A chat-bot based version of the WHO-validated intervention Self-Help+ for stress management: Pilot Study

Valentina Fietta, Silvia Rizzi, Chiara De Luca, Lorenzo Gios, Maria Chiara Pavesi, Silvia Gabrielli, Merylin Monaro, Stefano Forti

Submitted to: JMIR Human Factors
on: July 22, 2024

Table of Contents

A chat-bot based version of the WHO-validated intervention Self-Help+ for stress management: Pilot Study

Valentina Fietta[1, 2]; Silvia Rizzi[3]; Chiara De Luca[4]; Lorenzo Gios[3]; Maria Chiara Pavesi[4]; Silvia Gabrielli[3]; Merylin Monaro[2]; Stefano Forti[3]

[1]Fondazione Bruno Kessler Trento JM
[2]University of Padova Department of General Psychology Padova IT
[3]Fondazione Bruno Kessler Trento IT
[4]Istituto Pavoniano Artigianelli Trento IT

Corresponding Author:
Valentina Fietta
Fondazione Bruno Kessler
Via Sommarive, 18, 38123 Povo TN
Trento
JM

Abstract

Background: Advancements in technology offer new opportunities to support vulnerable populations such as pregnant women and women diagnosed with breast cancer during physiologically and psychologically stressful periods.

Objective: This study aims to adapt and co-design the World Health Organization's Self Help Plus into a m-health intervention for these target groups.

Methods: Based on the ORBIT and CeHRes models, low-fidelity and high-fidelity prototypes were developed. Prototypes were evaluated by 13 domain experts from diverse sectors and 15 participants from the target groups to assess usability, attractiveness, and functionality through semantic differential scales, the uMARS questionnaire, and semi-structured interviews.

Results: Feedback from participants indicated positive perceptions of the m-health intervention, highlighting its ease of use, appropriate language, and attractive multimedia content. Areas identified for improvement include enhancing user engagement through reminders, monitoring features, and increased personalization. The quality of the content and adherence to initial protocols were positively evaluated.

Conclusions: This research provides valuable insights for future studies, aiming to enhance usability, efficacy, and effectiveness of the app suggesting a potential role of chat-bot delivered Self Help Plus intervention as a supportive tool for pregnant women and women with breast cancer diagnosis.

(JMIR Preprints 22/07/2024:64614)
DOI: https://doi.org/10.2196/preprints.64614

Original Manuscript

A chat-bot based version of the WHO-validated intervention Self-Help+ for stress management: Pilot Study

Abstract

Background: Advancements in technology offer new opportunities to support vulnerable populations such as pregnant women and women diagnosed with breast cancer during physiologically and psychologically stressful periods.

Objective: This study aims to adapt and co-design the World Health Organization's Self Help Plus into a m-health intervention for these target groups.

Methods: Based on the ORBIT and CeHRes models, low-fidelity and high-fidelity prototypes were developed. Prototypes were evaluated by 13 domain experts from diverse sectors and 15 participants from the target groups to assess usability, attractiveness, and functionality through semantic differential scales, the uMARS questionnaire, and semi-structured interviews.

Results: Feedback from participants indicated positive perceptions of the m-health intervention, highlighting its ease of use, appropriate language, and attractive multimedia content. Areas identified for improvement include enhancing user engagement through reminders, monitoring features, and increased personalization. The quality of the content and adherence to initial protocols were positively evaluated.

Conclusions: This research provides valuable insights for future studies, aiming to enhance usability, efficacy, and effectiveness of the app suggesting a potential role of chat-bot delivered Self Help Plus intervention as a supportive tool for pregnant women and women with breast cancer diagnosis.

Keywords: ACT; well-being; pregnancy; breast cancer; eHealth; mHealth; development; usability; user-centred design.

Introduction

Background

The growing awareness of the profound significance of mental health for individuals and society has spurred an expanding body of research to scrutinise global population trends and the strategies employed to address this issue. Empirical evidence consistently reveals an enduring surge in requests for psychological support, yet this burgeoning demand remains largely unmet due to scarce available resources and services [1]. Consequently, people's needs remain unmet. The factors contributing to the challenge of accessing mental health services are multifaceted. In general, these impediments encompass issues such as suboptimal service quality, inadequate levels of mental health literacy, pervasive stigma, and formidable cost barriers [2]. Within this context, developing and implementing strategies to fortify and enhance the healthcare system becomes increasingly imperative, rendering it more accessible to the population. Significantly, particular attention is being devoted to the prospective role of digital technologies, which can enhance the sustainability of the healthcare system by providing 24/7 support to patients and optimizing healthcare provider interventions [3]. These innovations aim to surmount the aforementioned impediments, advancing digital health as an integral and foundational strategy to foster equitable, affordable, and universally accessible mental health care [4]. A flourishing body of literature corroborates the potential of emergent technologies, encompassing telemedicine, mobile health (m-health) initiatives, and digital therapies (DTx), in

facilitating a seamless continuum of care, extending from clinical settings to patients' homes while embracing a staged care approach [3].

Furthermore, digital health may be particularly suitable for low-intensity mental health interventions [5]. This terminology refers to specific programs wherein the active engagement of healthcare professionals and specialists is not necessarily required. These interventions are grounded in empirically validated [6], evidence-based psychological practices seamlessly integrated into a self-help paradigm, whether guided or unguided. This intervention genre is conventionally designed to be transdiagnostic, offers facile adaptability across diverse contexts, and is readily implementable by non-professional operators. Given their structural attributes and overarching mission, low-intensity interventions represent a valuable conduit for augmenting access to pragmatic, evidence-based psychological interventions, catering to a broad spectrum of recipients. These encompass individuals from the general population to those presenting with limited or mild symptomatic manifestations associated with distress and mental illness [6]. In summary, low-intensity interventions are positioned as a pivotal resource for addressing situations characterized by mild distress, where failure to intervene effectively could potentially precipitate the escalation of these conditions into pathological states (7).

Psychological distress frequently co-occurs with physical illnesses, such as breast cancer, encompassing stress, anxiety, and depression [8]. Cultivating a more optimistic outlook has been demonstrated to play a role in disease management and recovery, underscoring the significance of holistically addressing physical conditions and mental health issues [9,10].

Even a completely different health condition from the above, such as pregnancy, can expose women to similar psychological symptoms. Pregnancy is characterized by major transformations that significantly impact the woman physically, mentally, and socially. How the woman adapts to these changes determines the quality of her life and her levels of well-being [11]. Where adaptation is not functional, symptoms of psychological distress may occur; the most common conditions are anxiety, stress, and depression [11-14]. To date, psychoeducational interventions that promote women's psychological well-being during pregnancy are scarce and tend to focus mainly on samples of women with psychiatric symptomatology (e.g., perinatal depression disorder) [15]. For this reason, our study is part of the digital health framework to support health prevention strategies in the first 1000 days of life [16].

Evidence of the effectiveness of psychological interventions targeting pregnant women or women with breast cancer is increasing [17,18], but access to care services still presents several challenges. Many women face geographic barriers, with specialised centres often far from their homes. In addition, a shortage of qualified personnel, such as psychologists, further limits access to specialised care [19]. The stigma associated with mental health problems can also prevent women from seeking psychological support [20]. Therefore, in numerous instances, these targets are excluded from accessing the requisite services. However, low-intensity interventions emerge as pivotal in addressing this lacuna and the strategic combination with m-health methodologies can yield an effective, sustainable, and inclusive framework for augmenting the scalability of mental health interventions.

This model stands poised to cater comprehensively, encompassing prevention for individuals at potential risk of mental distress and intervention for patients grappling with mild to moderate mental distress [7].

Self Help Plus (SH+)

Self-Help Plus (SH+) is a low-intensity group intervention for stress management initially developed to target populations that are numerous and/or hard to reach by healthcare professionals, under the principle of improving and facilitating access to healthcare

interventions [21]. The SH+ package has been incorporated into the expanding array of low-intensity psychological interventions endorsed by the World Health Organization (WHO) [21]. By design, SH+ is a transdiagnostic intervention that is applicable, meaningful and safe for people with and without mental disorders. SH+ is based on Acceptance and Commitment Therapy (ACT) [22,23], a form of Cognitive-Behavioural Therapy (CBT) [24,25].
The SH+ intervention package has three main components: a pre-recorded audio course, a facilitator manual, and a self-help booklet for participants. This material has been translated and can be easily accessed online in multiple languages at the WHO website [26]. The audio material imparts key information about stress management and guides participants through individual exercises and small group discussions.The intervention is structured into five sessions focused on acceptance- and mindfulness-based techniques for stress management: i) *grounding* (mindfulness), ii) *unhooking* (cognitive distancing), iii) *acting on your values* (value-based behavioural activation), iv) *being kind* (gratitude), and v) *making room* (acceptance).
Preliminary studies report positive effects of SH+, with potential impact on mental well-being, also considering long-term efficacy in a target population of refugees and asylum seekers exposed to stressful situations [27]. Other studies show a still debatable effect of SH+ when applied to healthcare professionals during the COVID-19 pandemic, whilst there emerges a need for further examining the potential role of confounding effects of non-specific factors [28].

Present Research

The present research fits into the landscape of WHO strategies by adapting the stress management intervention developed by WHO itself, SH+, with two main goals, that is, i) to assess the viability of this intervention when targeting specific sub-groups (women with breast cancer and pregnant women) and ii) to validate the intervention as a chat-bot delivered and preventive action. This SH+ intervention, which has already been validated and tested on some specific vulnerable populations (e.g. asylum seekers), [29] will be fully available to users through digital tools. In particular, it will be delivered by a mobile application and guided by a virtual assistant, ALBA. The present research aims to assess the prototype of the ALBA application to gather feedback and needs from key stakeholders to further refine the application from a qualitative perspective of usability, accessibility, and acceptability of the intervention delivered via chatbot.

Methods

Overview

The stress management intervention program is developed iteratively, following the ORBIT model [30], as illustrated in Figure 1, which depicts the pathway followed to translate a human-guided intervention into a possible DTx. In particular, the design and development process encompassed a multidisciplinary approach and continuous, systematic evaluation throughout, as the Center for eHealth Research and Disease Management (CeHRes) comprehensive roadmap approach recommended to improve the uptake and impact of eHealth technologies [31].

