Evaluating the Impact of Conversational and Animation Virtual Agent Features in a Mental Health App on Depressive Symptoms and User Experience in College Students: A Randomized Controlled Trial

Stephanie Six, Elizabeth Schlesener, Victoria Hill, Sabarish Babu, Kaileigh Byrne

Submitted to: JMIR Mental Health
on: October 08, 2024

Table of Contents

Evaluating the Impact of Conversational and Animation Virtual Agent Features in a Mental Health App on Depressive Symptoms and User Experience in College Students:

A Randomized Controlled Trial

Stephanie Six[1] PhD; Elizabeth Schlesener[2]; Victoria Hill[1]; Sabarish Babu[2] PhD; Kaileigh Byrne[1] PhD

[1]Department of Psychology Clemson University Clemson US
[2]Department of Human-Centered Computing Clemson University Clemson US

Corresponding Author:
Kaileigh Byrne PhD
Department of Psychology
Clemson University
418 Brackett Hall
Clemson
US

Abstract

Background: Numerous mental health applications (MHealth apps) purport to alleviate depressive symptoms. Strong evidence suggests that brief cognitive behavioral therapy (bCBT)-based MHealth apps can decrease depressive symptoms, yet there is limited research elucidating the specific features that may augment its therapeutic benefits. One potential design feature that may influence effectiveness and user experience is the inclusion of virtual agents that can mimic realistic, human face-to-face interactions.

Objective: The goal of the current experiment was to determine the effect of conversation and animation virtual agent features within a bCBT-based MHealth app on depressive symptoms and user experience in college students with and without depressive symptoms.

Methods: College students (N=209) completed a two-week intervention in which they engaged with a bCBT-based MHealth app with a customizable therapeutic virtual agent that varied in conversational and animation features. A 2 (Time: Baseline vs. Two-Week Follow-Up) x 2 (Conversational vs. Non-Conversational Agent) x 2 (Animated vs. Non-Animated Agent) randomized controlled trial was utilized to assess mental health symptoms (PHQ-8, PSS-10, and RRS questionnaires) and user experience (MAUQ, WAI, VAI questionnaires) in college students with and without current depressive symptoms. MHealth app usability and qualitative questions regarding users' perceptions of their therapeutic virtual agent interactions and customization process were assessed at follow-up.

Results: Mixed ANOVA results demonstrated a significant decrease in symptoms of depression ($P = .002$; $M = 5.50 \pm 4.86$ at follow-up vs. $M = 6.35 \pm 4.71$ at baseline), stress ($P = .005$; $M = 15.91 \pm 7.67$ at follow-up vs. $M = 17.02 \pm 6.81$ at baseline), and rumination ($P = .028$; $M = 40.42 \pm 12.96$ at follow-up vs. $M = 41.92 \pm 13.61$ at baseline); however, no significant effect of conversation or animation was observed. Findings also indicate a significant increase in user experience in animated conditions, primarily comfort in sharing goals with the agent ($F(1, 205) = 190.93$, $P < .001$) and perceived knowledgeability of the agent ($F(1, 205) = 691.82$, $P < .001$). This significant increase in animated conditions is also reflected in the user's ease of use and satisfaction ($F(1, 201) = 102.60$, $P < .001$), system information arrangement ($F(1, 201) = 123.12$, $P < .001$), and usefulness of the application ($F(1, 201) = 3667.62$, $P < .001$).

Conclusions: The current experiment provides support for bCBT-based MHealth apps featuring customizable, humanlike therapeutic virtual agents and their ability to significantly reduce negative symptomology over a brief timeframe. The app intervention reduced mental health symptoms, regardless of whether the agent included animation or conversational features, but animation features enhanced user experience. These effects were observed in both users with and without depressive symptoms.

(JMIR Preprints 08/10/2024:67330)
DOI: https://doi.org/10.2196/preprints.67330

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

Evaluating the Impact of Conversational and Animation Virtual Agent Features in a Mental Health App on Depressive Symptoms and User Experience in College Students: A Randomized Controlled Trial

Stephanie G. Six[a], Elizabeth A. Schlesener[b], Victoria Hill[a],
Sabarish Babu[b], & Kaileigh A. Byrne[a]
[a] *Department of Psychology*
[b] *Department of Human-Centered Computing*

ABSTRACT

Background: Numerous mental health applications (MHealth apps) purport to alleviate depressive symptoms. Strong evidence suggests that brief cognitive behavioral therapy (bCBT)-based MHealth apps can decrease depressive symptoms, yet there is limited research elucidating the specific features that may augment its therapeutic benefits. One potential design feature that may influence effectiveness and user experience is the inclusion of virtual agents that can mimic realistic, human face-to-face interactions.

Objective: The goal of the current experiment was to determine the effect of conversation and animation virtual agent features within a bCBT-based MHealth app on depressive symptoms and user experience in college students with and without depressive symptoms.

Methods: College students (N=209) completed a two-week intervention in which they engaged with a bCBT-based MHealth app with a customizable therapeutic virtual agent that varied in conversational and animation features. A 2 (Time: Baseline vs. Two-Week Follow-Up) x 2 (Conversational vs. Non-Conversational Agent) x 2 (Animated vs. Non-Animated Agent) randomized controlled trial was utilized to assess mental health symptoms (PHQ-8, PSS-10, and RRS questionnaires) and user experience (MAUQ, WAI, VAI questionnaires) in college students with and without current depressive symptoms. MHealth app usability and qualitative questions regarding users' perceptions of their therapeutic virtual agent interactions and customization process were assessed at follow-up.

Results: Mixed ANOVA results demonstrated a significant decrease in symptoms of depression ($P = .002$; $M = 5.50 \pm 4.86$ at follow-up vs. $M = 6.35 \pm 4.71$ at baseline), stress ($P = .005$; $M = 15.91 \pm 7.67$ at follow-up vs. $M = 17.02 \pm 6.81$ at baseline), and rumination ($P = .028$; $M = 40.42 \pm 12.96$ at follow-up vs. $M = 41.92 \pm 13.61$ at baseline); however, no significant effect of conversation or animation was observed. Findings also indicate a significant increase in user experience in animated conditions, primarily comfort in sharing goals with the agent ($F(1, 205) = 190.93$, $P < .001$) and perceived knowledgeability of the agent ($F(1, 205) = 691.82$, $P < .001$). This significant increase in animated conditions is also reflected in the user's ease of use and satisfaction ($F(1, 201) = 102.60$, $P < .001$), system information arrangement ($F(1, 201) = 123.12$, $P < .001$), and usefulness of the application ($F(1, 201) = 3667.62$, $P < .001$).

Conclusions: The current experiment provides support for bCBT-based MHealth apps featuring customizable, humanlike therapeutic virtual agents and their ability to significantly reduce negative symptomology over a brief timeframe. The app intervention reduced mental health symptoms, regardless of whether the agent included animation or conversational features, but animation features enhanced user experience. These effects were observed in both users with and without depressive symptoms.

Keywords: Depression, mental health applications, virtual agents, virtual therapeutic coach, cognitive behavioral therapy

Note. This study is based partially on dissertation work by lead author S. Six.

Evaluating the Impact of Conversational and Animated Virtual Agent Features in a Mental Health App on Depressive Symptoms and User Experience in College Students: A Randomized Controlled Trial

Depression is a debilitating, non-discriminatory mental disorder that can significantly impact an individual's physical health, relationships, motivation, and work performance [1-5]. The prevalence of depressive symptoms within the United States drastically increased from 17 million to 21 million--a nearly 25% increase --from 2018 to 2020 during the COVID-19 pandemic [6] with young adults and women disproportionately affected [7]. Despite the prevalence and negative consequences of depressive symptoms, less than half of individuals with a depression diagnosis receive professional treatment [8] due to barriers such as lack of access to care, time, and knowledge about where to go for help [9]. To address depressive symptoms, mental health applications (MHealth apps) have emerged to offer assistance and therapeutic techniques to the public.

Cognitive behavioral therapy (CBT) is a commonly utilized evidence-based form of therapy that has successfully improved the quality of life of individuals of varying ages and clinical severity levels [10-14]. Indeed, with over three-fourths of the American population owning or having regular access to smartphone technology, CBT MHealth apps represent a viable option to improve access to mental health resources [15,16]. An estimated 20,000 - 30,000 MHealth apps are available for download on various platforms [17] advertising an assortment of services directed at improving specific or overarching negative emotions, moods, or symptoms [18-21]. A form of CBT, brief cognitive behavioral therapy (bCBT) has been suggested for depressive individuals as a means of maintaining the user's attention while not requiring large amounts of the user's time or energy. This form of CBT has successfully delivered therapeutic interventions in a time-efficient manner – around 4 –16 brief sessions [22] in both sub-clinical [21,23,24] and clinical populations [25]. Several bCBT-based apps, such as MoodMission [26], Pacifica [27], and SuperBetter [28], have demonstrated effectiveness in reducing depressive symptoms. Despite their effectiveness, MHealth app usability and user experience are often ignored or viewed as secondary [29] and face multiple issues, such as high attrition rates and features which lack the support of evidence-based psychological practices and empirical research. It is therefore critical to identify MHealth app features that can enhance user experience to maximize user engagement and therapeutic benefits.

The use of virtual agents represents one potential avenue that may enhance MHealth user experience, as virtual agents can be leveraged to mimic realistic human interactions and model social connection. Virtual agents may be useful in MHealth apps for depressive symptoms because the demand for mental health resources far exceeds the supply of human therapists and clinicians [30]. Moreover, MHealth apps are typically intended as an adjuvant to standard treatment that can be used daily, while daily access to mental health professionals is nearly impossible to acquire for most individuals due to access, cost, and time constraints. The term “virtual agent” refers to a non-controllable virtual entity which includes a form of artificial intelligence (AI) [31-33]. Virtual agents in mental health applications were introduced in 1964 when ELIZA, a chatbot display using natural language processing, was used as a Rogerian Psychotherapist [34]. Since then, virtual agents have evolved into a variety of designs, including chatbots, chatbots which incorporate speech capabilities, and embodied conversational agents (ECAs) [35]. Chatbot textual displays are a common form of conversational agent design, in which the user and the agent communicate via textual interface design. This display type has shown potential in a variety of CBT-based MHealth apps, including Woebot [36], Wysa [37], and Tess [38]. Through the implementation of speech-to-text engines, chatbot interactions can utilize speech input for more natural, speech-based conversational interactions, compared to traditional chatbot interactions [39]. ECAs are similar as they are AI-driven virtual agents meant for conversational interaction; however, compared to chatbots or chatbots with speech capabilities, ECAs are even more synonymous to human face-to-face conversation [40]. ECAs provide a more natural social presence than traditional chatbot displays, due to the inclusion of both verbal and nonverbal communication. These human-like interactions and characteristics of ECAs have been shown to affect users’ positive perception of the agent in mental health interventions [41,42]. Specifically, agents with higher levels of human-like qualities improve user experiences and enjoyment [43]. These higher ratings of experience with agents are unsurprising, given that 86% percent of people prefer interactions with humans over chatbots [44].

