Lee J, Tai ES, Fleisch E, von Wangenheim F, Tudor Car L, Müller-Riemenschneider F, Kowatsch T. Development of “LvL UP 1.0”: a smartphone-based, conversational agent-delivered holistic lifestyle intervention for the prevention of non-communicable diseases and common mental disorders. Front Digit Health 2023;5:1039171. PMID:37234382

51. Ulrich S, Gantenbein AR, Zuber V, Von Wyl A, Kowatsch T, Künzli H. Development and Evaluation of a Smartphone-Based Chatbot Coach to Facilitate a Balanced Lifestyle in Individuals With Headaches (BalanceUP App): Randomized Controlled Trial. J Med Internet Res 2024 Jan 24;26:e50132. doi: 10.2196/50132

52. Maciejewski J, Smoktunowicz E. Low-effort internet intervention to reduce students’ stress delivered with Meta’s Messenger chatbot (Stressbot): A randomized controlled trial. Internet Interventions 2023 Sep 1;33:100653. doi: 10.1016/j.invent.2023.100653

53. Gabrielli S, Rizzi S, Bassi G, Carbone S, Maimone R, Marchesoni M, Forti S. Engagement and Effectiveness of a Healthy-Coping Intervention via Chatbot for University Students During the COVID-19 Pandemic: Mixed Methods Proof-of-Concept Study. JMIR Mhealth Uhealth 2021 May 28;9(5):e27965. doi: 10.2196/27965

54. Liu H, Peng H, Song X, Xu C, Meng Z. Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. Internet Interventions 2022

Jan 6;27:100495. doi: 10.1016/j.invent.2022.100495

55. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: A review and meta-analysis. Journal of Affective Disorders 2013 May;148(1):1–11. doi: 10.1016/j.jad.2012.11.026

56. MobileCoach. Get Started | MobileCoach. Available from: https://www.mobile-coach.eu/ [accessed Nov 27, 2021]

57. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. Journal of Consulting and Clinical Psychology 2017 Feb;85(2):147–159. doi: 10.1037/ccp0000138

58. Haug S, Paz Castro R, Kowatsch T, Filler A, Schaub MP. Efficacy of a technology-based, integrated smoking cessation and alcohol intervention for smoking cessation in adolescents: Results of a cluster-randomised controlled trial. Journal of Substance Abuse Treatment 2017 Nov;82:55–66. doi: 10.1016/j.jsat.2017.09.008

59. Stieger M, Eck M, Rüegger D, Kowatsch T, Flückiger C, Allemand M. Who wants to become more conscientious, more extraverted, or less neurotic with the help of a digital intervention? Journal of Research in Personality 2020 Aug;87:1–11. doi: 10.1016/j.jrp.2020.103983

60. Firebase. Firebase. Available from: https://firebase.google.com/?hl=de [accessed Nov 27, 2021]

61. Apple. Testflight. Available from: https://testflight.apple.com/join/oAvL0b5N [accessed Nov 27, 2021]

62. Seidl M-H, Limberger MF, Ebner-Priemer UW. Entwicklung und Evaluierung eines Stressbewältigungsprogramms für Studierende im Hochschulsetting. Zeitschrift für Gesundheitspsychologie 2016 Jan;24(1):29–40. doi: 10.1026/0943-8149/a000154

63. Schwarzer R. Psychologie des Gesundheitsverhaltens: Einführung in die Gesundheitspsychologie. 3. Auflage. Göttingen: Hogrefe; 2004.