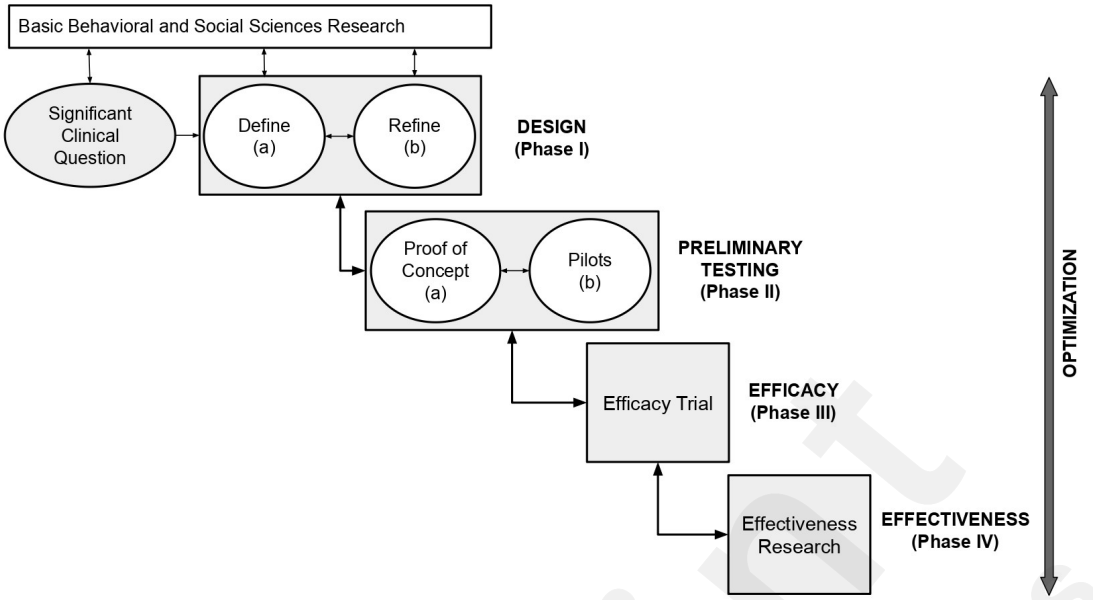

Figure 1. The ORBIT model [30].

The multidisciplinary project team, consisting of psychology, eHealth research, and communication experts, had biweekly meetings during the design and development phase. User-centred design methodologies ensured user involvement throughout the design and development process. Patient representatives, healthcare providers and security experts were consulted throughout. The stress management intervention was developed in iterative processes through a combination of (i) intervention content development, identified and adjusted from the evidence-based cognitive behavioural stress management concept, and (ii) iterative software development (phase 1: low-fidelity prototypes; phase 2: high-fidelity prototypes) and formative evaluation.

Intervention Content Development

A primary goal of this study was to adapt a validated stress management intervention, SH+ [21], into a new technology-based stress management intervention for pregnant women and women with breast cancer. To do so, the development involved a wide range of expertise encompassing psychology, eHealth information technology, interaction design, and specialized knowledge in pregnancy and oncology. A multidisciplinary team was assembled to address diverse user needs, ensure psychological coherence, and integrate technological requirements and user-design principles. The focus was on achieving adaptability of the proposed tools in the digital domain [32], guided by two pivotal methodologies: user-centred design [33,34] and service design [35].

The development process comprised the following stages: i) literature review of WHO protocols, papers on digital mental health and the specific psychological needs of target populations; ii) exploration gathering insights from user representatives (e.g., breast cancer patients and pregnant women), healthcare providers, and eHealth experts, including designers and developers; iii) content adaptation customizing SH+ intervention manual content addressing software development, privacy and security potential issues. The intervention content was adapted and tailored by the entire research team through iterative processes to fit a 5-module-based intervention in electronic format. Adjustments were made to ensure easy language, short sentences, and focus on clear content for small screens.

This holistic approach, merging diverse expertise and user-centric methodologies, underscores the dedication to crafting a robust and efficient chatbot-driven intervention tailored for women dealing with breast cancer and pregnancy.

Phase 1: Iterative Development and Low-Fidelity Prototypes

The adaptation of the SH+ interventions through the implementation of the ALBA chatbot was based on a novel approach to delivering psychological support. Users can engage in a comprehensive and effective intervention through gamification, personalised sessions, reminders, and feedback. Therefore, a key aspect represented by a multi-level structure to consider the different sections of the app and, simultaneously, guarantee proper levels of user engagement, adherence, and overall impact on users' well-being. In the first iteration, in 2023, three psychologists and two communication experts tested and gave feedback on the prototype to ensure that the intervention program was logically built and would meet the stakeholder requirements.

On this basis, the following methodology was applied. The group of experts was divided into pairs, where one person assumed the role of the chatbot while the other adopted the user's perspective, reading their respective segments of the dialogue aloud.

The pairs were reorganised for each of the five distinct modules to ensure a diverse spectrum of interactions and exhaustive coverage of potential dialogue scenarios and avoid biases. This method facilitated the exploration of varied interaction dynamics and the collection of data on multiple communication styles. The oral recitation of dialogues served as a mechanism to evaluate several aspects of the chatbot's effectiveness, such as dialogue realism, rhythmicity/repetition of the texts and communication fluency [36]. The verbalization process aids in gauging how seamlessly the chatbot replicates a human-like conversation, identifying any inconsistencies or unnatural responses. Listening to the dialogue's progression thus allows for the assessment of the conversation's smoothness, encompassing the coherence and pertinence of the chatbot's replies. Another function of this method was to identify any ambiguities or misinterpretations that might emerge during interactions with the chatbot [36]. Simulating the conversation enabled the experts to offer immediate critiques on every facet of the dialogue, contributing to rapid and focused content refinement.

In a nutshell, this procedure aimed to analyze the chatbot's capability to engage in realistic, empathetic, and psychologically suitable conversations, leveraging direct feedback and expert psychologists' insights as key metrics for evaluation.

Phase 2: Iterative Development and High-Fidelity Prototypes

After minor adjustments, the paper prototype was implemented into an electronic format using the LandBot tool [37]. Landbot is a chatbot generator that allows you to create, test and deploy conversational chatbots via WhatsApp and other chat channels.

During the period spanning the end of 2023 and the beginning of 2024, a two-phase evaluation of the ALBA prototype was conducted involving an overall sample of 28 participants, as elaborated in Figure 2. In particular, participants involved were: 13 domain experts (6 psychologists, 21.43% of the total sample), 1 SH+ experts (3.57%), 3 communication experts (10.71%), and 3 usability experts (10.71%)) and 15 users (8 targets: 4 women who are currently pregnant or have given birth within the last year (14.29%) and 4 women who are in current breast cancer disease status or follow-up from it (14.29%); 7 target clinicians: 4 gynaecology clinicians (e.g., obstetricians, gynaecologists; 14.29%) and 3 oncology clinicians (e.g., oncologists, case-managers; 10.71%). The recruitment of participants was based on personal contacts of researchers, selected based on representativeness of the key target user groups addressed by the ALBA solution.

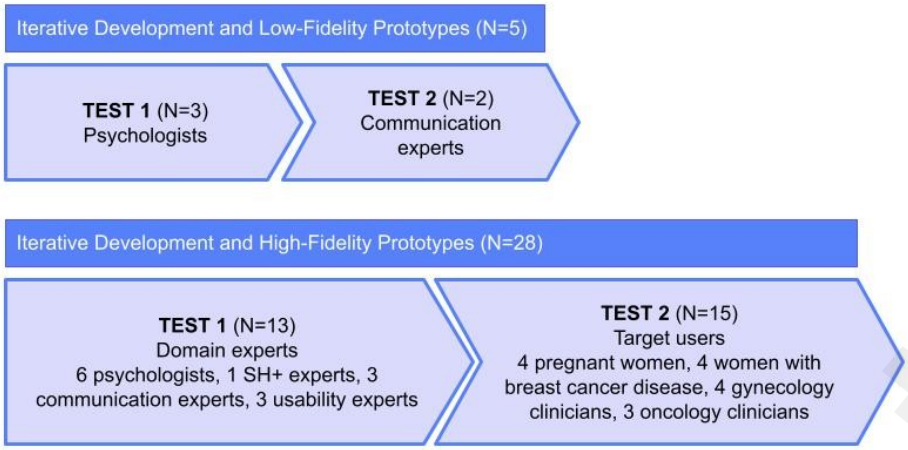

Figure 2. Software development and formative evaluation (*N*=37)

Variables identification

In an attempt to gather the necessary information, key variables were identified for investigation: communication, session structure, materials, engagement, functionality, aesthetics, information, subjective and perceived impact, interaction, communication mode, involvement constancy, general comments, technical implementation's amelioration and content adherence. The first four variables were assessed through the semantic differential tool [38], the following five through the User Version of the Mobile Application Rating Scale (uMARS) [39] and the latter six through an ad hoc semi-structured interview.
The semantic differential is an instrument consisting of a series of scales, each of which is composed of a pair of bipolar adjectives between which a rating scale (5 positions) is placed. Given the study's variables, a list of sub-variables was chosen to be created ad hoc items for the research. Table 1 shows the chosen variables and their respective sub-variables. Based on the target subject domain, each person was asked to evaluate and determine variables, as reported in Table 1 and single items are reported in Appendix Table 1.

Table 1. List of variables investigated through semantic differential (rows) and people involved in evaluating the individual variables (columns).

Variables	Sub-Variables	Psy.	SH+ exp.	Comm. exp.	Us. exp.	Target	Clin.
Communication	Empathy and listening	x	x	x	-	x	x
	Smoothness and fluidity	x	x	x	x	x	x
	Chatbot interaction	x	x	x	x	x	x
	Lexicon	x	x	x	x	x	x
Session structure	Interaction length	x	x	x	x	x	x
Materials	Audio tracks	x	x	x	-	x	x
	Infographics and videos	x	x	x	-	x	x

Note. Psy.: psychologists, SH+ exp.: SH+ expert, Comm. exp.: communication experts, Us. exp.: usability experts, Clin.: clinicians.

The uMARS questionnaire, on the other hand, evaluates mobile applications by covering four objective dimensions (engagement, functionality, aesthetics, and information) and one subjective dimension. Briefly, the questionnaire consists of 20 items covering the inquered variables as follows: engagement (N = 5), functionality (N = 4), aesthetics (N = 3), and information (N = 4), and 4 items belonging to the subjective quality domain. A section on perceived impact (6 items) also assesses users' perceptions of the app's usefulness. Each answer is rated on a 5-point scale (1-inadequate, 2-poor, 3-acceptable, 4-good, 5-excellent) measuring the usability of mobile health apps.
Interviews, instead, are suitable for a more in-depth investigation of users' attitudes and preferences toward new technological solutions since open-ended discussions with users can help researchers to understand better the issues and concerns related to the possible future adoption of these solutions [40]. In the Appendix Table 2 shows the list of topics and questions posed during the interviews.