Two key features can be embedded into virtual agents to convey the realism of human face-to-face interactions: conversation and animation features. Conversational verbal behaviors, including lip-sync and speech, are used to replicate natural, verbal communication actions. A systematic review assessed the role of digital conversational agents on psychological distress using various therapeutic techniques, mental health focuses, study designs, and technology formats. The review concluded that digital therapeutic conversational agents can successfully mitigate psychological distress [45]. Additionally, qualitative research revealed that individuals valued the virtual agent's conversational approach to therapy, high levels of empathy, human-like personality, and interactivity [45]. In contrast, if the agents became repetitive, confused, or showed a lack of understanding regarding the user's experiences or emotions, the level of connection between the user and the agent decreased and lead to feelings of frustration [45]. Furthermore, prior research reported that individuals felt more trust towards an agent if the voice matched the user's gender [46]. Lastly, users were more likely to recommend active agents to their friends and family over the agents who had lower levels of engagement [45,47,48]. However, many of the studies mentioned in the systematic review involved the assessment of chatbots, which utilized messaging to distribute conversational therapy, rather than human-like dialogue.

On the other hand, animation can be used to show facial expressions, co-speech gestures, eye movement, and nonverbal backchanneling, which are used to depict natural, conversational body actions [40,49]. Prior research has demonstrated a strong connection between animations and positive attributes of the virtual agent. Specifically, natural animations encouraged higher levels of acceptance, trust, credibility, and task appropriateness [50]. While overly expressive facial animations were deemed unrealistic and cartoon-like, users reported natural facial expressions as signifying more authoritarian traits, like respect and competency, as well as comforting traits, like calmness and warmth [51]. Furthermore, animations have shown to elicit greater emotional responses and sense of co-presence when compared to static conditions [52]. These positive traits combined, encourage natural animation to be used in collaboration with virtual agents. Despite the prevalence of Mhealth apps for depression and the use of chatbots in prior research, to the best of our knowledge, no studies have directly compared how different virtual agent features may influence mental health outcomes and user experience in users with and without depressive symptoms. Indeed, for depressive individuals who often experience negative perceptions of themselves, their situation, and the world around them [53], these comforting traits could help to improve the virtual therapy experience.

Embodied Conversational Agents (ECAs) with conversational, animation features may enhance therapeutic interactions in MHealth apps by mimicking realistic human-human interactions, which is particularly beneficial for users with depression. Prior research highlights that empathy is a crucial element in forming a strong therapeutic alliance between a client and therapist, which in turn, increases patients' ability to process their emotions significantly better than in therapeutic contexts without social interaction and empathy [54,55]. Furthermore, other work has shown that improving the working alliance between a therapist and client can improve depressive symptoms [56], which demonstrates that working alliance is critical for effective depressive symptom treatment outcomes. Through facial expressions, body language, and conversational cues, ECAs can express empathy in ways similar to human therapists, building rapport and trust that can improve therapeutic outcomes, particularly for clients with depressive symptoms [57]. By offering a lifelike presence, these agents may encourage users to feel understood and supported, much like in human-human therapy. In addition to empathy, ECAs contribute to the therapeutic process through social presence and humanlike interaction mechanics such as turn-taking, feedback, and nonverbal communication [40]. This lifelike interaction is crucial for fostering a strong working alliance, which has been linked to better therapy outcomes in real-life patient interactions [58]. Despite the potential benefits of incorporating these features into evidence-based MHealth apps, whether virtual agent features that create a lifelike interaction influence effectiveness and user experience in users with depressive symptoms remains unclear.

Contribution of Current Study and Hypotheses

This research endeavor will primarily investigate the effect of conversation and animation features of a virtual therapeutic coach over a two-week longitudinal period utilizing a smartphone MHealth app design. This study will be among the first to assess changes in symptoms of depression, stress, and rumination over a two-week period utilizing a bCBT-based MHealth app. To the best of our knowledge, this is the first study to further investigate the difference between varying levels of animation and conversation in a virtual agent on user preference and experience. The current study aims to investigate whether conversation and animation features of a virtual agent within a bCBT-based MHealth app can reduce symptoms of depression, stress, and rumination over two weeks. Additionally, the study assesses user experience using both quantitative and qualitative methods.

The study hypotheses for the quantitative analyses are outlined below:

H1: Individuals will exhibit significantly lower symptoms of depression after two weeks. It is expected this reduction will be more pronounced in the conversational compared to the non-conversation condition, and in the animated compared to the non-animated condition.

H2: Individuals with higher levels of stress and rumination symptoms will show significantly lower symptoms of depression after two weeks. It is expected this reduction will be more pronounced in the conversational compared to non-conversational condition, and in the animated compared to non-animated condition.

H3: Individuals will show higher levels of working alliance towards the agent in the conversational condition compared to the non-conversational condition, and in the animated compared to the non-animated condition.

H4: Individuals will exhibit a more positive impression of the agent in the conversational condition compared to the non-conversational condition, and in the animated compared to the non-animated condition.

H5: Individuals will have a more positive user experience with the agent in the conversational condition compared to the non-conversational condition, and in the animated compared to the non-animated condition.

In addition to these quantitative analyses, we will query participants' rationale in designing their virtual coaches in terms of gender and similarity to people they know through qualitative methods. This information will provide insight into user preferences for virtual coaches in mHealth app settings.

Rationale for Hypotheses

Prior research featuring bCBT-based MHealth apps have successfully demonstrated effectiveness in the reduction of depressive symptoms [21, 26-28]. Additionally, the utilization of conversational, animated virtual agents in such MHealth applications can impart the realism of human, face-to-face interactions, thus increasing levels of user acceptance, impressions of the agent, and perceived empathy [50,57]. Preceding literature highlights the importance of empathy in client-therapist working alliance, and its potential impact on the improvement of depressive symptoms [54,56]. Therefore, we hypothesize that the inclusion of both animated and conversational elements in the agent design for a bCBT-based MHealth app will promote lower depressive symptoms after two weeks (H1), specifically, amongst individuals with higher levels of stress and rumination symptoms (H2). Additionally, we expect that the inclusion of such design features will result in stronger working alliance (H3), more positive agent impressions (H4), and a more positive user experience (H5).

METHOD

Participants

Prior to the experiment commencement, a G*power analysis was conducted to determine the number of participants needed to maintain an 80% power level to detect a small to medium effect at the level of significance (*P*=.05). A total of 128 participants were needed to maintain the desired power level for determining whether a reduction in depressive symptoms would occur overall. An increase of 64 participants, (N=192), was needed to maintain the desired power level for determining whether a reduction in depressive symptoms would occur between the four conditions. Additional participants were recruited to account for potential data loss or attention check failure. Participants were incentivized to participate in this experiment with compensation in the form of course credit, extra credit, or a $20 Amazon gift card.

Exclusion Criteria: Participants were excluded from participating in the experiment for four reasons: 1) the individual was under the age of 18 and classified as a minor or older than 30, 2) the individual was not fluent in English, 3) the user did not have daily access to an iPhone or Android phone, and 4) the user participated in a study using a prior version of the MHealth app (Six et al., 2022). Data was excluded for three reasons: 1) the participant completed less than two CBT-based modules, 2) failed more than one attention check, or 3) did not submit the post-intervention survey.

Table 1
Participant Demographic Information by Depressive Group and Condition

<table>
<tr><th colspan="6">Total Sample: 209 Participants</th></tr>
<tr><th colspan="2">Gender</th><th colspan="2">Race</th><th colspan="2">Mental Health Diagnosis</th></tr>
<tr><td colspan="2">Female: 168/209
Male: 39/209
Non-Binary: 2/209</td><td colspan="2">White: 168/209
Asian: 19/209
Hispanic: 10/209
Black: 8/209
Bi-Racial: 3/209
American Indian or Native: 1/209</td><td colspan="2">Depression: 60/209
Anxiety: 59/209
ADHD: 21/209
OCD: 10/209
PTSD: 7/209
Bipolar II: 4/209
Eating Disorder: 2/209
Adjustment Disorder: 1/209
Trichotillomania: 1/209
Mood Disorder: 1/209</td></tr>
<tr><th colspan="3">Non-Depressive Group</th><th colspan="3">Depressive Group</th></tr>
<tr><td>M = 2.15</td><td colspan="2">SD = 1.34</td><td>M = 9.29</td><td colspan="2">SD = 3.91</td></tr>
<tr><td>M_{age} = 20.24</td><td colspan="2">SD_{age} = 2.49</td><td>M_{age} = 19.84</td><td colspan="2">SD_{age} = 2.03</td></tr>
<tr><td>Diagnosis: 16/86</td><td colspan="2">No Diagnosis = 70/86</td><td>Diagnosis: 49/123</td><td colspan="2">No Diagnosis = 74/123</td></tr>
</table>

A total of 280 participants were recruited for the study. Seventy-one participants were excluded (43 for failing to complete the post-survey, 14 for completing fewer than three modules, 11 for failing two or more attention checks, two for exceeding the pre-determined age range, and one for failing to complete the pre-survey). The final sample size contained 209 participants between the four conditions (M_{age} = 19.97, SD_{age} = 2.19; Table 1).

Measures: Demographics

A demographics questionnaire was used to collect information regarding age, gender, race, and highest level of education achieved in both the pre- and post-intervention questionnaires. Additional questions regarding frequency of gaming, eyesight, and phone usage were included in the pre-intervention questionnaire.

Measures: Mental Health Symptoms

These questionnaires were completed in both the pre- and post-intervention questionnaires, and within the MHealth app over the course of the two-week intervention.