64. Shah LBI, Klainin-Yobas P, Torres S, Kannusamy P. Efficacy of psychoeducation and relaxation interventions on stress-related variables in people with mental disorders: A literature review. Arch Psychiatr Nurs 2014 Apr;28(2):94–101. PMID:24673782

65. Wenzel A. Basic Strategies of Cognitive Behavioral Therapy. Psychiatric Clinics of North America 2017 Dec;40(4):597–609. doi: 10.1016/j.psc.2017.07.001

66. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, Merchant GC, Naughton F, Blandford A. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. American Journal of Preventive Medicine 2016 Nov 1;51(5):833–842. doi: 10.1016/j.amepre.2016.06.015

67. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, Van Daele T, Nordgreen T, Höfer S, Oeverland S, Jensen KL. Best Practices and Recommendations for Digital Interventions to Improve Engagement and Adherence in Chronic Illness Sufferers. European Psychologist 2019 Jan 1;24(1):49–67. doi: 10.1027/1016-9040/a000349

68. Klein EM, Brähler E, Dreier M, Reinecke L, Müller KW, Schmutzer G, Wölfling K, Beutel ME. The German version of the Perceived Stress Scale – psychometric characteristics in a representative German community sample. BMC Psychiatry 2016 Dec;16(159):1–10. doi: 10.1186/s12888-016-0875-9

69. Spitzer RL, Williams JBW, Kroenke K. Patient Health Questionnaire--SADS. doi: 10.1037/t06164-000

70. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J GEN INTERN MED 2001 Sep 1;16(9):606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

71. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, Herzberg PY. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266–274. PMID:18388841

72. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: Validity of a New Measure for Evaluating the Severity of Somatic Symptoms. Psychosomatic Medicine 2002 Apr;64(2):258–266. PMID:11914441

73. Instruction Manual. Instructions for Patient Health Questionnaire (PHQ) and GAD-7 Measures. Available from: https://www.pcpcc.org/sites/default/files/resources/instructions.pdf [accessed Feb 7, 2023]

74. Schwarzer R. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. Applied Psychology 2008 Jan;57(1):1–29. doi: 10.1111/j.1464-0597.2007.00325.x

75. Brenninkmeijer V, Lagerveld SE, Blonk RWB, Schaufeli WB, Wijngaards-de Meij LDNV. Predicting the effectiveness of work-focused CBT for common mental disorders: The influence of baseline self-efficacy, depression and anxiety. J Occup Rehabil 2019 Mar;29(1):31–41. PMID:29450678

76. Topkaya N, Gençoğlu C, Şahin E. General self-efficacy and forgiveness of self, others, and situations as predictors of depression, anxiety, and stress in university students. EDUC SCI-THEOR PRACT 2018;18(3):605–626. doi: 10.12738/estp.2018.3.0128

77. Büttner TR, Dlugosch GE. Stress im Studium: Die Rolle der Selbstwirksamkeitserwartung und der Achtsamkeit im Stresserleben von Studierenden. Präv Gesundheitsf 2013 May;8(2):106–111. doi: 10.1007/s11553-012-0369-7

78. Schwarzer R, editor. Self-efficacy: thought control of action. Washington: Hemisphere Pub. Corp; 1992. ISBN:978-1-56032-269-6

79. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): Psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2009; doi: 10.1002/cpp.658

80. Kelders SM, Kok RN, Ossebaard HC, Gemert-Pijnen JEV. Persuasive System Design Does Matter: A Systematic Review of Adherence to Web-Based Interventions. Journal of Medical Internet Research 2012;14(6):e152. doi: 10.2196/jmir.2104

81. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 06;19(12):e402. PMID:29212630

82. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, Michie S. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb 1;40(1):85–98. doi: 10.1007/s10865-016-9818-7

83. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. PMID:15829473

84. Christensen H, Mackinnon A. The Law of Attrition Revisited. Journal of Medical Internet Research 2006 Sep 29;8(3):e558. doi: 10.2196/jmir.8.3.e20

85. Ryan RM, Deci EL. Self-determination theory: basic psychological needs in motivation, development, and wellness. 2017. ISBN:978-1-4625-3896-6

86. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of

the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. doi: 10.2196/mhealth.5849

87. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325–336. doi: 10.1002/wps.20673

88. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 2007 May;39(2):175–191. doi: 10.3758/BF03193146

89. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. Journal of Affective Disorders 2020 Feb;263:413–419. doi: 10.1016/j.jad.2019.11.167

90. Zeger SL, Liang K-Y, Albert PS. Models for Longitudinal Data: A Generalized Estimating Equation Approach. Biometrics 1988 Dec;44(4):1049. doi: 10.2307/2531734

91. Mayring P. Qualitative Content Analysis. Forum: Qualitative Social Research 2000;1(2). doi: 10.17169/fqs-1.2.1089

92. Mayring P. Qualitative Inhaltsanalyse. In: Flick U, v.Kardoff E, Keupp H, Rosenstiel L, Wolff S, editors. Handbuch qualitative Forschung: Grundlagen, Konzepte, Methoden und Anwendungen München: Beltz; 1991. p. 209–213.

93. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77–101. doi: 10.1191/1478088706qp063oa

94. Goldberg SB, Bolt DM, Davidson RJ. Data Missing Not at Random in Mobile Health Research: Assessment of the Problem and a Case for Sensitivity Analyses. Journal of Medical Internet Research 2021 Jun 15;23(6):e26749. doi: 10.2196/26749

95. Ly KH, Ly A-M, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interventions 2017 Dec;10:39–46. doi: 10.1016/j.invent.2017.10.002

96. Daley K, Hungerbuehler I, Cavanagh K, Claro HG, Swinton PA, Kapps M. Preliminary Evaluation of the Engagement and Effectiveness of a Mental Health Chatbot. Front Digit Health 2020 Nov 30;2:576361. doi: 10.3389/fdgth.2020.576361

97. Clutterbuck D, Spence G. Working with goals in coaching. The Sage Handbook of Coaching Sage Publications, Inc.; 2016. p. 218–237.

98. Yang E, Schamber E, Meyer RML, Gold JI. Happier Healers: Randomized Controlled Trial of Mobile Mindfulness for Stress Management. J Altern Complement Med 2018 May;24(5):505–513. PMID:29420050

99. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. Journal of Medical Internet Research 2019 Jul 22;21(7):e12869. doi: 10.2196/12869

100. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 1;22(4):289–298. PMID:20628581

101. Heim E, Rötger A, Lorenz N, Maercker A. Working alliance with an avatar: How far can we go with internet interventions? Internet Interventions 2018 Mar 1;11:41–46. doi:

10.1016/j.invent.2018.01.005

102. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Apr 3;8(4):e15806. PMID:32242820

103. Statista Research Department. Available from: https://de.statista.com/statistik/daten/studie/306922/umfrage/verteilung-der-studierenden-an-fachhochschulen-in-der-schweiz-nach-geschlecht/

104. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254–267. PMID:27966189

105. Carlbring P, Hadjistavropoulos H, Kleiboer A, Andersson G. A new era in Internet interventions: The advent of Chat-GPT and AI-assisted therapist guidance. Internet Interventions 2023 Apr 1;32:100621. doi: 10.1016/j.invent.2023.100621

106. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell Nature Publishing Group; 2023 Jan;5(1):46–57. doi: 10.1038/s42256-022-00593-2

107. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56–67. doi: 10.1016/j.eswa.2019.03.054

108. Suta P, Lan X, Wu B, Mongkolnam P, Chan JH. An Overview of Machine Learning in Chatbots. IJMERR 2020;502–510. doi: 10.18178/ijmerr.9.4.502-510

109. Schachner T, Keller R, V Wangenheim F. Artificial Intelligence-Based Conversational Agents for Chronic Conditions: Systematic Literature Review. J Med Internet Res 2020 Sep 14;22(9):e20701. PMID:32924957

110. Lim SM, Shiau CWC, Cheng LJ, Lau Y. Chatbot-Delivered Psychotherapy for Adults With Depressive and Anxiety Symptoms: A Systematic Review and Meta-Regression. Behavior Therapy 2022 Mar;53(2):334–347. doi: 10.1016/j.beth.2021.09.007