Context-specific methodologies

Personas

Personas is a user-centred and service design method utilized to create and visualize fictional representations of the target group [41]. Personas is an effective method for all project team members to understand better the target group for which the app is built. Personas in this study contained information about the pregnancy background, challenges, and technology use. See Figure 3 for illustrated examples of study Personas. Psychologists had to take the perspective of one of these Personas before the reading.

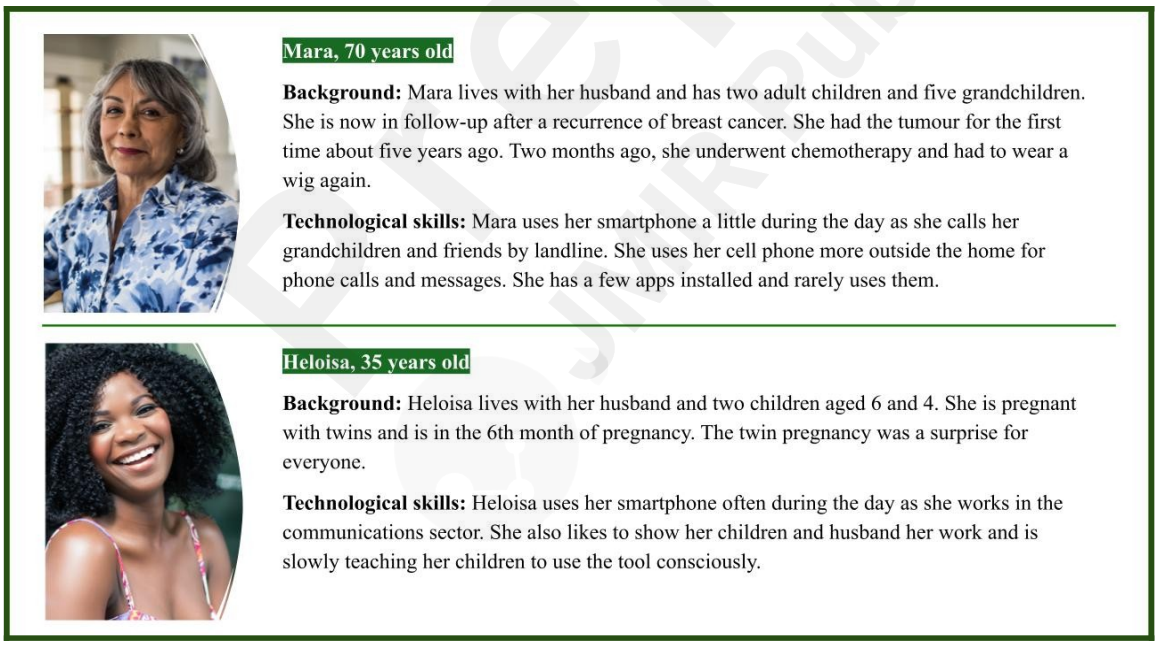

Figure 3. Examples of Personas

Focus on experience and expertise

The study aims to involve experts from various fields and does not require all participants to respond to every specific ad hoc item of the Semantic Differential and questions of semi-structured interview. Usability experts are not asked to evaluate specific intervention's content, whereas SH+ expert is queried about the fidelity of the content to the original intervention

during the interview. Clinicians and psychologists, as well as participants from the pregnancy context, are asked questions relevant to that context, while those from the oncology field respond to questions specific to their area of work or direct life experience.

Procedure

All data were collected confidentially with participants' informed consent. Recruitment was done through word of mouth and direct acquaintance, clearly stating the study's objectives. Only volunteers aged 18 or older were included. Before the study, participants received a privacy notice and consent form via Google Forms, allowing them to consent to participation and data processing.

Operationally, the experimental procedure was structured as follows:

1. Display of the information notice and presentation of informed consent.
2. Participants who decided to participate in the study were asked to fill out a questionnaire which collected some generic biographical information (age, schooling, gender, employment status) and some information related to their knowledge and use of mobile applications.
3. Next, subjects were interacting with the ALBA application prototype. During the study, the subject was asked to test the ALBA app; in particular, he/she had the opportunity to explore the interface and different sections of the app, read, listen to and interact with the chatbot during the dialogue session, and was also informed about the future implementation of reminders, feedback for the activities proposed by the chatbot.
4. Then, questionnaires were presented to the participant regarding their overall experience with the system. In the final stages, participants were asked to evaluate the experience they had with the ALBA app by answering two questionnaires. In particular, the Semantic Differential tool [38] and the Italian version of the User version of the uMARS [42] were used for the evaluation through questionnaires.
5. Finally, a brief semi-structured interview was conducted. After the usability assessment, participants were invited to join an online interview (duration 20 minutes each) to further report about their expectations, preferences, and concerns regarding the ALBA solution tested. A total of 28 interviews were conducted by a researcher and audio-recorded to enable a more detailed analysis of participants' responses. It was then analysed and processed using qualitative tools. The conductor initially provided a brief introduction to the interview objectives. Then, participants were asked to answer a series of semi-structured questions regarding their expectations and preferences to use the app.

All data were collected in Italian and pseudonymized, with participants' informed consent. Confidential audio recordings of semi-structured interviews were used for data analysis, and subjects were identified only by numeric codes. At the study's conclusion, participants can request the research outcomes from the research manager.

The study was approved by the University of Padua Ethics Committee for psychological research on 01/08/2024, with the unique reference number: 238-b.

Data Analysis

Data analysis for quantitative results from the Semantic Differential and uMARS questionnaires was conducted using JASP and R software [43,44]. Due to the small sample size, non-parametric tests were used [45]. When applicable, the Wilcoxon signed-rank test (W) was used as a non-parametric alternative to the one sample t-test [45, 46, 47], and the rank-biserial correlation (r) was reported to indicate the strength of association, along with the

corresponding 95% confidence interval (*CI*) [47]. All analysis results were considered significant with a critical *P*-value set at 0.05.

The data collected during the interviews were analysed using a qualitative method. Initially, all interview transcriptions were reviewed to obtain an overall understanding. Subsequently, thematic analysis was conducted [48], organising the themes into tables based on different contexts and participant types (e.g., psychologists, clinicians). Finally, a comprehensive report was created, highlighting the main findings, with references to the specific groups where applicable.

Results

Phase 1: Iterative Development and Low-Fidelity Prototypes

Specifically, therefore, the app is designed to include five sections: Chatbot, Exercises, Diary, Gallery, and Progress. The first low-fidelity prototype version of the app was developed (Figure 4).

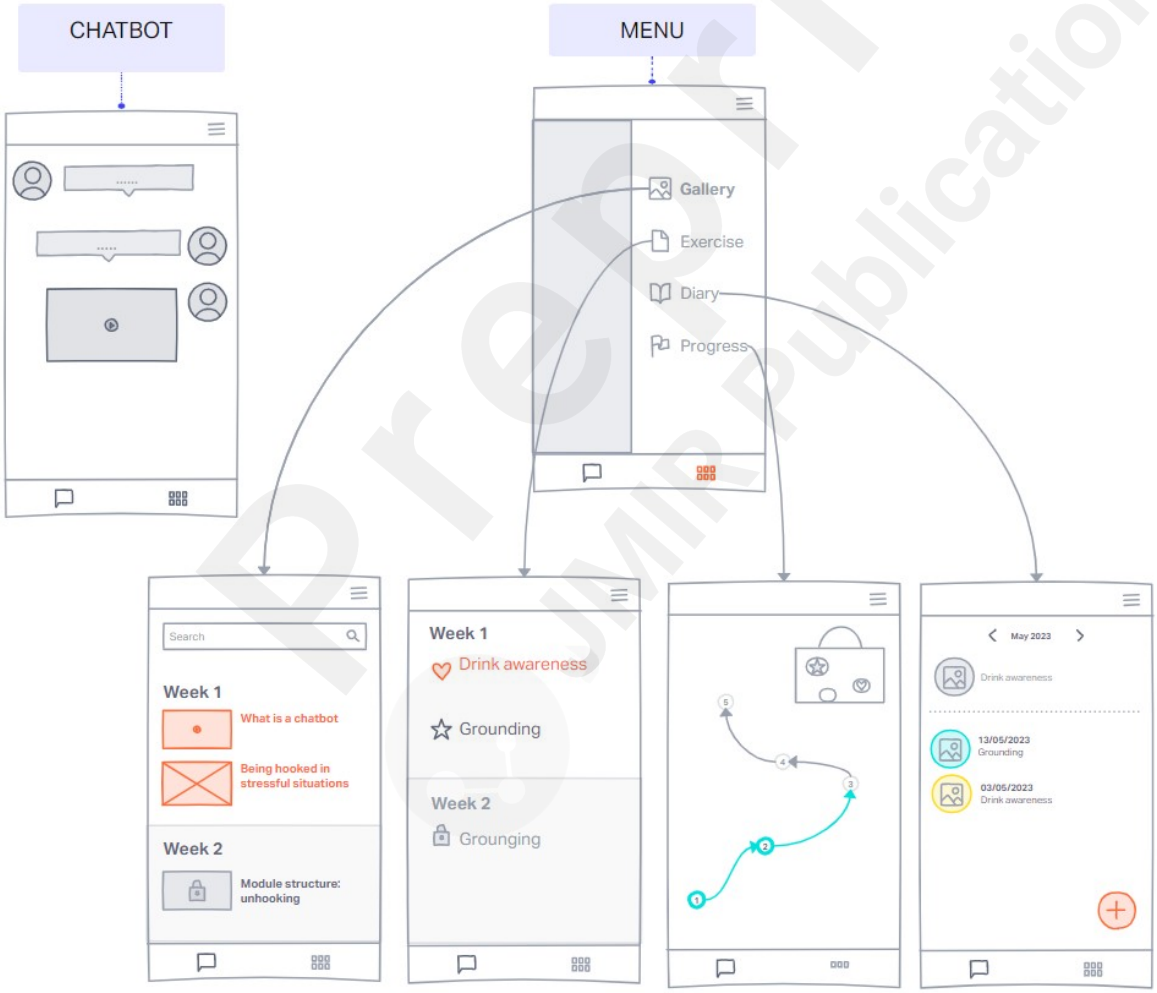

Figure 4. Low-fidelity prototype version of the ALBA app

ALBA assigns homework during the session delivery that the users should carry out during the following days, so the app Diary section will support self-monitoring of exercise progress and completeness of sessions, reinforcing users' empowerment. The completion of exercises is

monitored in the evenings before the next session. All the exercises, divided week per week, are shown in the Exercise section. The Gallery page features all multimedia material delivered by ALBA, like educational videos, images and intervention introductions. Additionally, the app features a section called Progress, which provides users with an overview of their “journey” towards increased well-being. Using the journey metaphor, users can see the sticker badges earned by practising exercises between sessions appearing on their suitcase illustration. These badges serve as long-term positive reinforcement and gratification for their efforts. Users receive badges for completion of exercises, and this is another way they can track their progress. Again, this gamification approach has been selected to further reinforce empowerment and sense of self-efficacy of the user.