***Depressive Symptoms Questionnaire*.** The Patient Health Questionnaire – 8 (PHQ-8) was utilized to estimate depressive symptom severity over the past two weeks ranging from mild (0-4) to severe (20+) [59,60]. Users responded to eight questions regarding negative emotions or moods and common symptoms of depression, such as feelings of hopelessness, self-hatred, anhedonia, and fatigue on a Likert scale from zero (not at all) to three (nearly every day) [59,60]. The PHQ-4, a shortened, four-item version, was utilized within the MHealth app; scores ranged from 0-12 with higher scores indicating more severe depressive symptoms [61].

***Stress Symptoms Questionnaire*.** The Perceived Stress Scale -10 (PSS-10) is a subjective assessment of the user’s stress symptoms during the past month [62,63]. The ten questions are assessed with a 5-point Likert scale (0-4); question topics include feeling out of control, confidence levels, and irritations [62]. Participants’ scores ranged from 0-40 with responses < 14 suggesting low stress levels and >26 suggesting high levels of stress. The PSS-4, a shortened, four-item version, was also included in the daily questionnaire within the MHealth app. This questionnaire included four items from the PSS-10 scale; scores ranged from 0-16 with a higher score indicating higher levels of perceived stress [64].

Rumination Symptoms Questionnaire. The Response Rumination Scale (RRS) is a 22-item questionnaire which measures subjective levels of rumination tendencies [65]. Participants are asked to indicate the frequency with which they respond, either cognitively or behaviorally, with the provided responses to negative moods, thoughts, or events on a 0 (*Almost Never*) to 4 (*Almost Always*) scale. Items include "Think 'what am I doing to deserve this?'" and "Write down what you are thinking about and analyze it" [66]. Responses are summed, ranging from 0-88 with higher scores indicating more ruminative tendencies.

Measures: User Experience and Usability

The mHealth App Usability Questionnaire (MAUQ). The MAUQ is a 21-item questionnaire comprised of three subscales: ease of use and satisfaction (MAUQ-E), system information arrangement (MAUQ-S) and usefulness (MAUQ-U) [67]. The MAUQ-E consists of 8 items that determine a participant's satisfaction with the mHealth application. The MAUQ-S is a 6-item subscale that evaluates the layout and presentation of system information. The third subscale, the MAUQ-U, is comprised of 7 items that gauge the usefulness of the app based on the user's needs and goals. Each subscale is scored on a 7-point response range, spanning from 1 (*Strongly Agree)* to 7 (*Strongly Disagree;* Zhou et al., 2019). Example items include: "I like the interface of the program" (MAUQ-E), "The program would be useful for my health and well-being" (MAUQ-U), and "Whenever I make a mistake using the program, I could recover easily and quickly" (MAUQ-S).

The Working Alliance Inventory (WAI). The WAI utilizes three subscales that measure the alliance between tasks, goals, and the bond between the participant and an agent [68]. The subscales that comprise the WAI each have twelve questions that are scored on a seven-point scale ranging from 1 (*Never*) to 7 (*Always)* [68]. This inventory is commonly used to analyze the bond between a participant and their virtual therapist, clinician or other medical expert, as well as their goals and tasks. This experiment utilizes the inventory to investigate user preference in the context of embodied conversational agents. Example questions include: “My virtual coach and I respect each other” and “I believe that my virtual coach is genuinely concerned for my welfare”.

Virtual Agent Impressions Questionnaire (VAI). The VAI uses eight items to assess user preference, specifically their willingness to continue working with the agent, level of trust, similarity and naturality of the agent, and lastly, their perception of likeability, knowledge and relationship with the virtual agent. This questionnaire is measured on a 5-point Likert scale format ranging from 1 (*Not at all*) to 5 (*Very much so*). Example questions of the VAI are “How knowledgeable was your virtual coach?” and “How natural was your conversation with your virtual coach?”.

Open-Ended Qualitative Questions

Participants were asked the follow open-ended questions: (1) Did you make your virtual coach resemble yourself or someone you know? If so, why? (2) When creating your virtual agent, you were asked to select either a masculine or feminine agent. Please explain how you selected your virtual coach's gender. What was your thought process behind the selection? (3) Do you have any suggestions for how to improve the virtual coach?

Materials

MHealth App. The MHealth app was designed, implemented, and tested by our research team. The app, called AirHeart, contains all themes and features of a version published in prior work [21] but included new features, such as a help section, additional customization options for the virtual agent, and an additional resources section.

Virtual Therapeutic Coach: The virtual agent joined the participants on their journey by helping guide them through the various topics of CBT or related psychotherapies. A dialogue framework utilizing the RTVoice Native (Android) + Amazon Web Services (AWS) Standard (iOS) text-to-speech (TTS) engines was enabled to provide audio-based dialogue to the participants. The dialogue system consisted of a custom dialogue object, which contained multiple lines of dialogue for the module scenario. A looping dialogue structure iterated through each line of text, which was converted into audio using the TTS engine. After the audio file finished playing, the loop continued onto the next line. This process continued until the end of the module's dialogue object. In the conversational agent condition, the Speech Recognition System speech-to-text (STT) engine was used to record the participant's response, then save the text to a local database. In the conversational condition's dialogue object, line numbers that required user input were cued to "pause" the dialogue loop. At this point, the STT engine was turned on to record user audio. To effectively collect all user input, the dialogue loop waited three seconds after the user finished speaking before continuing to the next line of dialogue. If the participant continued speaking before the three seconds were up, the timer would stop and the STT engine would continue recording, then combine the user's latest utterance with their previous statement. This loop would continue until the user had completed their entire response. Once the participant finished speaking, the agent would respond with one of five randomized backchanneling responses, such as "Okay" or "I see".

For the animations, the coach used both verbal and non-verbal body language, specifically lip-sync movements, facial expressions (Multimedia Appendix 2), mouth movements, and body movements. The coach was programmed to produce two different kinds of facial expressions: positive and neutral. Positive facial expressions were used after participants provided a response to an educational question, such as "What do you think Sarah should do in this instance?". Neutral facial expressions were used to convey attention on the participant when they were providing personal or serious responses to questions such as "Have you ever experienced a depressive episode?". Micro-expressions, such as eyebrow movements and blinking, were used randomly while talking and in idle position, to enhance the realism of the agent. The plugin SALSA LipSync Suite was attached to the virtual coach to match the visemes and phonemes with the audio, effectively providing realistic lip-sync animations in real time. Lastly, the coach used randomized arm and hand movements to mimic bodily non-verbal communication in realistic conversations. This non-verbal animation included gestural animations while the agent was speaking, and head nodding to visually convey that the agent heard and understood the user's responses.

The coach customization section utilized the online avatar creation service Ready Player Me (RPM). A RPM API was used to load avatar customization directly within the application. After completing the customization process, a web link for the RPM avatar was saved locally in a SQLite database. This link was then used by the RPM avatarLoader system to load the RPM avatar directly within the scene. RPM allowed participants access to various customization options within the application, such as clothes, hairstyles, hair, skin, eye colors, body and face shape, etc. Participants were asked to customize an agent after the initial creation of their account, but they could later change their agent's appearance at any time afterward within the app. The coach customizer section could be accessed from the map home page for convenience.

CBT Modules: A total of eight CBT modules were included in the MHealth app. The background displayed a realistic image of the specific location with the virtual agent in front of the current wonder of the modern world. A text box with the virtual coach's dialogue was located above their head, at the top. This MHealth app contained the same seven bCBT modules (psychoeducation, identifying and combatting maladaptive thoughts, mindfulness and meditation, problem-solving, behavioral activation, and an overall summary) as those in our group's previous design [21], with the addition of a module on episodic future thought. Episodic future thinking (EFT) involves combining prospective imagery, a part of CBT, with prompts asking about participants' details for future enjoyable events [69,70] and has been shown to increase anticipatory pleasure and joy regarding the upcoming event [69]. These modules were originally created using traditional CBT manuals and guidance books provided to therapists and clinicians [19, 71-73].

The eight modules took place at the seven wonders of the modern world plus an additional location featuring the students' university. Participants were encouraged to use their new skills and techniques learned from the app during the experiment. The order of the modules can be visualized in Multimedia Appendix 1.

TestFlight

While the AirHeart MHealth app could be placed directly on Android phones, a third-party application was necessary for downloading the app on Apple iPhones. iPhones include an extra level of security that prevents a user-created application from being directly downloaded onto the iPhone. TestFlight is a beta-testing application used to test and assess the usability, user-satisfaction, and overall quality of new and unreleased applications. This app allows Apple users access to unfinished applications directly on their iPhone or other Apple products for free. The app can be found on the Apple App Store for free and requires 5.7 MB of storage prior to download [74].

Experimental Conditions

The current experiment included four experimental conditions differing based on the level of animation (animated vs. non-animated) and conversation between the virtual agent and the participant (conversational vs. non-conversational). All conditions had access to all app features (i.e. CBT modules, journaling, mood tracker, agent customization, help section, and additional resources section).

Animation: The two levels of animation within the experiment refer to the dynamic body movements and facial expressions exhibited by the virtual coach. The animated condition included human-like non-verbal body movements, mouth movements, and gestures in association with the information provided by the virtual agent. For example, if the virtual agent informing the user about the process of the depressive spiral, the body language will be informative or educational. The facial expressions mimicked the tone of the information or conversation. For example, when the participant converses with the virtual agent. they will have an interested facial expression. The non-animated condition displayed a static, non-moving virtual agent with a blank facial expression.

Conversational: The two levels of conversation within the experiment mimicked the active and passive conditions from the previous experiment. The active conversation included a question and response engagement spread throughout the educational portions of the CBT modules. The virtual agent asked questions or instructed the participant to complete activities aloud. The passive, non-conversational condition, did not allow the user to add their input or respond to questions. Figure 1 shows the visualization of the virtual agent in the four different conditions.