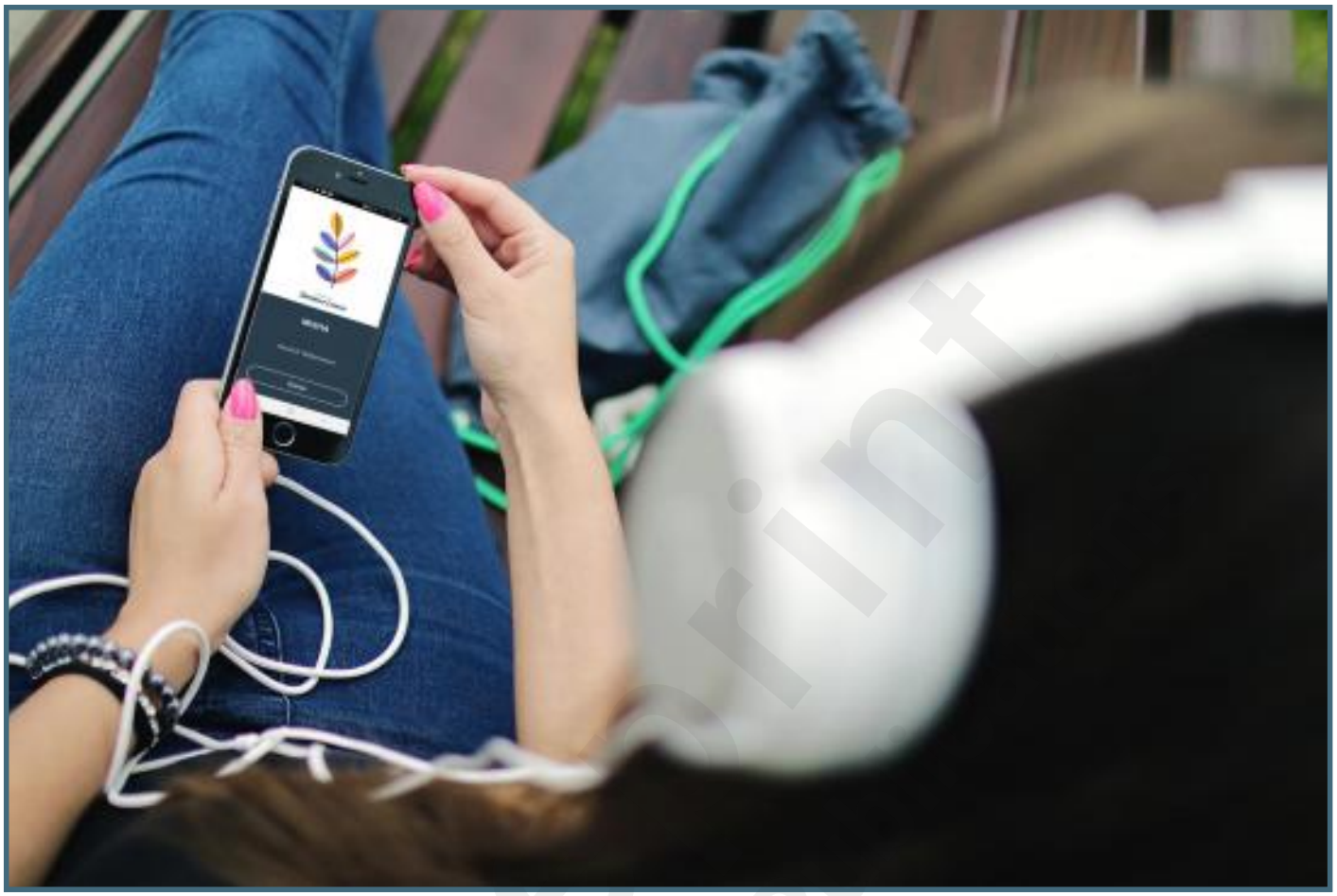
111. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial Intelligence–Based Chatbots for Promoting Health Behavioral Changes: Systematic Review. J Med Internet Res 2023 Feb 24;25:e40789. PMID:36826990

112. Stade EC, Stirman SW, Ungar LH, Schwartz HA, Yaden DB, Sedoc J, DeRubeis R, Willer R, Eichstaedt JC. Artificial intelligence will change the future of psychotherapy: A proposal for responsible, psychologist-led development. PsyArXiv; 2023 Apr. doi: 10.31234/osf.io/cuzvr