Regarding the interaction with the low-fidelity prototyped session in the co-design phase, two main themes emerged.

Firstly, it was possible to review the communication style through role play. In order to make the protocol more realistic, some parts of the dialogue have been revised. The changes were made from a grammatical and syntactic point of view to make the text more fluent and fluid in reading. The changes also took empathy into account. This attention creates a feeling of trust in the relationship with ALBA. Through empathic communication, the person can feel in a safe space, within which she is reassured and protected. For example, after these changes, ALBA allows the option to skip a particularly sensitive answer, it also lets the user find a calm place before doing audio exercises, and moreover, it gives the possibility to review the concepts of the previous session or not.

A second critical result emerging from the specialists who tested the low-fidelity prototype is allowing the user to divide the session into two mini-sessions. Given that the time needed to complete a single session is approximately 40 minutes, it is generally considered excessive to dedicate time to interaction with an application. For this reason, appropriate changes were made to the dialogue to provide a partial and optional closure after 20 minutes and following a gradual resumption of the session by the subsequent day. So, to support adherence, the session can be divided into two mini-sessions, preventing user fatigue and improving the usability of both the chatbot and the app itself.

At the same time, writing errors were corrected, and the remaining parts of the original group-made protocols were adapted to the individual intervention. Mainly, attention was placed on the translation from group to individual gratitude exercises, the final exercise of each session.

Phase 2: Iterative Development and High-Fidelity Prototypes

To better understand the application and the entire intervention, more realistic mockups were created (see Figure 5), after minor adjustments on the basis of the information gathered. This allowed users to better represent the application in its final appearance.

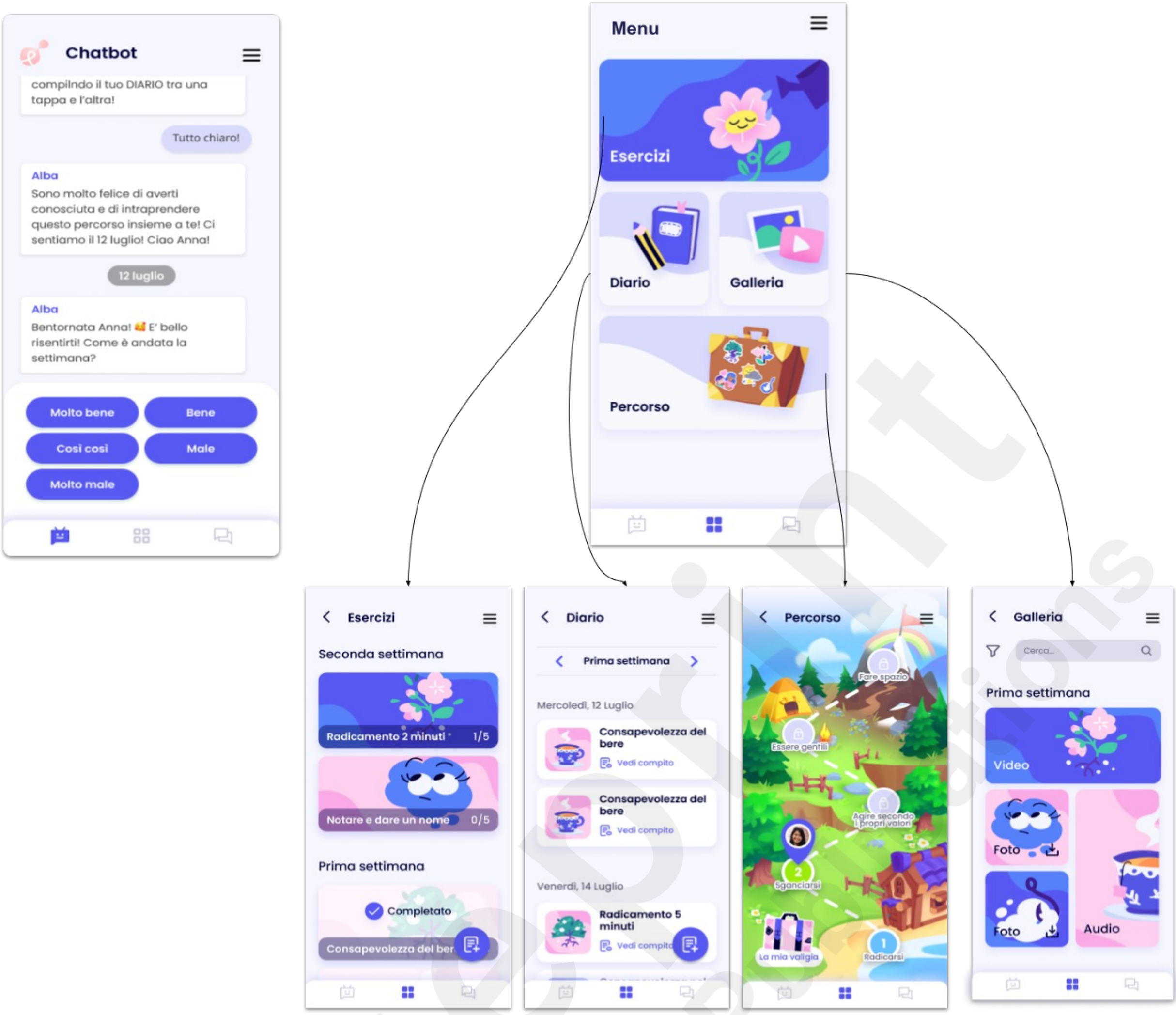

Figure 5. More realistic mockups of the ALBA application with its different sections.

In total, 28 individuals (4 men and 24 women) participated in the pilot evaluation; the average age was 41.39 years old (*SD*= 11.66, ranging from a minimum of 28 to a maximum of 72 years old). The average amount of schooling years of the sample was 18.98 (*SD*= 2.88, range 13-23 years of education).

Quantitative results

Semantic Differential. In this case, also, using a semantic differential-based questionnaire facilitates the observation of the respondents' average positioning with respect to the three macro-variables under investigation. In particular, concerning the sub-variables, several significant results emerged from the Wilcoxon analysis, as reported in Table 2. The graphical representation of mean values derived from the Semantic Differential for each item is reported in Appendix Figure 1. As seen from Figure 1 in the Appendix, a tendency toward the positive semantic pole (right pole) emerges in the feedback in these groups of participants.

Table 2. Results from the Semantic Differential.

	95% CI for Effect Size

	N	*M*	*Mdn*	*SD*	*W*	*P*	*r*	*Lower*	*Upper*
Empathy and listening	25	3.55	3.60	0.67	213.50	<.001	0.85	0.64	0.94
Smoothness and fluidity	28	3.89	4.00	0.69	266.00	<.001	0.93	0.83	0.97
Chatbot interaction	28	3.72	3.75	0.55	271.50	<.001	0.97	0.92	0.99
Lexicon	28	4.41	4.50	0.61	404.00	<.001	0.99	0.98	1.00
Session structure	28	2.95	3.00	0.63	64.50	.87	-0.05	-0.54	0.47
Audio tracks	25	3.97	4.00	0.44	300.00	<.001	1.00	1.00	1.00
Infographics and videos	25	4.08	4.00	0.50	300.00	<.001	1.00	1.00	1.00

Note. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

The results indicate that participants generally responded positively across various sub-variables. Significant positive responses were observed for empathy and listening, smoothness and fluidity, chatbot interaction, lexicon, audio tracks, infographics, and videos. The Session structure sub-variable did not show a significant positive trend, indicating a neutral or mixed response. The effect sizes ranged from moderate to large, suggesting varying degrees of impact for the different sub-variables under investigation.

uMARS. The uMARS evaluated the respondents' average positioning in four key dimensions. Detailed results from the Wilcoxon tests are summarised in Table 3.

Table 3. Results from the uMARS.

								95% CI for Effect Size	
	N	***M***	***Mdn***	***SD***	***W***	***P***	***r***	***Lower***	***Upper***
Engagement	28	3.55	3.60	0.43	300.00	<.001	1.00	1.00	1.00
Functionality	28	4.16	4.25	0.51	378.00	<.001	1.00	1.00	1.00
Aesthetics	28	3.86	4.00	0.49	300.00	<.001	1.00	1.00	1.00
Information	28	4.20	4.25	0.52	405.00	<.001	1.00	0.99	1.00
Subjective Items	28	3.20	3.25	0.50	233.50	.06	0.44	0.02	0.72
Perceived Impact	28	3.66	3.83	0.60	264.00	<.001	0.91	0.79	0.97

Note. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

The results indicate consistently high ratings across all dimensions of the uMARS. Significant positive responses were noted for *Engagement, Functionality, Aesthetics, Information, and Perceived Impact*. Effect sizes were uniformly high, particularly for the Wilcoxon tests, indicating a strong positive skew in user perceptions for each dimension evaluated. However,

items like “Customization” and “Interactivity” did not show significant positive trends in the Engagement scale, with “Customization” even indicating a negative effect size, suggesting variability or mixed responses from users.

Considering singularly the *Subjective Items* (which scale did not show a significant trend), the items “Would you recommend” (*Mdn*= 4.00, *SD*= 0.62, *P*<.001, *Effect Size*= 1.00) and “Overall rating” (*Mdn*= 4.00, *SD*= 0.52, *P*<.001, *Effect Size*= 1.00) received a strong positive response. This suggests that most users would recommend the app to others, with a robust positive effect size indicating widespread satisfaction with the app’s performance, utility, and perceived value. However, for the item “How many times” (*Mdn*= 3.00, *SD*= 0.77, *P*=.64, *Effect Size*= 0.13), responses were mixed regarding the frequency of app use, with no significant trend emerging. This indicates that users are likely to use the app on average 3-10 times in the next year. Moreover, for the item “Would you pay” (*Mdn*= 2.00, *SD*= 0.72, *P*<.001, *Effect Size*= -1.00), there was a significant negative response, indicating that users are generally unwilling to pay for the app. The strong negative effect size underscores a consistent reluctance to incur costs for app use.