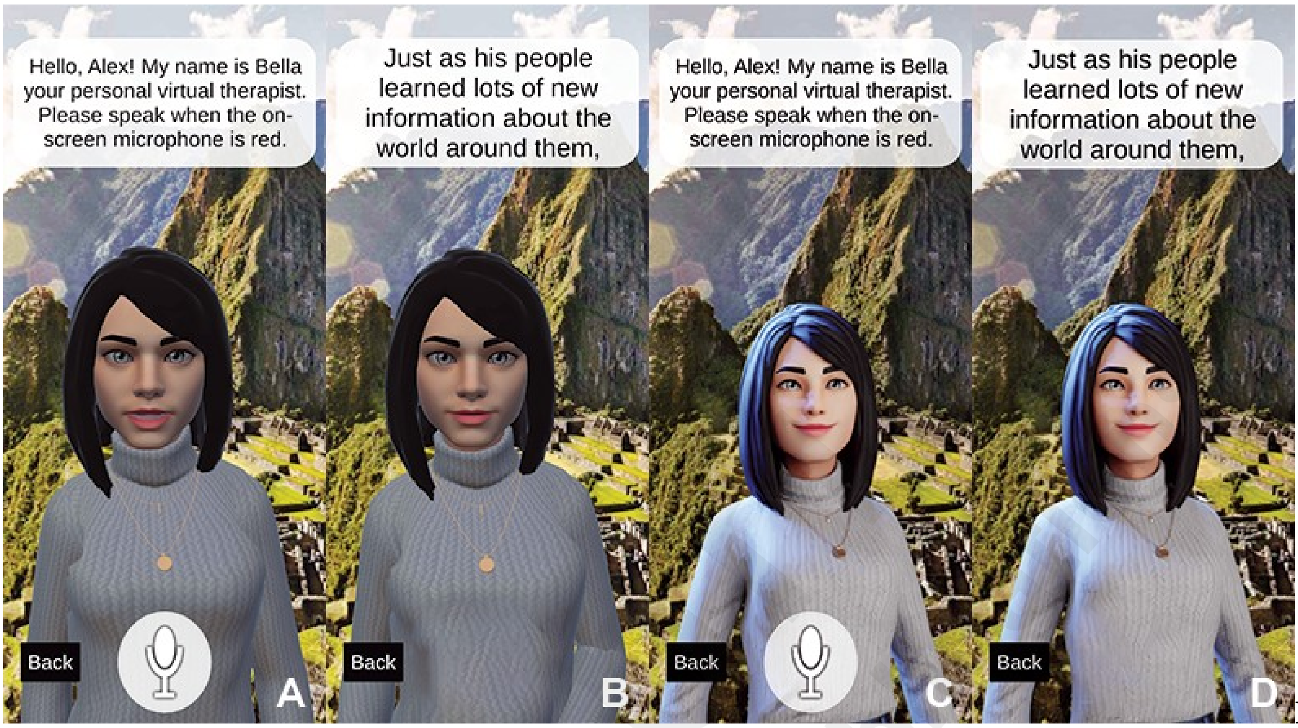

Figure 1. Example of a customized virtual agent in the (A) conversational, animated condition, (B) non-conversational, animated condition, (C) conversational, non-animated condition, and (D) non-conversational, non-animated condition.

Experimental Design

A 2 (Conversation: Present vs. Absent) x 2 (Animation: Present vs. Absent) X 2 (Time: Baseline vs. Post-Intervention Follow-Up) mixed experimental design was utilized for this experiment.

Procedure

Participants were first randomized to one of the four virtual agent conditions that varied in conversational and animation features. When participants entered the laboratory, they were instructed to read and sign a consent form before starting the experiment. Once signed, the participants completed a pre-experimental survey consisting of demographics questions, and mental health-related questionnaires (the PHQ-8, the GAD-7, the PSS-10, and RRS). Comprehension questions were hidden throughout the survey to assess attention. Upon completion of the pre-experimental survey, individuals were instructed to download the AirHeart MHealth app via Test Flight for iPhones or via Google Drive for Android phones. Users were instructed to create an account and complete the tutorial document consisting of creating a personalized virtual agent, answering the daily questionnaire, viewing their mood tracker statement, writing a short entry in their journal, and completing the first introductory CBT module. When the participant finished the tutorial, they were instructed to use the app every other day for two weeks and completed the post-survey emailed to them after the two-week mark.

Intervention Period: During the following two weeks, participants used the MHealth app a minimum of eight times for full completion, but additional usage was encouraged. When participants logged into the app for the first time that day, they were prompted to complete the daily questionnaire, view their mood tracker, and then taken to the home page where they had access to the CBT modules. Should the participant log in for a second time on the same day, they were taken straight to the home page. Additional features like agent customization and journaling will be offered at the user's discretion.

End of Intervention Assessment: After fourteen days, the participant received a follow-up survey in their email inbox. This survey included mental health related (PHQ-8, the GAD-7, and the PSS-10) and usability questionnaires as well as open-ended questions regarding user experience. Attention check questions were hidden throughout the survey to assess attention.

Data Analysis

To investigate H1 and H2, separate 2 (Conversational Status) X 2 (Animation Status) X 2 (Baseline vs. Post-Intervention Symptoms) mixed effects ANOVAs was used to analyze changes in depressive, stress, and rumination symptoms, respectively. Sensitivity analyses were conducted that focused only on participants who reported experiencing depressive symptoms (PHQ-8 scores of 5 and above).

Next, to assess H3 – H5 for the user experience predictions, separate 2 (Conversational

Status: Present vs. Absent) X 2 (Animation Status: Present vs. Absent) X 2 (Depressive Status: Depressive vs. Non-Depressive State) ANOVAs were performed. Using the validated cutoff scores established in prior work [61], PHQ-8 scores ranging from 0 – 4 were considered normal (or non-depressive) and scores of 5 and above were considered in a depressive state. Inclusion of this factor allowed for distinguishing whether individuals with and without current depressive symptoms had user experience preferences for the virtual agent characteristics.

For the open-ended qualitative questions (H6), a thematic analysis was performed similar to the proceed outlined in past work [75]. Two researchers reviewed de-identified responses and created categories using axial coding. After coding was completed, the study conditions (i.e., conversational, animated) and depressive group (Depressive vs. Non-Depressive) were re-attached to the responses to create a frequency data table.

RESULTS

Mental Health Symptoms

Change in Depressive Symptoms. The 2 (Conversational vs. Non-Conversational) X 2 (Animated vs. Non-Animated) X 2 (Time: Baseline vs. Post-Intervention) mixed ANOVA results demonstrated a statistically significant main effect of time ($F(1, 205) = 10.06$, $P = .002$; $\eta p^2 = .05$), indicating that depressive symptoms were lower at two-week follow-up ($M = 5.50$, $SD = 4.86$) compared to baseline ($M = 6.35$, $SD = 4.71$) across all four experimental conditions. There was no significant main effect of animation condition ($F(1, 208) = .02$, $P = .91$; $\eta p^2 < .001$), conversational condition ($F(1, 208) = .25$, $P = .62$, $\eta p^2 = .001$), nor any of the interaction effect (Ps > .05)[1]. Thus, H1 is partially supported. Multimedia Appendix 3 shows the means and full ANOVA results.

Change in Stress Symptoms. Mixed ANOVA results showed a significant main effect of time ($F(1, 205) = 8.09$, $P = .005$; $\eta p^2 = .038$), such that self-reported stress levels were lower at two-week

[1] When the 2 (Conversation: Present vs. Absent) X 2 (Animation: Present vs. Absent) X 2 (Time: Pre vs. Post) analysis is performed separately for those that meet criteria of depressive symptoms at baseline (PHQ-9 scores <6) and those that do not, the results do not differ. Thus, animation and conversation features do not significantly affect change in depressive symptoms for those with or without depressive symptoms.

follow-up (M = 15.91, SD = 7.67) than baseline (M = 17.02, SD = 6.81) across all four experimental conditions. The animation condition (F(1, 208) = .007, P = .93; $\eta p^2 < .001$), conversational condition (F(1, 208) = .113, P = .74; $\eta p^2 = .001$), and all interaction effects, (Ps >.05) were non-significant (Multimedia Appendix 4).

Change in Rumination Symptoms. A main effect of Time indicated that post-intervention rumination scores were significant lower after the two-week intervention (M = 40.42, SD = 12.96) when compared to the pre-intervention scores (M = 41.92, SD = 13.61) (F(1, 205) = 4.88, P = .03; $\eta p^2 = .023$) across all four conditions. No significant effects were ascertained for animation condition (F(1, 208) = .09, P = .76; $\eta p^2 < .001$) nor the conversational condition (F(1, 208) = .37, P = .54; $\eta p^2 =$ .002). The interaction effect was also non-significant (Ps >.05; Multimedia Appendix 5).[2] These results partially support H2.

Working Alliance Inventory

Data from three subscales of the Working Alliance Inventory (WAI) are task, bond, and goal were subjected to a 2 (Conversational vs. Non-Conversational) X 2 (Animated vs. Non-Animated) X 2 (Depressive Status: Depressed vs. Non-Depressed) ANOVA. The analyses for WAI perceived that comfort in sharing tasks and bonding with the agent did not reveal any significant effects (Multimedia Appendix 6 shows the full ANOVA results).

However, the ANOVA for WAI perceived comfort in sharing goals with the agent revealed a significant main effect of Animation F(1, 205) = 190.93, P < .001, $\eta p^2 = 0.59$. Table 2 shows the results of the ANOVA.

Table 2
ANOVA results for Working Alliance Inventory: Goal

Means (M)		Standard Deviation (SD)	
Animated, N = 106	11.79	Animated	4.89
Non-Animated, N = 103	19.41	Non-Animated	2.74
Conversational, N = 105	15.79	Conversational	4.97

[2] When the analysis is performed separately for those that meet criteria of depressive symptoms at baseline (PHQ-9 scores <6) and those that do not, the results do not differ. Thus, animation and conversation features do not significantly affect change in stress or rumination symptoms for those with or without depressive symptoms.

Non-Conversational, *N* = 104	15.30	Non-Conversational	6.01
Depressed, *N* = 46	15.37	Depressed	5.67
Not Depressed, *N* = 163	15.60	Not Depressed	5.48
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp^2)**
Conversational Main Effect	1.53	.22	.008
**Animation Main Effect*	*144.37*	*<.001*	*.420*
Depressive Status Main Effect	.15	.70	.001
Animated X Conversational Interaction Effect	2.52	.11	.012
Conversation X Depressive Status Interaction Effect	1.97	.16	.010
Animation X Depressive Status Interaction Effect	1.04	.31	.005
Conversation X Animated X Depressive Status Interaction Effect	1.04	.31	.005

*Significant using alpha = 0.05

Interestingly, post-hoc pairwise comparisons using Tukey's HSD method revealed that mean WAI perceived comfort in sharing goals with the agent was significantly higher when agent animation was absent (*M* = 19.38, *SD* = 2.74) as compared to when animation was present (*M* = 11.89, *SD* = 4.97), *P* < .001.