113. Torous J, Firth J. The digital placebo effect: Mobile mental health meets clinical psychiatry. The Lancet Psychiatry Netherlands: Elsevier Science; 2016;3(2):100–102. doi: 10.1016/S2215-0366(15)00565-9

Supplementary Files

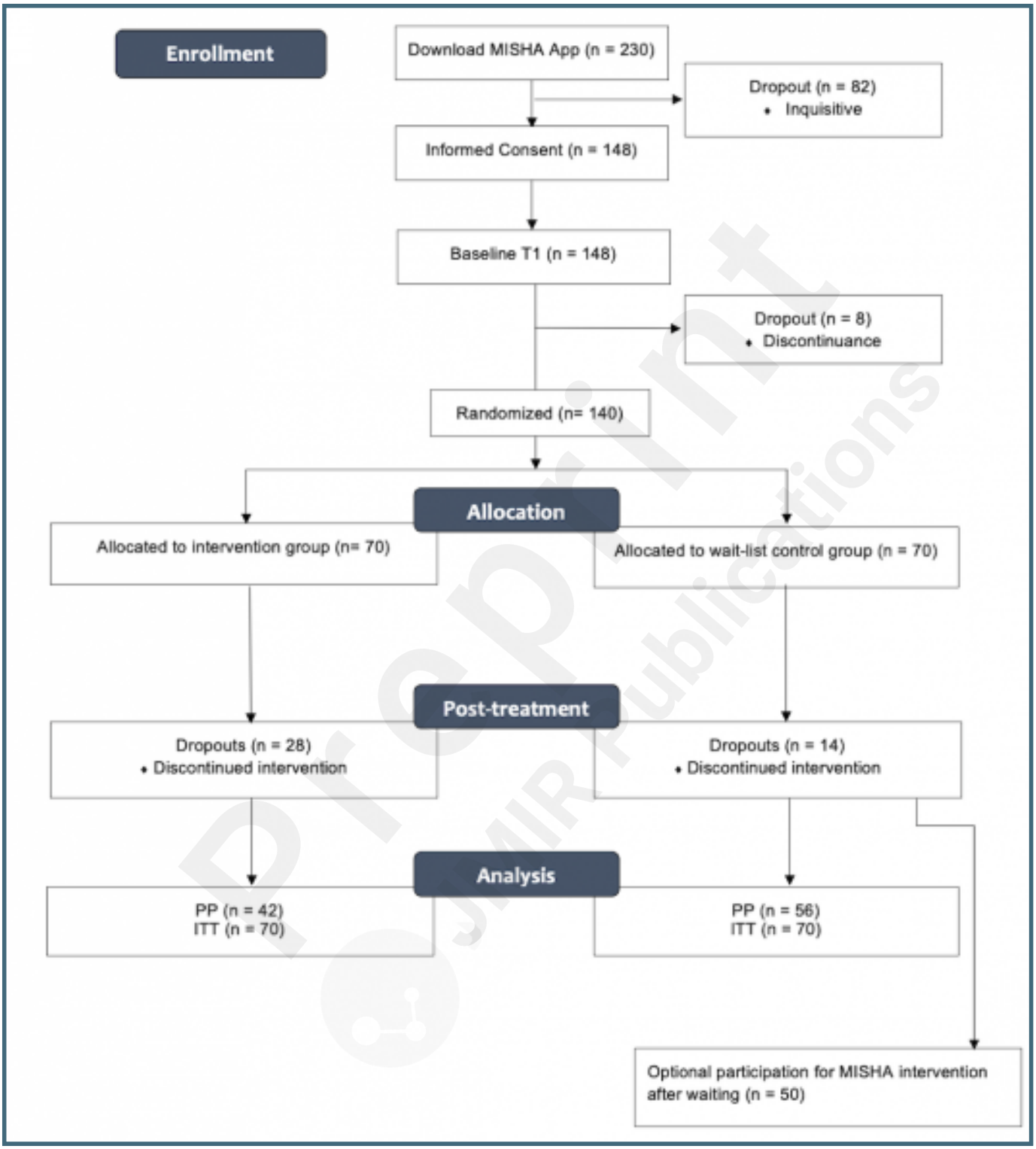
Untitled.