Regarding the relevant *Perceived Impact* scale of uMARS, items such as “Awareness” (*Mdn*= 4.00, *SD*= 0.74, *P*<.001, *Effect Size*= 0.24), “Knowledge” (*Mdn*= 4.00, *SD* = 0.79, *P*<.001, *Effect Size*= 0.26), “Attitudes” (*Mdn*= 3.50, *SD*= 0.73, *P*=.02, *Effect Size*= 0.26), “Intention to Change” (*Mdn*= 4.00, *SD*= 0.69, *P*=.001, *Effect Size*= 0.26), and “Help Seeking” (*Mdn* =4.00, *SD*= 1.09, *P*=.003, *Effect Size*= 0.23) were rated positively. This indicates that the app positively impacted users, though the effect sizes are lower compared to other items, suggesting moderate consensus and some variability in responses. Conversely, the “Behavior Change” item result (*Mdn*= 4.00, *SD*= 0.64, *P*<.001, *Effect Size*= 0.88) highlights that the app had a significant positive impact on users, with a high effect size indicating strong agreement among users about the app’s effectiveness in promoting behavioral modifications.

Every item means and statistical significance are reported in the Appendix Table 3. Figure 6 represents the graphical distribution of items’ scores.

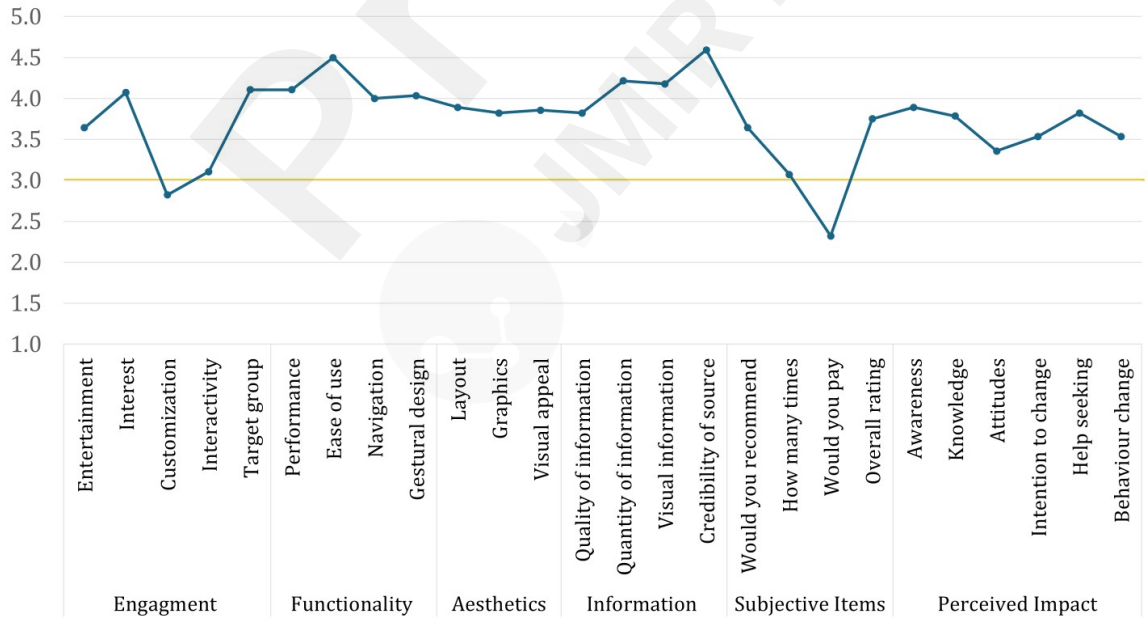

Figure 6. Graphical distribution of uMARS items’ scores.

Qualitative results: Semi-structured interviews.

One psychologist conducted the qualitative interview to gather additional information.
Interviews were conducted and analysed according to the identified thematic variables.
Set 1: Interaction. The interaction with the chatbot was generally well-received. Clinicians appreciated the good overall interaction, pointing out the effectiveness of the interactive videos and images, as well as the presence of multiple responses. However, they found it difficult to go back to previous answers, noting repetitiveness and lack of novelty in the content. The communication experts praised the fluidity of the interaction and particularly appreciated the personality of the app, but criticised the excessively fragmented design of the messages.
The SH+ expert highlighted the confidence given by the chatbot, which calls the user by name and provides relevant answers, but found a lack of clarity as to when answers could be free and expressed a preference for the human figure in the videos. The usability experts appreciated the guidance and psycho-educational support offered by the chatbot, but pointed out the one-way interaction and the excessive number of messages in a row.
Among the features preferred by those who tested the app are videos with exercises, the personality of the app, heterogeneous content, and images with goals and values. Among the elements to be improved, however, are some aspects of the videos. In particular, the psychologists pointed out that professional voices could greatly improve the effectiveness of the videos, which were nevertheless appreciated for their content. The SH+ expert, on the other hand, found the videos and audio a little too slow and monotonous, but greatly appreciated the chatbot's confidence and relevant answers. Finally, the usability experts found the videos and graphics useful, but criticised the excessive length of the messages.
There is some discord regarding the possibility of correcting oneself and going back. In fact, SH+ experts and psychologists consider it very useful and suggest it should be included, while communication experts do not find it particularly useful to go back. The usability experts did not find errors frequent, and the psychologists considered the consequences of errors not serious. On the other side, the target group highlighted that while the interaction was varied, it was tiresome due to the length of the videos and messages, suggesting a reduction.
Set 2: Communication mode. Concerning clarity and thoroughness, all categories of experts appreciated the possibility to better investigate concepts with examples and videos. In particular, the communication experts emphasised the clarity of the content, although they suggested using more concise answers. The same suggestion was made by the SH+ expert, who found some dialogues too complex and articulate. Clinicians disagreed, who instead pointed out simple and accessible language.
Finally, the usability experts also noted that the repetition of information was positive, especially for weekly use, but found the audio and video a little too slow and monotonous. Similarly, psychologists appreciated the repetition of videos to better explain concepts, and suggested alternating text with images or videos to lighten the information load.
The language used by the chatbot was generally well-received. Clinicians appreciated the simple language and short sentences, while communication and usability experts praised the clear, appropriate and friendly language. Psychologists and SH+ experts also found the language appropriate to the content and easy to understand. However, SH+ experts considered the tone to be a little slow. It was suggested to keep the language simple, but to consider more complexity and variety in the content, using a more active and engaging tone.
The use of emojis was well received. All categories appreciated and found useful the visual mode of emoji, which was also considered effective by usability experts. However, it was suggested to offer both emojis and words as a response option to accommodate different preferences. Finally, the target group pointed out that while the length of sentences and appropriate terms were satisfactory, the messages were sometimes too long. They suggested reducing the length of the messages and incorporating both words and emojis to effectively

express emotions.

Set 3: Involvement and constancy. User involvement and consistency were strengths for some groups, while others experienced difficulties. Clinicians found the tone positive and encouraging, and the communication fast. The SH+ experts found the communication engaging, while the psychologists found good engagement in the videos. However, clinicians found it difficult to feel engaged without a physical person, while usability experts found the interaction similar to reading a book, needing more interaction and exchange. For these reasons, it was suggested that the level of personalisation and interaction be improved to increase engagement.

Customisation of the chatbot was considered good, but with room for improvement. The clinicians appreciated the personalisation by name and the space for personal choices, but noted standard answers for all and the lack of specific personalisation elements. The SH+ expert gave a score of 8/10 to personalisation, while the psychologists noted that the answers were standard. It was suggested to offer the possibility to add photos and avatars, and to improve the customisation of answers. The target group found the content engaging and motivating to continue, with good personalization using names. However, they noted that it became less engaging in the long term, with repetition feeling lengthy and tedious, and exercises sometimes monotonous.

Set 4: General questions. Communication experts noted that the timing should be personalised according to individual circumstances, emphasising that some users might benefit more from reminders and progress tracking.

Psychologists found consistency with the ACT model and good care in the messages, but reported that negative responses were not always considered.

Clinicians have suggested that the ideal time for chatbot use in oncology is during and after chemo/radiation treatment, while in pregnancy it is in the first and second trimester. SH+ experts expressed that the chatbot was generally useful but highlighted the need for it to be adaptable across different stages of treatment and pregnancy. Target users indicated that the chatbot was helpful during times of high stress and change, such as during chemo/radiation and after childbirth, but less useful immediately post-diagnosis or in early pregnancy.

Set 5: Technical implementation. Reminders and notifications were considered useful for maintaining consistency. However, it was emphasised that reminders should not be too intrusive. Communication experts emphasised the importance of sending reminders in a non-disruptive way, while SH+ experts stressed the need for reminders to maintain constancy. Usability experts agreed with clinicians but suggested integrating reminders and immediate feedback to enhance user engagement.

The preference for gradual colouring of the stickers was expressed by most groups. Clinicians suggested that gradual colouring gives the idea of progress, while communication experts preferred immediate gratification. SH+ experts indicated a preference for receiving the sticker at the end, while usability experts suggested giving the reward immediately to maintain engagement. The target group prefers gradual colour progression of stickers as they complete exercises. This approach, indeed, helps visualise progress, providing satisfaction and motivation.

Set 6: Content adherence. The SH+ expert reported that our intervention on a scale of 1 to 5 adheres to a level of 4 with the original intervention. In terms of content it takes them up and is very innovative, however through the methods of transmission.

Table 4 briefly highlights key positive and negative aspects derived from the user interviews.

Table 4. Main positive and negative aspects emerged from user interviews.

Variables	Sub-variables	Positive aspects	Negative aspects (and suggestions)
Set 1: Interaction	General + alternatives for answers	Positive interaction (5); Multiple options for answers (10)	Limited alternatives in some cases (5); Fatigue in going back to previous answers (2);
	Best and worst features	Confident and personalised chatbot; Interactive videos and images (9)	Long and sometimes monotonous videos (4)
	Go back for mistakes	-	Need for option to correct mistakes (13)
Set 2: Communication mode	Clarity	Clear language (10); Suitable for all users (2)	Sometimes overly simplistic and repetitive (1);
	Emoji	Visual and easy to express emotions (8)	Offer both emojis and words for responses to cater to different preferences (6)
	Length and terms used in messages	Terms appropriate to the content (13)	Messages can be too lengthy (5)
Set 3: Involvement and constancy	Engagement	Positive and encouraging tone (1)	Improve personalization (7)
	Personalization	Use of user's name (4); Space for personal input (3)	Lacks deeper personalization (6)
Set 4: General questions	Concerns and criticism	Reflective questions post-exercises (1);	Even if user could read the message, still she/he can decide not to proceed (1)
	Ideal time	Oncology: During (11) and after (8) the treatment ; Pregnancy: second trimester (10) and after (12) childbirth	Not ideal immediately post-diagnosis (5) or in early pregnancy (4)
Set 5: Technical implementati	Reminders	Helpful for maintaining consistency (22)	Ensure reminders are supportive and not overwhelming (5)

on	Pop-up positive feedbacks	Provides instant gratification and motivation (17)	Offer immediate feedback but consider a weekly summary for sustained engagement (1)
	Progress in stickers	Gradual colouring indicates progress and provides satisfaction (19)	-
Set 6: Content adherence	Adherence to SH+ protocol	Generally aligns well with SH+ protocol (1)	Content repetitive (3)

Note. Numbers in brackets indicate the response frequency.