Virtual Agent Interaction Questionnaire

ANOVA results for the question, "Knowledgeable was your virtual coach?" revealed a significant main effect of Animation, $F(1, 205) = 691.82$, $P < .001$, $\eta p^2 = 0.77$ (Table 3). Tukey's HSD post-hoc pairwise comparisons showed that mean perceived knowledgeability of the virtual therapeutic coach was significantly higher when the agent was animated (*M* = 4.01, *SD* = .83) as compared to when the agent was not (*M* = 1.99, *SD* = .10), *P* < .001.

Table 3

ANOVA results for perceived knowledgeability of the virtual therapeutic coach

Means (*M*)		**Standard Deviation (*SD*)**	
Animated, *N* = 106	4.04	Animated	0.78
Non-Animated, *N* = 103	1.98	Non-Animated	0.14
Conversational, *N* = 105	2.98	Conversational	1.14
Non-Conversational, *N* = 104	3.07	Non-Conversational	1.21
Depressed, *N* = 46	3.07	Depressed	1.24
Not Depressed, *N* = 163	3.01	Not Depressed	1.16
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp^2)**

Conversational Main Effect	.09	.77	.0004
**Animation Main Effect*	*502.75*	*<0.001*	*.71*
Depressive Status Main Effect	.60	.44	.003
Animated X Conversational Interaction Effect	.10	.75	.001
Conversation X Depressive Status Interaction Effect	1.40	.24	.007
Animation X Depressive Status Interaction Effect	1.26	.26	.006
Conversation X Animated X Depressive Status Interaction Effect	.70	.40	.003

*Significant using alpha = 0.05

The ANOVA analyses of the mean scores to the questions, “like to continue working with your coach,” “extent of trust your virtual coach,” “like your virtual coach,” “perceived relationship with your coach,” and ““How similar do you feel you are to your virtual coach?”

did not reveal any significant effects. Full results for these ANOVAs are shown in Multimedia Appendix 7.

User Experience and Usability Results

Table 4

ANOVA Results for MAUQ-Ease of Use and Satisfaction.

Means (*M*)		**Standard Deviation (*SD*)**	
Animated, *N* = 106	39.91	Animated	9.51
Non-Animated, *N* = 103	23.35	Non-Animated	9.81
Conversational, *N* = 105	32.37	Conversational	12.53
Non-Conversational, *N* = 104	31.12	Non-Conversational	12.93
Depressed, *N* = 46	32.76	Depressed	12.82
Not Depressed, *N* = 163	31.46	Not Depressed	12.71
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared** (ηp^2)
Conversational Main Effect	1.23	.27	.006
**Animation Main Effect*	*102.60*	*<0.001*	*.34*
Depressive Status Main Effect	.86	.36	.004
Animated X Conversational Interaction Effect	.32	.57	.002
Conversation X Depressive Status Interaction Effect	.024	.88	.0001
Animation X Depressive Status Interaction Effect	.024	.88	.0001
Conversation X Animated X Depressive Status Interaction Effect	.54	.46	.003

*Significant using alpha = 0.05

mHealth App Usability Questionnaire-Ease of Use and Satisfaction (MAUQ-E). The

ANOVA analysis on MAUQ-E scores revealed a significant main effect of Animation, $F(1, 201) = 102.60$, $P < .001$, $\eta p^2 = 0.34$. Table 4 displays the full ANOVA results for each main effect and interaction. Tukey's HSD post-hoc pairwise comparisons indicated that mean MAUQ-Ease of Use and Satisfaction scores was significantly higher when the agent was animated ($M = 39.95$, $SD = 9.48$) as compared to when the agent was not ($M = 23.14$, $SD = 9.61$), $P < .001$.

mHealth App Usability Questionnaire-System Information Arrangement (MAUQ-S). As presented in Table 5, ANOVA results for the MAUQ-S scores showed a significant main effect of animation, $F(1, 201) = 123.12$, $P < .001$, $\eta p^2 = .38$. Mean MAUQ-System Information Arrangement scores was significantly higher ($M = 31.00$, $SD = 6.84$) when the agent was animated as compared to when the agent was not ($M = 17.11$, $SD = 7.28$), $P < .001$.

Table 5

ANOVA Results for MAUQ-System Information Arrangement.

Means (*M*)		**Standard Deviation (*SD*)**	
Animated, *N* = 106	30.97	Animated	6.87
Non-Animated, *N* = 103	17.27	Non-Animated	7.43
Conversational, *N* = 105	24.35	Conversational	9.74
Non-Conversational, *N* = 104	24.09	Non-Conversational	10.11
Depressed, *N* = 46	23.43	Depressed	10.00
Not Depressed, *N* = 163	24.44	Not Depressed	9.89
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared** (ηp^2)
Conversational Main Effect	.16	.69	.001
**Animation Main Effect*	*123.12*	*<0.001*	*.38*
Depressive Status Main Effect	.44	.51	.002
Animated X Conversational Interaction Effect	1.24	.27	.006
Conversation X Depressive Status Interaction Effect	.027	.87	.0001
Animation X Depressive Status Interaction Effect	.34	.56	.002
Conversation X Animated X Depressive Status Interaction Effect	2.81	.096	.014

*Significant using alpha = 0.05

mHealth App Usability Questionnaire-Usefulness (MAUQ-U). The ANOVA analysis on MAUQ-U scores revealed a significant main effect of animation, revealed a significant main effect of Animation, $F(1, 201) = 3667.62$, $P < .001$, $\eta p^2 = .17$, such that mean MAUQ-Usefulness scores were significantly higher when the agent was animated (*M* = 32.27, *SD* = 9.41) than when the agent was not animated (*M* = 22.01, *SD* = 9.59), *P* < .001.

Table 6

ANOVA Results for MAUQ-Usefulness.

Means (*M*)		Standard Deviation (*SD*)	
Animated, *N* = 106	32.21	Animated	9.43
Non-Animated, *N* = 103	22.17	Non-Animated	9.70
Conversational, *N* = 105	27.81	Conversational	10.28
Non-Conversational, *N* = 104	26.71	Non-Conversational	11.29
Depressed, *N* = 46	28.59	Depressed	11.85
Not Depressed, *N* = 163	26.89	Not Depressed	10.78
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared** (ηp^2)
Conversational Main Effect	.69	.41	.003
**Animation Main Effect*	*39.91*	*<0.001*	*.16*
Depressive Status Main Effect	1.40	.24	.007
Animated X Conversational Interaction Effect	.85	.36	.004
Conversation X Depressive Status Interaction Effect	.001	.97	.000007
Animation X Depressive Status Interaction Effect	.002	.97	.000009
Conversation X Animated X Depressive Status Interaction Effect	2.70	.10	.013

*Significant using alpha = 0.05

Frequency Analysis for Agent Characteristic Selections.

***Agent Representativeness Selections*.** Ninety-five participants (45.5% of total sample) indicated that they designed the virtual agent to resemble themselves; of these participants, 55 (57.8%) were experiencing depressive symptoms than those that were not experiencing depressive symptoms (Z=1.54, *p*=.12). Seventy-seven participants (36.8% of total sample) reported that they designed the virtual agent to resemble someone they know, such as a friend, sibling, parent, or current/former therapist. Of these participants, 40 (51.9%) reported experiencing depressive symptoms (Z=0.33, *p*=.74). The remaining 37 participants (17.7%) reported making the virtual agent resemble a celebrity (n=3), a doctor or professional (n=2), or did not have a specific reason for their virtual agent design (n=32).

Agent Gender Selections. Of all participants, 84% chose a female virtual agent, and 16% chose a male. The majority of participants selected an agent's gender so that it aligned with their own gender: all but three female participants (98.2%) chose a female virtual agent, 31 of the 39 males (79.5%) selected a male virtual agent, and both non-binary participants chose a female agent.

Qualitative Results

Participants were asked to explain the reason they selected the gender of their virtual coach. Responses were collected from all 209 participants, but three were excluded for failing to supply a usable response. Two key themes emerged: relatability (n=89; 42.3%) and trust and/or comfort in talking with a particular gender about one's mental health concerns (n=160; 77.7%); note that some participants listed both reasons. Example quotes to illustrate the relatability theme are listed below:

"I chose a masculine agent because I was making a model of myself." (p #5)
"Female; I am also female." (p #116)
"I chose the same gender as mine to connect better with the therapist." (p #176)

Quotes describing the comfortability preference with a particular gender are included below:

"I selected a female therapist because I feel more comfortable talking to females about my problems. This is just my personal preference." (p #161)
"I selected female because I associate women with a more nurturing nature." (p #76)
"I chose a female because my previous therapist was female and it felt more comfortable." (p. #67)

Suggestions for improving the virtual agent were collected from all 209 participants, but 39 of them failed to provide a viable answer. Utilizing axial coding, the 170 responses were sorted into four different categories: 1) Robotic Voice/Interaction, 2) Lack of Personalization/ Customization, 2) More Engagement/Realism, 3) Technical Issues, and 4) Dislike for the Coach. Similar to the previous free response question, Z-score proportion tests were conducted for the depressive and non-depressive participants in each category. The Robotic Voice/Interaction (Z = 3.36, P <.001) was the sole categories to reach significance. A frequency data table was created to help visualize this information (see Table 7).