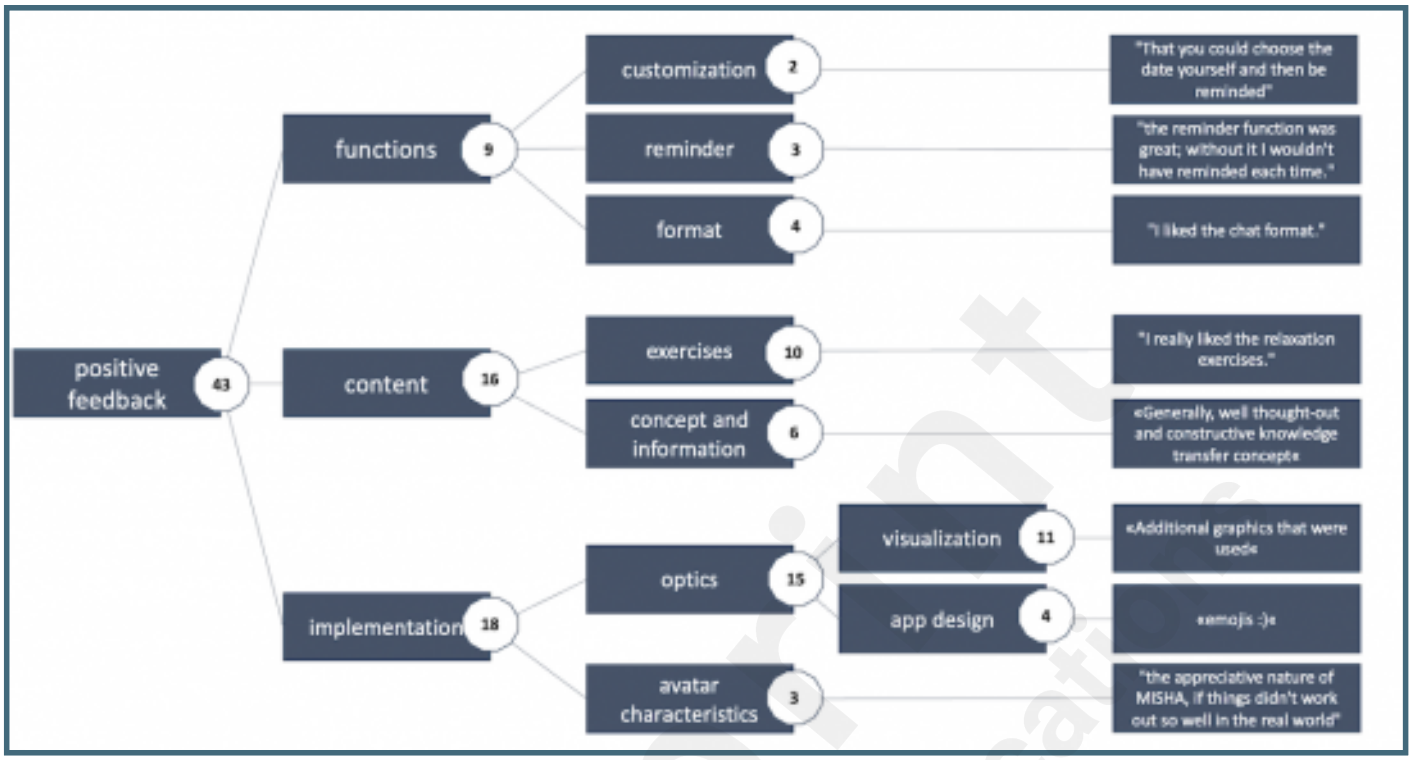
Figures

Screenshots of the MISHA app.