Discussion

This study evaluated the adaptation of the WHO SH+ intervention for stress management. The first step involved developing the SH+ protocol, which was implemented through a mobile application, with the support of an interactive ALBA chatbot. ALBA guides users through the 5-week program, corresponding to the five SH+ sessions, and facilitates navigation through various app sections. Notably, this adaptation introduced several innovations: the intervention was designed to further reinforce interaction and feedback to foster empowerment and self-efficacy, and it was tailored to female users, explicitly targeting two populations of interest in our study—pregnant women and women diagnosed with breast cancer.

Principal Findings

The ORBIT and the CeHRes comprehensive roadmap approaches were adopted to evaluate the app, starting with a preliminary phase focused on refining the dialogues and low-fidelity mock-ups of the app's sections. This initial phase was crucial for developing coherent, accurate, and engaging dialogues and ensuring a reliable adaptation of the original SH+ contents. Given that the protocol had already been validated, the chatbot's structured and standardized dialogues allowed for effective transmission of the proven content without the risk of artificial hallucinations, undertaken risks, or biases, which can occur with more advanced chatbot models based on large language models (LLMs) [49]. Giving value to the methodology adopted, about content adherence, the SH+ expert rated the intervention's adherence to the original at 4 out of 5, appreciating the innovative delivery methods while maintaining core content. However, an explicit limitation of our chatbot emerged—the lack of flexibility in personalizing responses and interactions, as highlighted by participants in Phase II. Indeed, it's worth noting that psychologists found consistency with the ACT model but noted that negative responses were not always considered by ALBA. This rigidity presents a double-edged sword: while it ensures the psychological rigour of the initial intervention, it also restricts the flexibility of personalized response options [50].

Additional relevant findings from this low-fidelity review pertain to the app's organization, featuring a well-defined structure in sections evaluated during the second phase regarding usability. The integration of gamification and feedback aspects, based on user progress monitoring to maintain engagement and immediate or delayed reinforcement from classical behaviourism to sustain motivation, was also significant. The app provided customization in

session management by proposing interruptions and clarification moments for the presented content, which users could accept or decline.

From the second phase of this study, further essential results emerged that will guide the iterative development of the app. Valuable insights were gathered by involving various stakeholder groups identified for both the pregnancy and oncology contexts. Both expert groups and target users, the final app users, provided quantitative feedback through semantic differential and uMARS questionnaires and qualitative feedback through semi-structured interviews. The evaluation of dialogues and mock-ups, following modifications in Phase I, confirmed a generally positive assessment of the app and the ALBA chatbot.

In particular, the semantic differential results indicated that ALBA's communication was empathetic and fluid, the interaction with users was deemed appropriate and acceptable. Also the interviews' report highlighted that the interaction with the chatbot was generally well received, though criticisms included difficulties in navigating back to previous answers, repetitiveness, lack of novelty, and fragmented message design. Specific feedback highlighted the chatbot's confidence, relevant responses, and psycho-educational support, while suggesting improvements such as professional voices for videos, reducing videos and messages length. User satisfaction significantly improves when chatbots provide quick, relevant, and friendly responses [51]. This capability reduces wait times and enhances the perception of service efficiency. Moreover, positive interactions with chatbots not only increase short-term satisfaction but also contribute to long-term users loyalty, as users with positive experiences are more likely to return and use the service again, thereby strengthening their relationship with the app [52].

About the communication modality, from the semantic differential emerged that the language was clear and understandable. Experts valued the clarity and thoroughness of using examples and videos to investigate concepts, suggesting more concise answers and alternating text with images or videos. So, multimedia content, such as images, videos, and audio provided by the chatbot, was also positively recognized both in semantic differential and interviews outputs. While the chatbot's language was praised for simplicity and clarity also in interviews, some found the tone slow and recommended a more active tone, and the use of emojis was well-received with a recommendation to offer both emoji and word response options; the target group found terms satisfactory but suggested reducing message length. Overall, the session structure was considered appropriately lengthy and moderately light in content according to the semantic differential results. The importance of using inclusive and comprehensible language in chatbots is increasingly recognized in academic literature, emphasizing how this can enhance user experience and engagement. Studies show that chatbots employing such language can significantly improve user satisfaction and accessibility, making interactions more effective and welcoming for a diverse audience [53,54].

Regarding the uMARS standardised questionnaire, equally encouraging results were obtained for engagement, functionality, aesthetics, and information variables. However, improvements are needed for customization and interactivity, which could have been rated more clearly positive. According to this during interviews' usability experts appreciated the guidance and psycho-educational support but noted one-way interaction and too many consecutive messages. Additional insights about engagement from interviews responses were mixed, with some finding the tone positive and encouraging, while others noted a need for more interaction and suggested improving personalization. Although personalization by name was appreciated, there was a need for more specific customization, including adding photos, avatars, and improving answer customization, and while the content was initially engaging and motivating for the target group, it became less engaging over time, leading to suggestions for reducing repetition and monotonous exercises. Enhancements are expected in these areas by better

integrating the diary and exercise sections with the chatbot's messages regarding weekly exercise management once the app is fully implemented. Additionally, features like reminders and feedback, which were not available to testers, are expected to improve the perception of app personalization. The literature emphasizes the effectiveness of engagement strategies and reminders in improving user interaction with chatbots. Recent research shows that chatbots employing personalized engagement techniques and timely reminders can significantly enhance user commitment over time and adherence to recommended actions, leading to better outcomes and increased user satisfaction [55].

Additional information from the interviews gives insights about future technical implementation: reminders and notifications were considered useful but should not be intrusive, with preferences for non-disruptive reminders and maintaining consistency, while most groups preferred gradual colouring of stickers to indicate progress, providing satisfaction and motivation.

Two very positive aspects highlighted were the app's ease of use and the high perceived credibility of the source. Both results will be further investigated in future feasibility studies using standardized tools to measure app usability [56] and user trustworthiness [57].

A good overall rating emerged from the uMARS items regarding subjective impact and perception, with behaviour change and help-seeking initiatives aligning with the intervention principles. These results are promising for the app's effective use but should be considered with the potential bias from psychological experts who favour such interventions. Additionally, participants indicated a reluctance to pay for the app hypothetically and suggested they would use it 3 to 10 times per year. While the first point may seem moderate, the app will be provided free by the healthcare system, and some willingness to pay adds value. The second point warrants further investigation, as the app is designed for continuous use over five weeks, and it is unclear if further use throughout the year implies single accesses or restarting the intervention. Interviews indicated that the app could be proposed to target women at various stages of breast cancer care or before or after childbirth in other contexts of interest. Clinicians suggested using the chatbot during and after chemo/radiation for oncology patients and during the first and second trimester for pregnant women. This flexibility is due to the different triggers and timings of stress-related issues in both contexts [58,59].

Strengths, Limitations, and Future Directions

The study highlights several strengths, limitations, and future directions for the app. Among the strengths, the chatbot's structured and standardised dialogues ensured an effective delivery of the validated content, empathetic dialogues and integration of gamification and feedback mechanisms were positively received from participants. Additionally, the app's ease of use and high perceived credibility of the contents coming from a validated WHO protocol were crucial for user adoption. However, notable limitations included the chatbot's rigidity in personalising responses and interactions and some repetitive content.

There was also a clear need for more specific customization options and improving answer personalization. While initially engaging, the content became less motivating over time, prompting suggestions to reduce repetition and monotonous exercises.

Future directions involve enhancing the chatbot's flexibility and personalization by incorporating user-specific elements. Implementing reminders and feedback mechanisms is also expected to improve personalization perception. Furthermore, conducting future feasibility studies using standardised tools to measure app usability and user trustworthiness will be essential. Exploring the frequency and context of app use over a year will help better understand user needs and improve continuous engagement and further tailoring for improved

effectiveness to pregnancy and oncological contexts.

Conclusions

This research evaluated the adaptation of SH+ intervention for stress management through a mobile application guided by the ALBA chatbot. The implementation of the protocol tailored for pregnant women and women with breast cancer diagnosis showcased several innovations, including interactive elements, gamification, and personalised feedback mechanisms. Utilising the ORBIT and CeHRes methodologies, the study validated the structured dialogues of the chatbot for effective content transmission, while acknowledging limitations such as rigidity in personalization. Despite these challenges, the app's organisation, user-friendly interface, and perceived credibility were notable strengths identified through participant feedback. Moving forward, addressing customization shortcomings, enhancing engagement strategies, and conducting further usability studies will be critical to refining the app's effectiveness and user satisfaction across diverse healthcare contexts.

Acknowledgements

We extend our gratitude to all study participants who dedicated their time, shared their knowledge, and contributed with their invaluable personal and professional experiences to the research.

Conflicts of Interest

None declared.

Appendix 1

Table 1. List of variables, sub-variables and adjectives investigated through semantic differential.

Variables	Sub-Variables	Adjectives
Communication	Empathy and listening	judgmental - welcoming passive listening - active listening alarming - reassuring indifferent - sensitive cold - warm
	Smoothness and fluidity	non-flowing - flowing
	Chatbot interaction	boring - engaging inefficient - efficient slow - fast pressing - adequate
	Lexicon	abstruse - understandable technical - common
Session structure	Interaction length	long - short demanding - light

Materials	Audio tracks	unpleasant - pleasant stressful - relaxing in the way - supportive useless - functional
	Infographics and videos	unimaginative - creative hindering - supportive useless - functional

Table 2. Questions posed to participants attending the evaluation

Topic investigated	**Questions**
Set 1: Interaction	1.1 How did you feel about the interaction with ALBA? Were there enough alternatives among the replies to ALBA? 1.2 What was the feature of the interaction that you liked the most? And the one you liked the least? 1.3 Have you ever made the mistake of clicking the answer button on a question? Have you wished you could have gone back?
Set 2: Communication mode	2.1 If something was not clear to you at first, do you find there is then a way to investigate the topic further? 2.2 Did you like answering the question "How are you?" from session 2 with emojis, or did you prefer to answer with words? 2.3 Is the mode of communication (length of sentences, terms used) appropriate for the content?
Set 3: Involvement and constancy	3.1 Was the communication with the chatbot engaging? Did it entice you to get involved and be consistent in activities? 3.2 Overall, was the intervention personalized for you? Give a level of personalization from 0 to 10.
Set 4: General questions	4.1 Do you have any concerns? Do you have any criticism? ONLY CLINICIANS AND TARGETS OF THE SPECIFIC CONTEXT 4.2 When would be the ideal time to take this intervention for a pregnant woman or a breast cancer patient?
Set 5: Technical implementation	5.1 Would reminders help you be more consistent in doing exercises or completing dialogues? 5.2 Would you like to receive the pop-up message that you reached the goal as soon as you did? Or would you prefer to receive it at the end of the week? 5.3 Would you like to receive the sticker in black and white for completing the dialogue and then coloured for completing the exercises? Or would you prefer it to be gradually coloured as you proceeded to complete the exercises?
Set 6: Content adherence	ONLY SH+ EXPERT 6.1 Did the intervention comply with the original SH+ protocol? On a scale from 1 to 5?