Table 7

Visualization of Qualitative Data: Suggestions for Virtual Therapeutic Coach Improvement

Themes	Animation vs. Non-Animation	Conversation vs. Non-Conversation	Examples	Depressive vs. Non-Depressive
Robotic Voice/ Interaction	Animated: 44/102 Non-Animated: 58/102	Conversational: 53/102 Non-Conversational: 49/102	“Make it less robotic” (p. 80) “make the voice left stiff-sounds like a robot.” (p. 101) “Possibly make the voice more realistic and not as robotic” (p. 177)	Depressive: 63/102 Non-Depressive: 39/102
More Engagement/ Interaction/ Connection	Animated: 7/19 Non-Animated: 12/19	Conversational: 9/19 Non-Conversational: 10/19	“It didn’t really feel like we were having a conversation or that she was listening to my responses” (p. 21) “Maybe be more engaging then just talking.” (p. 59)	Depressive: 11/19 Non-Depressive: 8/19
Lack of Personalization	Animated: 23/41 Non-Animated: 18/41	Conversational: 27/41 Non-Conversational: 14/41	“… they did not change their answers based on whether or not I responded so it did not feel very real.” (p. 83) “It seemed very scripted, and like I was just typing into a box.” (p. 150)	Depressive: 23/41 Non-Depressive: 18/41
Tech Issues/ UX	Animated: 8/16 Non-Animated: 8/16	Conversational: 11/16 Non-Conversational: 5/16	“Map wasn't lining up” (p. 54) “I think there should be the opportunity to rewind what the therapist says. If I missed something I would have to restart the whole module and that is frustrating.” (p. 139)	Depressive: 10/16 Non-Depressive: 6/16
No Suggestions	Animated: 21/38 Non-Animated: 17/38	Conversational: 14/38 Non-Conversational: 24/38	“No.” (p. 30) “NA” (p. 89)	Depressive: 20/38 Non-Depressive: 18/38

DISCUSSION

Principal Findings

The current randomized controlled trial sought to investigate how conversational and animated components of a virtual agent within a bCBT-based MHealth app might affect change in depressive symptoms and perceived user experience. Critically, this study compared these effects in college students both with and without current depressive symptoms. Given that individuals experiencing depressive symptoms may have negative views of themselves and/or others and may struggle with anhedonia, low energy, amongst other symptoms [1], it is reasonable that individuals experiencing depressive symptoms may have different intervention needs or preferences compared to those who are not experiencing such symptoms. The results demonstrated that bCBT delivered through a virtual agent within an MHealth app significantly reduced symptoms of depression, stress and rumination over a two-week period, regardless of whether the agent included conversational or animation features. Consequently, these results did not lend support for H1 or H2. The animation feature did enhance user experience, while the conversation feature had no significant impact. Specifically, animation increased users' comfort level in sharing information with agent and perception of the agent's knowledgeability, partially supporting H3 and H4. Users in the animation condition were also more satisfied with their overall experience using a bCBT-based MHealth application, which is consistent with H5. Furthermore, all effects for change in symptoms as well as user experience were observed overall, regardless of whether individuals reported depressive symptoms at baseline or not. These findings are particularly interesting as they indicate that the intervention was equally effective in reducing symptoms and enhancing user experience, regardless of whether users had depressive symptoms, highlighting the broad applicability of virtual agent-delivered bCBT interventions.

The findings that a two-week bCBT-based MHealth app intervention can significantly decrease depressive symptoms, stress and rumination. These results are consistent with past work showing that CBT-based MHealth apps, such as MoodMission & MoodKit [26], iCouch [76], Pacifica [27], and SuperBetter [28], can be effective in mitigating negative mental health symptoms. Moreover, small pilot studies on virtual agent-based self-monitoring technologies have shown promise in demonstrating the feasibility and preliminary efficacy in reducing depressive symptoms [77-79]. The current study advances this work by demonstrating that virtual agent-based bCBT technology can effectively reduce depressive symptoms through a sufficiently-powered randomized controlled trial. While animation and conversational features were expected to enhance the effectiveness of the intervention, particularly among those experiencing depressive symptoms, no added benefit of these features was observed on changes in depressive symptoms, stress or rumination. Past work has shown that ECAs that mimic human-human interactions may enhance perceived empathy and working alliance with the user [54-55], which may, in turn, improve intervention outcomes [56-57]. The results of the present study suggest that conversational and animation features may not be critical for establishing a meaningful connection between the virtual agent and the user in the context of bCBT MHealth apps for depression. Instead, the social presence of the human-like virtual agent alone may foster a strong working alliance.

Individuals were expected to show higher levels of working alliance towards the conversational, animated condition compared to the non-conversational, non-animated condition. However, results demonstrated that users felt that the virtual agent mutually endorsed and valued their goals more so in the non-animated conditions than the animated conditions. There were no other effects in the remaining working alliance subscales (i.e., task and bond), or any significant difference between conversational and non-conversational conditions; therefore, H3 is not supported. This effect is potentially due to perceived repetitive behaviors in the agent design, which may have led to perceived lack of empathy or understanding in the user's emotions. Prior research has shown that such user perceptions of the agent design can lead to user frustration and disconnection between the user and the agent [45]. Even so, this potential disconnection in the user-agent working alliance did not have a detrimental impact on other facets of the user's impression of the virtual agent. It was initially expected that users would exhibit a more positive impression of the agent in the conversational and animated conditions, compared to the non-conversational and non-animated conditions. Results from this study indicate that users exhibited more positive impressions of an animated virtual agent compared to a non-animated design. More specifically, users found the animated agent more knowledgeable than the non-animated versions. No significant difference was found between conversational and non-conversational conditions; therefore, H4 is partially supported. Prior literature involving embodied conversational agent (ECA) design indicates that the inclusion of animated features can positively impact user impressions of the agent, specifically regarding trust, credibility, and task performance [50]. This is likely due to the inclusion of both verbal and nonverbal behaviors, which have shown to elicit more positive user impressions in mental health interventions [41,43]. As such, this study provides results that correspond with prior work involving animated, virtual agents in mental health interactions.

It was initially assumed that individuals would have a more positive user experience with the agents in the conversational and animated conditions, compared to the non-conversational and non-animated conditions. Results from this study indicated that users' ease of use and satisfaction (MAUQ-E), system information arrangement (MAUQ-S) and usefulness (MAUQ-U) were higher in the animated agent conditions, compared to the non-animated conditions. There was no significant difference in conversational versus non-conversational conditions; therefore, H5 is partially supported. These results provide evidence that the inclusion of animation in the agent design led to stronger user ease of use and overall satisfaction with the application. Additionally, results show that the inclusion of animation led to higher perception of the application's system information arrangement, indicating that users felt that the system had higher quality. Moreover, users in the animated conditions found the overall application more useful. Prior research indicates that the inclusion of both verbal and non-verbal communication behaviors leads to user-agent interactions that are more synonymous to human face-to-face interactions, thus leading to more positive user impressions in mental health interventions [41,43]. Additionally, the inclusion of such animation design has previously demonstrated a strong connection to higher levels of agent acceptance, trust, credibility, and task appropriateness [50]. Therefore, the results from this study coincide with prior literature revolving around animation and user impressions.

The qualitative findings from the study provide insight into users' preferences in customizing their virtual therapeutic coach. Over 80% of participants selected and female virtual coach and reported designing the virtual coach so that it either resembled themselves or someone they knew, such as a friend, family member, or therapist. Numerous participants indicated a greater sense of comfort when discussing mental health concerns with females, attributing this preference to a perceived ability to relate more effectively to women on matters related to emotional and psychological well-being. The gender preference finding is consistent with research from an employer-based mental health support company that found that the majority of their customers preferred a female psychologist [80]. Social role theory proposes that women are stereotyped as more nurturing, emotionally aware, and empathetic than men [81-83]. It is possible that individuals may leverage these traditional stereotypes of women to guide their choice of therapist, both in person and in virtual form. Interestingly, however, a meta-analysis on gender differences in empathy observed that, although females self-reported higher empathy, no differences emerged on objective measures [84]. Thus, the perception of females as more empathetic, particularly in therapeutic contexts, may lack objective merit. This conflicting evidence highlights the need for further research to assess whether individuals' preferences for female therapists or therapeutic coaches impact the actual quality of therapy, both in human and virtual settings. In addition to gender preferences, participants preference for designing a virtual coach that resembled someone they knew is consistent with work showing that familiarity is comforting, particularly in contexts when people vulnerable or threatened. Familiarity conveys a sense of security and reassurance [85]. Past research with human research has demonstrated that clients prefer a therapist who shares similar attitudes and personality to themselves [86]. The current study extends this work by showing that similar preferences can be observed when choosing a virtual therapeutic coach. mHealth app users appear to find ways to identify with a virtual therapeutic coach by customizing the agent to resemble their notion of individuals who are familiar and empathetic. Thus, the ability to customize one’s virtual therapeutic coach may be critical to elicit a positive user experience in mHealth settings.

Findings from the qualitative analysis revealed that a large majority of participants cited the virtual agent as having a robotic voice and recommended improvements to the voice quality. The

MHealth app in the present study utilized a combination of AWS Polly Standard Voice (iOS) and RTVoice Native (Android) TTS technologies for speech synthesis. Both options can provide a somewhat robotic, synthetic tonality, reminiscent of Siri or Google assistant. Prior research on the perception of voices, both human-like and synthetic, has shown that synthetic, artificial voices induce an eerie feeling for participants [87,88]. Other work using a TTS conversational agent as a digital coach to help promote emotional regulation skills using CBT demonstrated a similar finding with some participants reporting that the speech felt robotic [89]. Future research should replicate this experiment utilizing a higher quality TTS or pre-recorded human voice to determine to improve interactions with the virtual therapeutic coach.

Limitations and Future Directions

As reflected in participants' qualitative feedback, one limitation of the virtual agent was that the voice had some robotic characteristics. It is possible that the quality of the virtual agent's voice may have impacted the results of the conversational feature, and future research is needed to assess whether more human-like voices in virtual agents may influence MHealth intervention effectiveness. Despite having thousands of choices for customizing the agent's facial features, hair color, and clothing, some participants still noted that they would have liked more customization options. Participants were not given extensive options for accessories or items that foster individuality, such as piercings and tattoos. It should be noted, however, that past work has shown that the ability to customize a virtual agent within a bCBT-based MHealth app does not significantly influence change in depressive symptoms or user engagement over a non-custom generic gender-neutral avatar [21]. While virtual agent customization is a preferred user feature that appears to enhance user experience, it may not be necessary to foster app engagement and effectiveness. Moreover, the study included pre- and two-week post-intervention measurements, but long-term follow-ups assess whether the effects of the intervention are sustained over time were not included in the study design. Additional research is needed to determine the duration of the benefits from the virtual agent-delivered bCBT

MHealth intervention following the conclusion of app use.

