

Evaluating the Impact of Virtual Agent Conversational and Animation 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 09, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript	5
Supplementary Files	38
Multimedia Appendixes	39
Multimedia Appendix 1.....	39
Multimedia Appendix 2.....	39
Multimedia Appendix 3.....	39
Multimedia Appendix 4.....	39
Multimedia Appendix 5.....	39
CONSORT (or other) checklists.....	40
CONSORT (or other) checklist 0.....	40
TOC/Feature image for homepages	41
TOC/Feature image for homepage 0.....	42

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

Stephanie Six¹ PhD; Elizabeth Schlesener²; Victoria Hill¹; Sabarish Babu² PhD; Kaileigh Byrne¹ PhD

¹Department of Psychology Clemson University Clemson US

²Department 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 questionnaire) 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. 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 09/10/2024:67381)

DOI: <https://doi.org/10.2196/preprints.67381>

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 visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [JMIR Publications](#).

No. Please do not make my accepted manuscript PDF available to anyone.



Original Manuscript

Preprint
JMIR Publications

Evaluating the Impact of Virtual Agent Conversational and Animation 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 conversational and animation features of a virtual agent 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) \times 2 (Conversational vs. Non-Conversational Agent) \times 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 questionnaire) 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. 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 conversational or animation features, but animation features enhanced the user experience. These effects were observed in both users with and without depressive symptoms.

Registration: OSF.IO/QXSK3

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

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

INTRODUCTION

Background

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 [1] with young adults and women disproportionately affected [2]. To address depressive symptoms, mental health applications (MHealth apps) have emerged to offer assistance and therapeutic techniques to the public. Cognitive behavioral therapy (CBT)-based MHealth apps represent a viable option to improve access to mental health resources [3]. 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 [4-5] in both sub-clinical [6-8] and clinical populations [9]. Several bCBT-based apps, such as MoodMission [10], Pacifica [11], and SuperBetter [12], have demonstrated effectiveness in reducing depressive symptoms. Despite their effectiveness, it is unclear how specific app features may enhance user experience to maximize therapeutic benefits.

The use of virtual agents represents one avenue that may enhance MHealth user experience, as virtual agents can be leveraged to mimic realistic human interactions and model social connection [13-15]. The term “virtual agent” refers to a non-controllable, AI-driven virtual entity, such as chatbots and embodied conversational agents (ECAs) designed to interact with users [16-18]. Chatbots communicate with the user via a textual or voice-based interface design but typically lack a visual embodiment [19]. ECAs are characterized by a human-like visual presence and have the capability to include both verbal and/or nonverbal communication behaviors [14-15]. While chatbots have demonstrated potential in numerous bCBT-based mHealth apps, such as Woebot, Wysa, and Tess [20–22], ECAs offer a richer, more natural social presence [15], making them particularly suited for mental health interventions. Surprisingly, exceptionally few studies have evaluated the effectiveness of bCBT-based mHealth apps with ECAs [6]. This study addresses this gap by incorporating an ECA-style virtual agent into the app design.

Given that a key component of ECA-style virtual agents is visual embodiment, the physical characteristics of a virtual agent may impact user experience. Research shows that similarity between

a user's demographics and an agent's characteristics, such as gender and voice, fosters positive interactions by building trust and enhancing user motivation [23-26]. This aligns with the similarity-attraction effect, where users often prefer agents that mirror their own demographics, appearance, and voice [24]. In mental health contexts, such similarity has been shown to significantly increase users' willingness to engage in support activities [25]. To leverage these benefits, the MHealth app in this study includes customization options for the agent's physical characteristics, aiming to create a greater sense of connection and comfort during interactions.