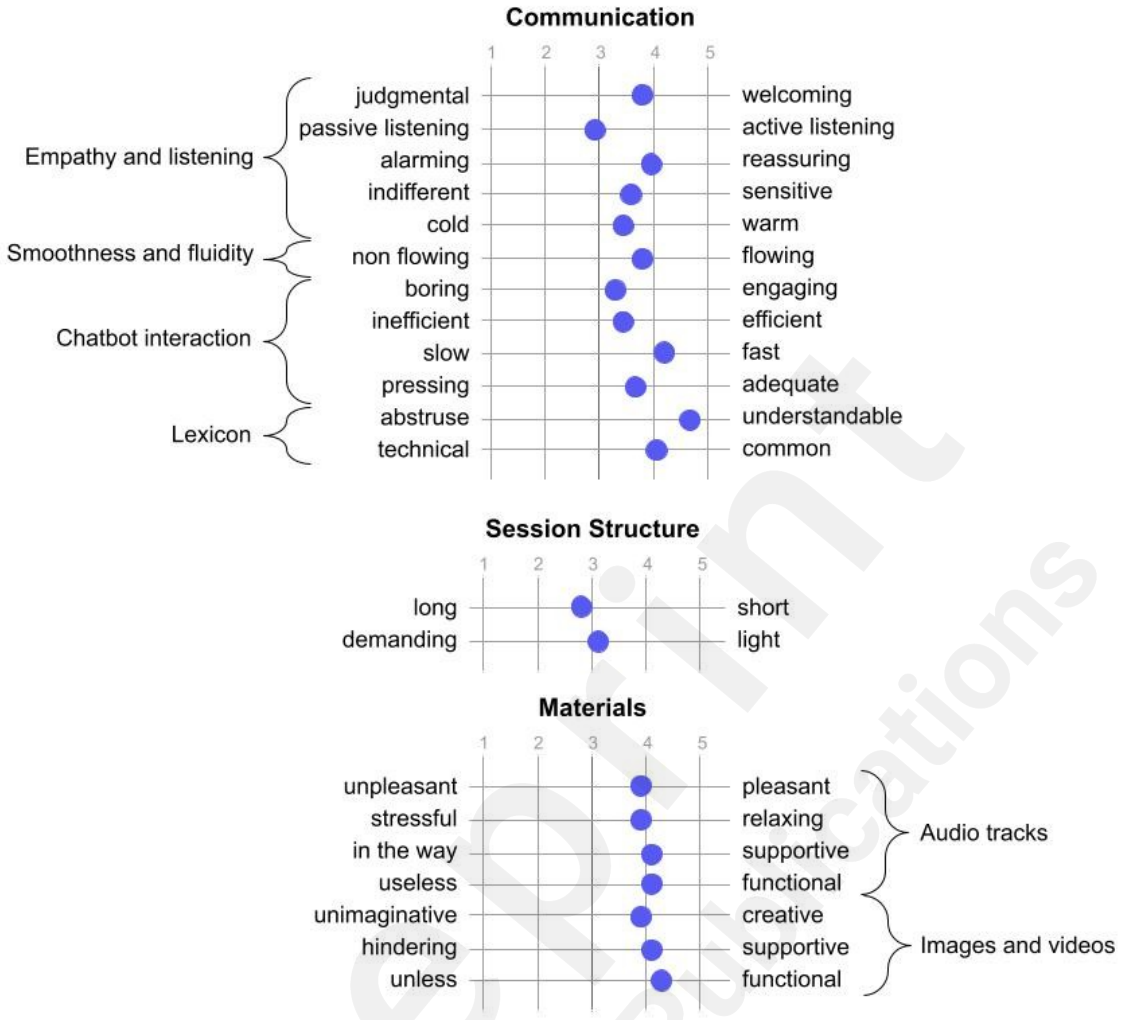

Figure 1. Graphical representation of mean values derived from the Semantic Differential.

Table 3. uMARS items Wicoxon signed ranked test analysis

	uMARS						*95% CI for Effect Size*	
	N	***M***	***Mdn***	***SD***	***W***	***P***	***r***	***Lower***	***Upper***
Entertainment	28	3.64	4.00	0.56	199.50	<.001	0.90	0.75	0.96
Interest	28	4.07	4.00	0.54	325.00	<.001	1.00	1.00	1.00
Customization	28	2.82	3.00	0.86	42.00	0.30	-0.29	-0.70	0.27
Interactivity	28	3.11	3.00	0.74	72.00	0.46	0.20	-0.36	0.65
Target group	28	4.11	4.00	0.79	231.00	<.001	1.00	1.00	1.00
Performance	28	4.11	4.00	0.69	276.00	<.001	1.00	1.00	1.00
Ease of use	28	4.50	5.00	0.64	351.00	<.00	1.00	1.00	1.00

Navigation	28	4.00	4.00	0.54	300.00	1 <.001	1.00	1.00	1.00
Gestural design	28	4.04	4.00	0.64	276.00	<.001	1.00	1.00	1.00
Layout	28	3.89	4.00	0.74	190.00	<.001	1.00	1.00	1.00
Graphics	28	3.82	4.00	0.72	171.00	<.001	1.00	1.00	1.00
Visual appeal	28	3.86	4.00	0.53	253.00	<.001	1.00	1.00	1.00
Quality of information	28	3.82	4.00	0.67	243.00	<.001	0.92	0.81	0.97
Quantity of information	28	4.21	4.00	0.69	300.00	<.001	1.00	1.00	1.00
Visual information	28	4.18	4.00	0.61	395.00	<.001	0.95	0.88	0.98
Credibility of source	27	4.59	5.00	0.69	300.00	<.001	1.00	1.00	1.00
Would you recommend	28	3.64	4.00	0.62	136.00	<.001	1.00	1.00	1.00
How many times	28	3.07	3.00	0.77	76.50	.64	0.13	-0.41	0.59
Would you pay	28	2.32	2.00	0.72	0.00	<.001	-1.00	-1.00	-1.00
Overall rating	28	3.75	4.00	0.52	210.00	<.001	1.00	1.00	1.00
Awareness	28	3.89	4.00	0.74	244.00	<.001	0.24	0.83	0.92
Knowledge	28	3.79	4.00	0.79	182.50	<.001	0.26	0.79	0.97
Attitudes	28	3.36	3.50	0.73	133.00	.02	0.26	0.10	0.82
Intention to change	28	3.54	4.00	0.69	153.00	.001	0.26	0.50	0.92
Help seeking	28	3.82	4.00	1.09	252.00	.003	0.23	0.36	0.86
Behaviour change	28	3.54	4.00	0.64	128.00	<.001	0.88	0.68	0.96

Note. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

Abbreviations

M-health: mobile health
DTx: digital therapies
SH+: Self-Help Plus

WHO: World Health Organization
ACT: Acceptance and Commitment Therapy
CBT: Cognitive-Behavioural Therapy
ORBIT: Obesity-Related Behavioral Intervention Trials
CeHRes: Center for eHealth Research and Disease Management
uMARS: User Version of the Mobile Application Rating Scale
W: Wilcoxon signed-rank test
r: rank-biserial correlation
CI: confidence interval
P: P value
N: number
M: mean
Mdn: median
SD: standard deviation
LLMs: large language models

References

1.World Health Organization. World Mental Health Report: Transforming Mental Health for All.; 2022. https://www.who.int/publications/i/item/9789240049338

2.World Health Organization. Comprehensive Mental Health Action Plan 2013–2030.; 2021. https://www.who.int/publications/i/item/9789240031029

3.Cripps M, Scarbrough H. Making Digital Health “Solutions” Sustainable in Healthcare Systems: A Practitioner Perspective. Frontiers in Digital Health. 2022;4. doi:https://doi.org/10.3389/fdgth.2022.727421

4.Park SY, Nicksic Sigmon C, Boeldt D. A Framework for the Implementation of Digital Mental Health Interventions: The Importance of Feasibility and Acceptability Research. Cureus. 2022;14(9). doi:https://doi.org/10.7759/cureus.29329

5.Mediavilla R, McGreevy KR, Felez-Nobrega M, et al. Effectiveness of a stepped-care programme of internet-based

psychological interventions for healthcare workers with psychological distress:

Study protocol for the RESPOND healthcare workers randomised controlled

trial. Digital Health. 2022;8:20552076221129084. doi:https://doi.org/10.1177/20552076221129084

6.Chen C, Wang X, Xu H, Li Y. Effectiveness of digital psychological interventions in reducing perinatal depression: a systematic review of meta-analyses. Archives of women’s mental health. 2023;26(4):423-439. doi:https://doi.org/10.1007/s00737-023-01327-y

7.World Health Organization. Scalable Psychological Interventions for People in Communities Affected by Adversity: A New Area of Mental Health and Psychosocial Work at WHO.; 2017. https://www.who.int/publications/i/item/WHO-MSD-MER-17.1

8.Guarino A, Polini C, Forte G, Favieri F, Boncompagni I, Casagrande M. The Effectiveness of Psychological Treatments in Women with Breast Cancer: A Systematic Review and Meta-Analysis. Journal of Clinical Medicine. 2020;9(1):209. doi:https://doi.org/10.3390/jcm9010209

9.Fasano J, Shao T, Huang H, Kessler AJ, Kolodka OP, Shapiro CL. Optimism and coping: do they influence health outcomes in women with breast cancer? A systemic review and meta-analysis. Breast Cancer Research and Treatment. 2020;183(3):495-501. doi:https://doi.org/10.1007/s10549-020-05800-5

10.Stewart DE, Cheung AM, Duff S, et al. Attributions of cause and recurrence in long-term breast cancer survivors. Psycho-Oncology. 2001;10(2):179-183. doi:https://doi.org/10.1002/pon.497