Conclusions

This study is among the first to compare the effectiveness and user experience of a virtual agent bCBT-based MHealth app in *both* users with and without depressive symptoms. The key findings from the study demonstrated that the app intervention was effective in reducing mental health symptoms, regardless of whether the agent included animation or conversational features, but animation features enhanced user experience. These effects were observed in both users with and without depressive symptoms. This work suggests that college students experiencing depressive symptoms may not have unique user experience requirements in MHealth apps, and such findings may apply more broadly to wellness apps. The finding that virtual agent animation improves user experience in MHealth apps but does not affect the intervention's effectiveness offers valuable insight for optimizing app design, which can help guide future development of digital mental health tools that are both effective and user-friendly.

Conflicts of Interest

The authors report no conflicts of interest.

Data Availability

The data will be available on the Open Science Framework upon publication acceptance.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Fifth Edition. American Psychiatric Association; 2013. doi:10.1176/appi.books.9780890425596
2. Lee H, Lee JA, Brar JS, Rush EB, Jolley CJ. Physical activity and depressive symptoms in older adults. *Geriatric Nursing*. 2014;35(1):37-41. doi:10.1016/j.gerinurse.2013.09.005
3. Orgilés M, Espada JP, Delvecchio E, et al. Anxiety and Depressive Symptoms in Children and Adolescents during COVID-19 Pandemic: A Transcultural Approach. *Psicothema*. 2021;(33.1):125-130. doi:10.7334/psicothema2020.287
4. Verhagen M, Van Der Meij A, Van Deurzen PAM, et al. Meta-analysis of the BDNF Val66Met polymorphism in major depressive disorder: effects of gender and ethnicity. *Mol Psychiatry*. 2010;15(3):260-271. doi:10.1038/mp.2008.109
5. Wang M, Ma Y, Yuan W, Su K, Li MD. Meta-Analysis of the COMT Val158Met Polymorphism in Major Depressive Disorder: Effect of Ethnicity. *J Neuroimmune Pharmacol*. 2016;11(3):434-445. doi:10.1007/s11481-016-9651-3
6. NIMH. Major Depression. National Institute of Mental Health. 2022. Updated July 2023. Accessed April 5, 2024. https://www.nimh.nih.gov/health/statistics/major-depression.
7. World Health Organization. COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. World Health Organization. March 2. 2022. Accessed April 5, 2024. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide
8. Puyat JH, Kazanjian A, Goldner EM, Wong H. How Often Do Individuals with Major Depression Receive Minimally Adequate Treatment? A Population-Based, Data Linkage

Study. *Can J Psychiatry*. 2016;61(7):394-404. doi:10.1177/0706743716640288

9. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. *Health Informatics J*. 2020;26(3):2042-2066. doi:10.1177/1460458219896492
10. Chaves C, Lopez-Gomez I, Hervas G, Vazquez C. A Comparative Study on the Efficacy of a Positive Psychology Intervention and a Cognitive Behavioral Therapy for Clinical Depression. *Cogn Ther Res*. 2017;41(3):417-433. doi:10.1007/s10608-016-9778-9
11. Gould RL, Coulson MC, Howard RJ. Cognitive Behavioral Therapy for Depression in Older People: A Meta-Analysis and Meta-Regression of Randomized Controlled Trials. *J American Geriatrics Society*. 2012;60(10):1817-1830. doi:10.1111/j.1532-5415.2012.04166.x
12. Grosse Holtforth M, Krieger T, Zimmermann J, et al. A randomized-controlled trial of cognitive–behavioral therapy for depression with integrated techniques from emotion-focused and exposure therapies. *Psychotherapy Research*. 2019;29(1):30-44. doi:10.1080/10503307.2017.1397796
13. Oud M, De Winter L, Vermeulen-Smit E, et al. Effectiveness of CBT for children and adolescents with depression: A systematic review and meta-regression analysis. *Eur psychiatr*. 2019;57:33-45. doi:10.1016/j.eurpsy.2018.12.008
14. Tandon SD, Leis JA, Mendelson T, Perry DF, Kemp K. Six-Month Outcomes from a Randomized Controlled Trial to Prevent Perinatal Depression in Low-Income Home Visiting Clients. *Matern Child Health J*. 2014;18(4):873-881. doi:10.1007/s10995-013-1313-y
15. Pew Research Center. Mobile Fact Sheet. Pew Research Center. January 31. 2024. Accessed April 25, 2024. https://www.pewresearch.org/internet/fact-sheet/mobile/
16. Purkayastha S, Addepally SA, Bucher S. Engagement and Usability of a Cognitive Behavioral Therapy Mobile App Compared With Web-Based Cognitive Behavioral Therapy Among College Students: Randomized Heuristic Trial. *JMIR Hum Factors*.

2020;7(1):e14146. doi:10.2196/14146

17. Clay, RA. Mental health apps are gaining traction. American Psychological Association. January 1, 2021. Accessed April 5, 2024. https://www.apa.org/monitor/2021/01/trends-mental-health-apps
18. Cole DA, Nick EA, Zelkowitz RL, Roeder KM, Spinelli T. Online social support for young people: Does it recapitulate in-person social support; can it help? *Computers in Human Behavior*. 2017;68:456-464. doi:10.1016/j.chb.2016.11.058
19. Miner, K., Tischler, S., DeSena, D., & Rimer, J. CBT Group Program for Depression Patient Manual. Cognitive-Behavioral Therapy (CBT) Group Program for Depression Adult Patient Manual. University of Michigan Medical School. 2016. Accessed October 29, 2021. https://medicine.umich.edu/sites/default/files/content/downloads/CBT-Group-Program-for-Depression-Patient-Manual.pdf
20. Schueller SM, Tomasino KN, Mohr DC. Integrating Human Support Into Behavioral Intervention Technologies: The Efficiency Model of Support. *Clinical Psychology: Science and Practice*. 2016;24(1):27-45. doi:https://doi.org/10.1111/cpsp.12173
21. Six SG, Byrne KA, Aly H, Harris MW. The Effect of Mental Health App Customization on Depressive Symptoms in College Students: Randomized Controlled Trial. *JMIR Ment Health*. 2022;9(8):e39516. doi:10.2196/39516
22. US Department of Veterans Affairs, V. H. A. . A Provider's Guide to Brief CBT. Va.gov: Veterans Affairs. December 12,2013 . Accessed October 20, 2021. https://www.mirecc.va.gov/visn16/guide-to-brief-cbt-manual.asp.
23. Atik E, Stricker J, Schückes M, Pittig A. Efficacy of a Brief Blended Cognitive Behavioral Therapy Program for the Treatment of Depression and Anxiety in University Students: Uncontrolled Intervention Study. *JMIR Ment Health*. 2023;10:e44742. doi:10.2196/44742
24. Richards D, Richardson T. Computer-based psychological treatments for depression: a

systematic review and meta-analysis. *Clin Psychol Rev*. 2012;32(4):329-342. doi:10.1016/j.cpr.2012.02.004

25. Smith P, Scott R, Eshkevari E, et al. Computerised CBT for depressed adolescents: Randomised controlled trial. *Behav Res Ther*. 2015;73:104-110. doi:10.1016/j.brat.2015.07.009
26. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behav Res Ther*. 2018;109:75-83. doi:10.1016/j.brat.2018.08.003
27. Moberg C, Niles A, Beermann D. Guided Self-Help Works: Randomized Waitlist Controlled Trial of Pacifica, a Mobile App Integrating Cognitive Behavioral Therapy and Mindfulness for Stress, Anxiety, and Depression. *J Med Internet Res*. 2019;21(6):e12556. doi:10.2196/12556
28. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. *Games Health J*. 2015;4(3):235-246. doi:10.1089/g4h.2014.0046
29. Neary M, Schueller SM. State of the Field of Mental Health Apps. *Cogn Behav Pract*. 2018;25(4):531-537. doi:10.1016/j.cbpra.2018.01.002
30. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. *The Lancet*. 2007;370(9590):878-889. doi:10.1016/S0140-6736(07)61239-2
31. Balakrishnan K, Honavar V. Evolutionary and Neural Synthesis of Intelligent Agents. In: Patel M, Honavar V, Balakrishnan K, eds. *Advances in the Evolutionary Synthesis of Intelligent Agents*. The MIT Press; 2001:1-28. doi:10.7551/mitpress/1129.003.0003
32. Erikson, T. Designing agents as if people mattered. Software Agents. 1997. Accessed October

5, 2024. https://tomeri.org/agents.html

33. Von Der Pütten AM, Krämer NC, Gratch J, Kang SH. "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior*. 2010;26(6):1641-1650. doi:10.1016/j.chb.2010.06.012
34. Weizenbaum, J. *Computer power and human reason: From judgment to calculation*. W. H. Freeman and Company; 1976.
35. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*. 2018;25(9):1248-1258. doi:10.1093/jamia/ocy072
36. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health*. 2017;4(2):e19. doi:10.2196/mental.7785
37. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR Mhealth Uhealth*. 2018;6(11):e12106. doi:10.2196/12106
38. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. *JMIR Ment Health*. 2018;5(4):e64. doi:10.2196/mental.9782
39. Kiuchi K, Otsu K, Hayashi Y. Psychological insights into the research and practice of embodied conversational agents, chatbots and social assistive robots: a systematic meta-review. *Behaviour & Information Technology*. Published online November 27, 2023:1-41. doi:10.1080/0144929X.2023.2286528
40. Cassell J, Bickmore T, Campbell L, Vilhjálmsson H, Yan H. Human Conversation as a System Framework: Designing Embodied Conversational Agents. In: Cassell J, Sullivan J,

Prevost S, Churchill EF, eds. *Embodied Conversational Agents*. The MIT Press; 2000:29-63. doi:10.7551/mitpress/2697.003.0004

41. Salles A, Evers K, Farisco M. Anthropomorphism in AI. *AJOB Neuroscience*. 2020;11(2):88-95. doi:10.1080/21507740.2020.1740350
42. Six S. *An Exploration Of The Efficacy Of Virtual Agents Within Smartphone Mental Health Applications*. Dissertation. Clemson University; 2024. Accessed October 5, 2024. https://open.clemson.edu/all_dissertations/3623
43. Van Pinxteren MME, Wetzels RWH, Rüger J, Pluymaekers M, Wetzels M. Trust in humanoid robots: implications for services marketing. *JSM*. 2019;33(4):507-518. doi:10.1108/JSM-01-2018-0045
44. Press G. AI stats news: 86% of consumers prefer humans to Chatbots. Forbes. December 15, 2019. Accessed September 12, 2024. https://www.forbes.com/sites/gilpress/2019/10/02/ai-stats-news-86-of-consumers-prefer-to-interact-with-a-human-agent-rather-than-a-chatbot/
45. Gaffney H, Mansell W, Tai S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *JMIR Ment Health*. 2019;6(10):e14166. doi:10.2196/14166
46. Lee SK, Kavya P, Lasser SC. Social interactions and relationships with an intelligent virtual agent. *International Journal of Human-Computer Studies*. 2021;150:102608. doi:10.1016/j.ijhcs.2021.102608
47. 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;46(5):570-582. doi:10.1017/S1352465817000820
48. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*.