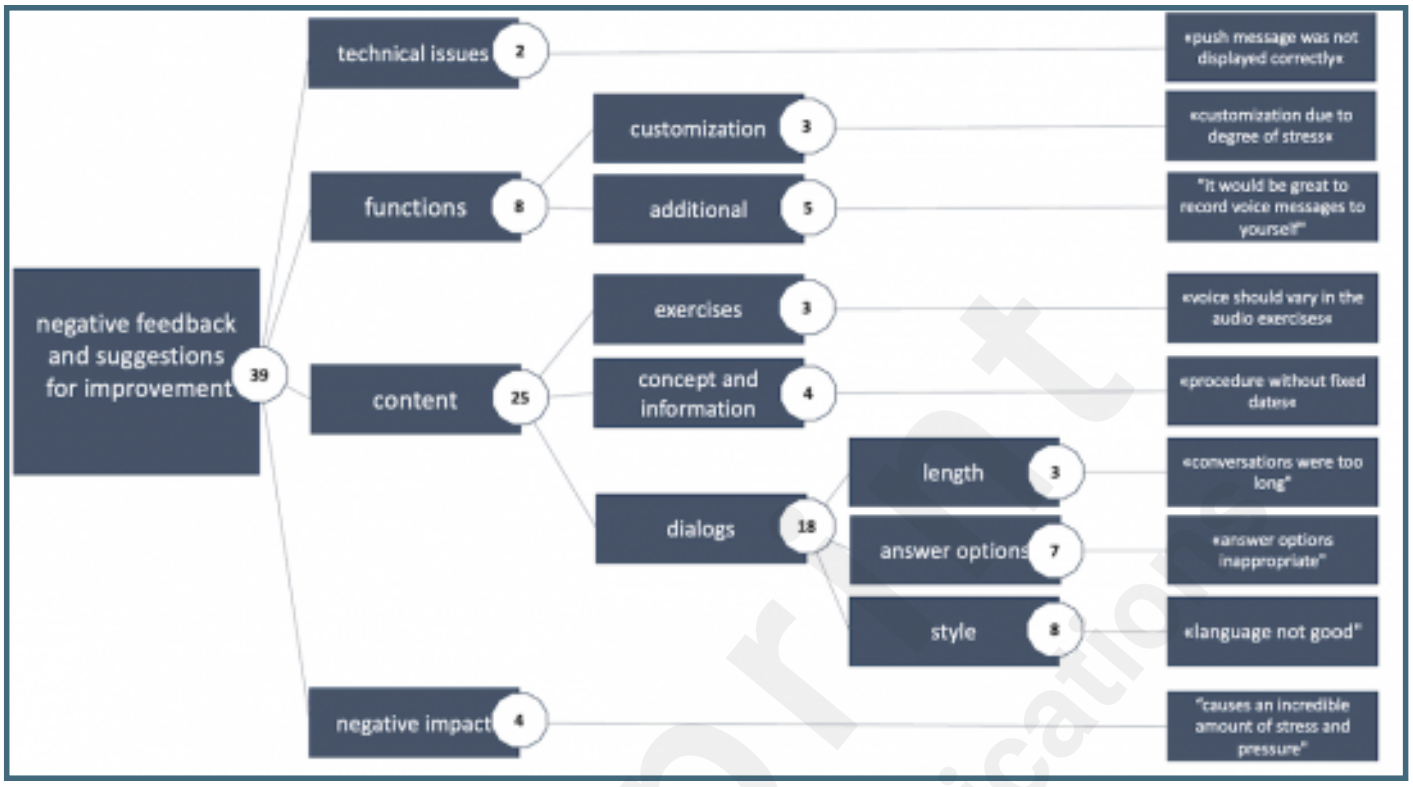
Intervention Flow.



Positive feedback.



Negative feedback.



Multimedia Appendixes

Overview coaching content.
URL: http://asset.jmir.pub/assets/1438af38f7e34b0677bbf8a8ce89224f.pdf

Outcome and timepoints.
URL: http://asset.jmir.pub/assets/e78631e0bc35142fbba19e1b826017e5.pdf

Reminder escalation.
URL: http://asset.jmir.pub/assets/2f7ae23fea85593e54d9148639605ede.pdf

CONSORT (or other) checklists

CONSORT-eHEALT checklist.
URL: http://asset.jmir.pub/assets/13ded50b3167a4fb707ed3235563aa95.pdf

TOC/Feature image for homepages

Misha app coaching.