Beyond visual embodiment, two key features can be embedded into ECA-style virtual agents to convey the realism of human face-to-face interactions: conversational and animation features. Conversational behaviors, including lip-sync and speech, are used to replicate natural, verbal communication actions [15,27-28]. Virtual agent verbal cues that align with social norms, such as greetings, small talk, and thanking, foster trust and perceived knowledgeability [29-30], particularly when the agent uses a formal, familiar voice quality and style [31]. Additionally, conversational agents can engage in turn-taking and provide feedback, mimicking the natural flow of human conversation [15]. Turn-taking allows users to feel that they are actively participating in the interaction, while feedback conveys that the agent is attentive and responsive [15]. A systematic review of mental health interventions leveraging conversational agents observed a significant reduction in psychological distress post-intervention compared to baseline [32]. These findings underscore the preliminary efficacy of virtual agents with conversational features on mental health symptoms and suggests that virtual agent conversational feature may afford empathy and interactivity that mimics therapeutic dynamics [32]. However, few studies reviewed were empirical RCTs [32], and the variability in mental health symptoms limits understanding of how conversational features affect individuals with and without depressive symptoms.

On the other hand, animation supports natural communication by conveying nonverbal behaviors, such as facial expressions, co-speech gestures, body movements, and eye gaze [14-15].

Nonverbal cues, such as nodding and eye gaze indicate active listening and foster rapport [33-35], while facial expressions can convey emotional responsiveness [36]. It is critical for animations to appear natural, as overly expressive facial animations can seem unrealistic [37]. Natural animations encourage positive attributions toward virtual agents, such as greater acceptance, trust, credibility, and task appropriateness [31, 38]. Natural animations also elicit stronger emotional responses and a greater sense of social presence compared to static or partially animated agents [39-40]. While natural animation cues, such as body movements and facial expressions, can enhance social presence, they are not always effective in conveying trustworthiness [31]. Factors like the user's age, the relevance of the animation to the task, and the context (e.g., interviews, learning, or commerce) influence how animation impacts perceived trust [31, 41], and the effectiveness of such animation features in mental health contexts remains underexplored.

In human-human therapeutic interactions, body language, tone, and other social cues are critical to conveying empathy and can influence therapeutic outcomes in individuals with depression [42-44]. Research with chatbots [13] and ECAs [45] has demonstrated that individuals experiencing depressive symptoms report high perceived virtual agent empathy and user-agent working alliance with levels mirroring that of CBT-based human interventions. These findings suggest that virtual agents may be able to mirror human-human therapeutic interactions by encouraging users to feel understood and supported. Such characteristics may be especially critical for individuals with depression, who often experience negative perceptions of themselves, others, and their environment [46]. However, no studies have directly compared how these virtual agent features (e.g., conversational vs. animation) may influence mental health outcomes and user experience in users with and without depressive symptoms.

Objectives

The current study builds on prior work [5-13,20-22, 45] in several ways. First, this study directly compares how virtual agent conversational and animation features influence user experience

in the context of MHealth apps using a randomized controlled trial design. Second, this study assesses whether either of these features within a bCBT-based MHealth app can reduce symptoms of depression, stress, and rumination over two weeks. Third, this study compares these features in a sample of users with and without depressive symptoms, addressing gaps in understanding how conversational versus animation features uniquely contribute to mental health outcomes in this population.

The study hypotheses for the quantitative analyses are outlined below:

H1: Individuals will exhibit significantly lower symptoms of depression, stress, and rumination after two weeks. This reduction will be more pronounced in the conversational and animated conditions.

H2: Individuals will have a more positive user experience with the agent in the conversational and animated conditions.

In addition to these quantitative analyses, we will query participants' rationale in designing their virtual agents in terms of gender and similarity to people they know through qualitative methods.

METHOD

Study Overview

The goal of this four-arm randomized controlled trial was to determine the effect of virtual agent conversational and animation features within a bCBT-based MHealth app on user experience and change in depressive symptoms over a two-week intervention period. The virtual agent conversational feature reflects dialogue-based interaction between the user and agent, and the virtual agent animation embodies dynamic body movements and facial expressions. Participants completed a baseline training and setup session along with baseline questionnaires via a face-to-face assessment; thereafter, participants completed the intervention and two-week post-intervention questionnaires remotely. In this section, we describe the design of the overall app and virtual agent with a focus on the manipulation of the conversational and animation features. We also describe the

methodology and analytic approach used to evaluate these features in a sample of college students with and without depressive symptoms.