2010;8(5):336-341. doi:10.1016/j.ijsu.2010.02.007

49. Bickmore T, Cassell J. Social Dialongue with Embodied Conversational Agents. In: Van Kuppevelt JCJ, Dybkjær L, Bernsen NO, eds. *Advances in Natural Multimodal Dialogue Systems*. Vol 30. Text, Speech and Language Technology. Springer Netherlands; 2005:23-54. doi:10.1007/1-4020-3933-6_2
50. Parmar D, Olafsson S, Utami D, Murali P, Bickmore T. Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Auton Agent Multi-Agent Syst*. 2022;36(1):17. doi:10.1007/s10458-021-09539-1
51. Hyde J, Carter EJ, Kiesler S, Hodgins JK. Evaluating Animated Characters: Facial Motion Magnitude Influences Personality Perceptions. *ACM Trans Appl Percept*. 2016;13(2):1-17. doi:10.1145/2851499
52. Wu Y, Babu SV, Armstrong R, et al. Effects of Virtual Human Animation on Emotion Contagion in Simulated Inter-Personal Experiences. *IEEE Trans Visual Comput Graphics*. 2014;20(4):626-635. doi:10.1109/TVCG.2014.19
53. Beckham EE, Leber WR, Watkins JT, Boyer JL, Cook JB. Development of an instrument to measure Beck's cognitive triad: the Cognitive Triad Inventory. *J Consult Clin Psychol*. 1986;54(4):566-567. doi:10.1037//0022-006x.54.4.566
54. Malin AJ, Pos AE. The impact of early empathy on alliance building, emotional processing, and outcome during experiential treatment of depression. *Psychotherapy Research*. 2015;25(4):445-459. doi:10.1080/10503307.2014.901572
55. Kim S. Therapist's Empathy, Attachment, and Therapeutic Alliance: Neurobiological Perspective. . *International Journal of Psychology & Behavior Analysis*. 2018;25(140):1-5. https://doi.org/10.15344/2018/2455-3867/140
56. Falkenström F, Ekeblad A, Holmqvist R. Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. *J*

Consult Clin Psychol. 2016;84(8):738-751. doi:10.1037/ccp0000119

57. Watson JC, Steckley PL, McMullen EJ. The role of empathy in promoting change. *Psychotherapy Research*. 2014;24(3):286-298. doi:10.1080/10503307.2013.802823
58. Bickmore T, Gruber A. Relational Agents in Clinical Psychiatry. *Harvard Review of Psychiatry*. 2010;18(2):119-130. doi:10.3109/10673221003707538
59. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x
60. Wu Y, Levis B, Riehm KE, et al. Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. *Psychol Med*. 2020;50(8):1368-1380. doi:10.1017/S0033291719001314
61. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ–4. *Psychosomatics*. 2009;50(6):613-621. doi:10.1016/S0033-3182(09)70864-3
62. Cohen S, Kamarck T, Mermelstein R. Perceived stress scale. *Measuring stress: A guide for health and social scientists*. 1994;10(2):1-2.
63. Lee EH. Review of the Psychometric Evidence of the Perceived Stress Scale. *Asian Nursing Research*. 2012;6(4):121-127. doi:10.1016/j.anr.2012.08.004
64. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983;24(4):385-396.
65. McMurrich SL, Johnson SL. Dispositional Rumination in Individuals with a Depression History. *Cogn Ther Res*. 2008;32(4):542-553. doi:10.1007/s10608-006-9093-y
66. Treynor W. Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*. 2003;27(3):247-259. doi:10.1023/A:1023910315561
67. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability

Questionnaire (MAUQ): Development and Validation Study. *JMIR Mhealth Uhealth*. 2019;7(4):e11500. doi:10.2196/11500

68. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*. 1989;36(2):223-233. doi:10.1037/0022-0167.36.2.223
69. Hallford DJ, Sharma MK, Austin DW. Increasing Anticipatory Pleasure in Major Depression through Enhancing Episodic Future Thinking: a Randomized Single-Case Series Trial. *J Psychopathol Behav Assess*. 2020;42(4):751-764. doi:10.1007/s10862-020-09820-9
70. Renner F, Werthmann J, Paetsch A, Bär HE, Heise M, Bruijniks SJE. Prospective Mental Imagery in Depression: Impact on Reward Processing and Reward-Motivated Behaviour. *Clin Psychol Eur*. 2021;3(2):e3013. doi:10.32872/cpe.3013
71. Ackerman CE. *25* CBT techniques and worksheets for cognitive behavioral therapy. PositivePsychology.com. May 20, 2021. Accessed October 5, 2024. https://positivepsychology.com/cbt-cognitive-behavioral-therapy-techniques-worksheets/
72. Cully JA, Dawson DB, Hamer J, Tharp AT. A therapist's guide to brief cognitive behavioral therapy. Houston: Department of Veterans Affairs South Central MIRECC. 2008. Accessed October 5, 2024. https://www.mirecc.va.gov/visn16/docs/therapists_guide_to_brief_cbtmanual.pdf
73. Muñoz RF, Miranda J. *Individual Therapy Manual for Cognitive-Behavioral Treatment of Depression*. Rand; 2000.
74. Test Flight, App Store. Date unknown. https://apps.apple.com/us/app/testflight/id899247664
75. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa
76. Dahne J, Collado A, Lejuez CW, et al. Pilot randomized controlled trial of a Spanish-language Behavioral Activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among united states Latinx adults with limited English proficiency. *Journal of*

Affective Disorders. 2019;250:210-217. doi:10.1016/j.jad.2019.03.009

77. Burton C, Szentagotai Tatar A, McKinstry B, et al. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. *J Telemed Telecare*. 2016;22(6):348-355. doi:10.1177/1357633X15609793
78. Pinto MD, Greenblatt AM, Hickman RL, Rice HM, Thomas TL, Clochesy JM. Assessing the Critical Parameters of eSMART-MH: A Promising Avatar-Based Digital Therapeutic Intervention to Reduce Depressive Symptoms: A Promising Avatar-Based Digital Therapeutic Intervention to Reduce Depressive Symptoms. *Perspect Psychiatr Care*. 2016;52(3):157-168. doi:10.1111/ppc.12112
79. Pinto MD, Hickman RL, Clochesy J, Buchner M. Avatar-based depression self-management technology: promising approach to improve depressive symptoms among young adults. *Applied Nursing Research*. 2013;26(1):45-48. doi:10.1016/j.apnr.2012.08.003
80. Nilo Health. Why do so many people prefer a female psychologist? Nilo Health. 2022. June 19, 2023. Accessed October 2, 2024. https://nilohealth.com/blog/why-do-so-many-people-prefer-a-female-psychologist/.
81. Christov-Moore L, Simpson EA, Coudé G, Grigaityte K, Iacoboni M, Ferrari PF. Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews*. 2014;46:604-627. doi:10.1016/j.neubiorev.2014.09.001
82. Eagly AH, Woo W, Diekman AB. Social Role Theory of Sex Differences and Similiarities: A Current Appraisal. In: Eckes T, Trautner HM, eds. *The Developmental Social Psychology of Gender*. Psychology Press; 2012:123-174.
83. Ellemers N. Gender Stereotypes. *Annu Rev Psychol*. 2018;69(1):275-298. doi:10.1146/annurev-psych-122216-011719
84. Eisenberg N, Lennon R. Sex differences in empathy and related capacities. *Psychological Bulletin*. 1983;94(1):100-131. doi:10.1037/0033-2909.94.1.100

85. De Vries M, Holland RW, Chenier T, Starr MJ, Winkielman P. Happiness Cools the Warm Glow of Familiarity: Psychophysiological Evidence That Mood Modulates the Familiarity-Affect Link. *Psychol Sci*. 2010;21(3):321-328. doi:10.1177/0956797609359878

86. Atkinson RK, Mayer RE, Merrill MM. Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*. 2005;30(1):117-139. doi:10.1016/j.cedpsych.2004.07.001

87. Abdulrahman A, Richards D. Is Natural Necessary? Human Voice versus Synthetic Voice for Intelligent Virtual Agents. *MTI*. 2022;6(7):51. doi:10.3390/mti6070051

88. Foukarakis M, Karuzaki E, Adami I, et al. Quality Assessment of Virtual Human Assistants for Elder Users. *Electronics*. 2022;11(19):3069. doi:10.3390/electronics11193069

89. Hopman K, Richards D, Norberg MM. A Digital Coach to Promote Emotion Regulation Skills. *MTI*. 2023;7(6):57. doi:10.3390/mti7060057

Supplementary Files

Multimedia Appendixes

Order of CBT Modules within the MHealth App.
URL: http://asset.jmir.pub/assets/8815ee69ec0a12db98c5c31fffab9e7d.docx

Positive (A) and Negative (B) Facial Expression Examples.
URL: http://asset.jmir.pub/assets/d57d416e017c0686c2115bd2910605b8.docx

Mixed ANOVA Results for Change in Depressive Symptoms.
URL: http://asset.jmir.pub/assets/e97892078a05676f6d74f00169886f41.docx

Mixed ANOVA Results for Change in Self-Reported Stress.
URL: http://asset.jmir.pub/assets/b7251a69aa6e28ef93d74f90a3f395e1.docx

Mixed ANOVA Results for Change in Rumination Symptoms.
URL: http://asset.jmir.pub/assets/8b1f2776d5b9a4531965e39f4f79ecc0.docx

ANOVA Results for Working Alliance Inventory.
URL: http://asset.jmir.pub/assets/ba5008399321b765dc1040934500c72d.docx

ANOVA Results for Virtual Agent Interaction Questionnaire.
URL: http://asset.jmir.pub/assets/b0a624cbda08aa249c58cd3149de423e.docx

TOC/Feature image for homepages

Example Virtual Agent in the MHealth app.