11.Goodman JH, Guarino A, Chenausky K, et al. CALM Pregnancy: results of a pilot study of mindfulness-based cognitive therapy for perinatal anxiety. Archives of Women's Mental Health. 2014;17(5):373-387. doi:https://doi.org/10.1007/s00737-013-0402-7

12.Woods SM, Melville JL, Guo Y, Fan MY, Gavin A. Psychosocial Stress during Pregnancy. Obstetric Anesthesia Digest. 2010;30(4):237. doi:https://doi.org/10.1097/01.aoa.0000389617.09252.47

13.Fonseca A, Gorayeb R, Canavarro MC. Women's help-seeking behaviours for depressive symptoms during the perinatal period: Socio-demographic and clinical correlates and perceived barriers to seeking professional help. Midwifery. 2015;31(12):1177-1185. doi:https://doi.org/10.1016/j.midw.2015.09.002

14.Li H, Bowen A, Bowen R, Muhajarine N, Balbuena L. Mood instability, depression, and anxiety in pregnancy and adverse neonatal outcomes. BMC Pregnancy and Childbirth. 2021;21(1). doi:https://doi.org/10.1186/s12884-021-04021-y

15.Fenwick J, Toohill J, Gamble J, et al. Effects of a midwife psycho-education intervention to reduce childbirth fear on women's birth outcomes and postpartum psychological wellbeing. BMC Pregnancy and Childbirth. 2015;15(1). doi:https://doi.org/10.1186/s12884-015-0721-y

16.Gabrielli S, Mayora Ibarra O, Forti S. A Holistic Digital Health Framework to Support Health Prevention Strategies in the First 1000 Days. JMIR Preprints. Published online December 6, 2023. Accessed July 19, 2024. https://preprints.jmir.org/preprint/55235

17.Zhang Q, Zhao H, Zheng Y. Effectiveness of mindfulness-based stress reduction (MBSR) on symptom variables and health-related quality of life in breast cancer patients—a systematic review and meta-analysis. Supportive Care in Cancer. 2018;27(3):771-781. doi:https://doi.org/10.1007/s00520-018-4570-x

18.Ghorbani V, Zanjani Z, Omidi A, Sarvizadeh M. Efficacy of Acceptance and Commitment Therapy (ACT) on depression, pain acceptance, and psychological flexibility in married people with breast cancer. Trends in Psychiatry and Psychotherapy. 2021;43(2). doi:https://doi.org/10.47626/2237-6089-2020-0022

19.Dilworth S, Higgins I, Parker V, Kelly B, Turner J. Patient and health professional's perceived barriers to the delivery of psychosocial care to adults with cancer: a systematic review. Psycho-Oncology. 2014;23(6):601-612. doi:https://doi.org/10.1002/pon.3474

20.maria.galante. Solo il 17% delle donne con tumore al seno usufruisce del supporto psico-oncologico. Europa Donna Italia. Published April 13, 2021. https://europadonna.it/2021/04/13/risultati-fortemente-2-supporto-psicologico-tumore-seno/

21.Epping-Jordan JE, Harris R, Brown FL, et al. Self-Help Plus (SH+): a new WHO stress management package. World Psychiatry. 2016;15(3):295-296. doi:https://doi.org/10.1002/wps.20355

22.Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and Commitment Therapy: Model, processes and outcomes. Behaviour Research and Therapy. 2006;44(1):1-25. doi:https://doi.org/10.1016/j.brat.2005.06.006

23.Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change. 2nd ed. Guilford Press; 2012.

24.Hayes SC, Hofmann SG. The third wave of cognitive behavioral therapy and the rise of process-based care. World Psychiatry. 2017;16(3):245-246. doi:https://doi.org/10.1002/wps.20442

25.Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The Efficacy of Cognitive Behavioral Therapy: a Review of Meta-Analyses. Cognitive Therapy and Research. 2012;36(5):427-440. doi:https://doi.org/10.1007/s10608-012-9476-1

26.SELF-HELP PLUS (SH+). www.who.int. https://www.who.int/publications/i/item/9789240035119

27.Turrini G, Purgato M, Tedeschi F, et al. Long-term effectiveness of Self-Help Plus in refugees and asylum seekers resettled in Western Europe: 12-month outcomes of a randomised controlled trial. Epidemiology and Psychiatric Sciences. 2022;31(e39). doi:https://doi.org/10.1017/s2045796022000269

28.Riello M, Purgato M, Bove C, et al. Effectiveness of self-help plus (SH+) in reducing anxiety and post-traumatic symptomatology among care home workers during the COVID-19 pandemic: a randomized controlled trial. Royal Society Open Science. 2021;8(11). doi:https://doi.org/10.1098/rsos.210219

29.Eirini Karyotaki, Marit Sijbrandij, Purgato M, et al. Self-Help Plus for refugees and asylum seekers: an individual participant data meta-analysis. BMJ Mental Health. 2023;26(1):e300672-e300672. doi:https://doi.org/10.1136/bmjment-2023-300672

30.Czajkowski SM, Powell LH, Adler N, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychology. 2015;34(10):971-982. doi:https://doi.org/10.1037/hea0000161

31.van Gemert-Pijnen JE, Nijland N, van Limburg M, et al. A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies. Journal of Medical Internet Research. 2011;13(4):e111. doi:https://doi.org/10.2196/jmir.1672

32.van Gemert-Pijnen JE, Nijland N, van Limburg M, et al. A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies. Journal of Medical Internet Research. 2011;13(4):e111. doi:https://doi.org/10.2196/jmir.1672

33.Abras C, Maloney-Krichmea D, Preece J. User-centered design. In: Bainbridge W, ed. Berkshire Encyclopedia of Human-Computer Interaction When Science Fiction Becomes Science Fact. Berkshire Publishing Group; 2004:757-768.

34.DABBS ADV, MYERS BA, MC CURRY KR, et al. User-Centered Design and Interactive Health Technologies for Patients. CIN: Computers, Informatics, Nursing. 2009;27(3):175-183. doi:https://doi.org/10.1097/ncn.0b013e31819f7c7c

35.Stickdorn M, Schneider J. This Is Service Design Thinking : Basics--Tools--Cases. Bis Publishers; 2010.

36.See A, Roller S, Douwe Kiela, Weston J. What makes a good conversation? How controllable attributes affect human judgments. North American Chapter of the Association for Computational Linguistics. Published online June 1, 2019. doi:https://doi.org/10.18653/v1/n19-1170

37.Intuitive Conversational Chatbot Builder. Landbot.io. https://landbot.io/

38.Charles Egerton Osgood, Suci GJ, Tannenbaum PH. The Measurement of Meaning. University Of Illinois Press; 1957.

39.Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR mHealth and uHealth. 2016;4(2):e72. doi:https://doi.org/10.2196/mhealth.5849

40.Morgan DL. Focus groups. In: Encyclopedia of Social Measurement. Elsevier; 2005:51-57.

41.Adlin T, Pruitt J. The Essential Persona Lifecycle: Your Guide to Building and Using Personas. Morgan Kaufmann; 2010.

42.Morselli S, Sebastianelli A, Domnich A, et al. Translation and validation of the Italian version of the user version of the Mobile Application Rating Scale (uMARS). Journal of Preventive Medicine and Hygiene. 2021;62(1):E243-E248. doi:https://doi.org/10.15167/2421-4248/jpmh2021.62.1.1894

43.JASP Team. JASP (Version 0.18.2)[Computer Software].; 2024.

44.RStudio Team. RStudio | Open source & professional software for data science teams. rstudio.com. Published 2020. http://www.rstudio.com/

45.Field AP, Miles J, Field Z. Discovering Statistics Using R. Sage; 2012.

46.King AP, Eckersley RJ. Statistics for Biomedical Engineers and Scientists. Academic Press; 2019.

47.Tomczak M, Tomczak E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. Trends in Sport Sciences. 2014;21(1):19.

48.Terry G, Hayfield N, Clarke V, Braun V. Thematic analysis. In: The SAGE Handbook of Qualitative Research in Psychology. SAGE Publications Ltd; 2017:17-36.

49.Ke L, Tong S, Cheng P, Peng K. Exploring the Frontiers of LLMs in Psychological Applications: A Comprehensive Review. arXiv.org. doi:https://doi.org/10.48550/arXiv.2401.01519

50.Pryss R, Kraft R, Baumeister H, et al. Using Chatbots to Support Medical and Psychological Treatment Procedures: Challenges, Opportunities, Technologies, Reference Architecture. Studies in Neuroscience, Psychology and Behavioral Economics. Published online 2019:249-260. doi:https://doi.org/10.1007/978-3-030-31620-4_16

51.Go E, Sundar SS. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior. 2019;97:304-316. doi:https://doi.org/10.1016/j.chb.2019.01.020

52.Jenneboer L, Herrando C, Constantinides E. The impact of chatbots on customer loyalty: A systematic literature review. Journal of Theoretical and Applied Electronic Commerce Research. 2022;17(1):212-229. doi:https://doi.org/10.3390/jtaer17010011

53.Hill J, Randolph Ford W, Farreras IG. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior. 2015;49:245-250. doi:https://doi.org/10.1016/j.chb.2015.02.026

54.Chaves AP, Egbert J, Hocking T, Doerry E, Gerosa MA. Chatbots Language Design: The Influence of Language Variation on User Experience with Tourist Assistant Chatbots. ACM Transactions on Computer-Human Interaction. 2022;29(2):1-38. doi:https://doi.org/10.1145/3487193

55.Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. Journal of Medical Internet Research. 2017;19(5):e151. doi:https://doi.org/10.2196/jmir.6553

56.Lewis JR. The System Usability Scale: Past, Present, and Future. International Journal of Human–Computer Interaction. 2018;34(7):577-590. doi:https://doi.org/10.1080/10447318.2018.1455307

57.Gulati S, Sousa S, Lamas D. Design, development and evaluation of a human-computer trust scale. Behaviour & Information Technology. 2019;38(10):1004-1015. doi:https://doi.org/10.1080/0144929x.2019.1656779

58.Chaix B, Bibault JE, Pienkowski A, et al. When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot. JMIR Cancer. 2019;5(1):e12856. doi:https://doi.org/10.2196/12856

59.Chua JYX, Choolani M, Chee CYI, et al. Insights of Parents and Parents-To-Be in Using Chatbots to Improve Their Preconception, Pregnancy, and Postpartum Health: A Mixed Studies Review. Journal of Midwifery & Women's Health. Published online February 3, 2023. doi:https://doi.org/10.1111/jmwh.13472