Participants

Following prior research evaluating AirHeart [6], an a priori power analysis (F-test, repeated measures ANOVA, within-between interaction) was conducted for H1. The analysis aimed for 80% power to detect effects at $P=.05$ with four groups and two time points, based on prior effect sizes [47]. Results indicated a required minimum sample of 136 participants. Multimedia Appendix 1 provides detailed information. We sought to recruit $\geq 25\%$ over the minimum to account for attrition and data exclusions.

Two hundred and nine college students completed the study and were randomized to one of the four experimental conditions ($M_{age}=19.97$, $SD_{age}=2.19$; Table 1). Participants were incentivized to participate with compensation in the form of course credit, extra credit, or a \$20 Amazon gift card, depending on their choice. Participants were excluded if they were outside the 18-30 age range or did not have daily access to a smartphone. 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

Total Sample: 209 Participants					
Gender		Race		Mental Health Diagnosis	
Female: 168/209 Male: 39/209 Non-Binary: 2/209		White: 168/209 Asian: 19/209 Hispanic: 10/209 Black: 8/209 Bi-Racial: 3/209 American Indian or Native: 1/209		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	
Non-Depressive Group				Depressive Group	
$M = 2.15$		$SD = 1.34$		$M = 9.29$	
				$SD = 3.91$	

$M_{age} = 20.24$	$SD_{age} = 2.49$	$M_{age} = 19.84$	$SD_{age} = 2.03$
-------------------	-------------------	-------------------	-------------------

Materials

MHealth App. AirHeart was designed using Unity 2021 and contains all themes and features of a version published in prior work [6] but included new features, such as a help section, additional customization options for the virtual agent, and an additional resources section. The virtual agent was introduced to participants as their ‘virtual coach’ who joined them on their journey and guided them through CBT topics. Given the importance of customization features to foster user-agent similarity [24-26] users were able to customize numerous features of their agent, including facial features, body shape, and clothing. The majority of the conversational and animation feature design choices were motivated by past research describing the importance of both verbal (i.e. lip sync animation, co-speech gesturing) and nonverbal (i.e. head nods, backchanneling) behaviors in conveying natural social communication information [14-15]. All user-agent communication was conducted via natural dyadic verbal exchanges. Users initiated verbal input using a speech-to-text (STT) engine, the Speech Recognition System plugin, while audio-based agent dialogue was created using the following text-to-speech (TTS) engines: RTVoice Native for Android + Amazon Web Services (AWS) Polly Standard for iOS. Additional technical details regarding app development and virtual agent customization, conversational feature, and animation feature development can be found in Multimedia Appendix 2.

Experimental Conditions.

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

The animation feature involved dynamic body movements and facial expressions exhibited

by the virtual agent. The animated condition included human-like non-verbal body movements, mouth movements, and gestures in association with the information provided by the virtual agent. The non-animated condition displayed a static, non-moving virtual agent with a blank facial expression.

The conversational feature was characterized by user-agent interactivity in which question-and-response style dialogue was embedded within the CBT modules. The virtual agent asked questions or instructed the participant to complete activities aloud. A microphone icon provided a visual cue for users to engage in conversation with the agent— when red, the microphone is *on*, and white when *off*. The 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. Users were able to customize their agents' clothes, hairstyle, hair, skin, eye colors, body and face shape, facial cosmetics, and accessories. Within the conversational condition, the microphone icon provided a visual cue for users such that red indicated the microphone was on, and white indicated when it was off).

Measures

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+) [48,49].

Stress Symptoms Questionnaire. The Perceived Stress Scale -10 (PSS-10) is a subjective assessment of the user's stress symptoms during the past month [50,51]. Participants' scores ranged from 0-40 with responses <14 suggesting low stress levels and >26 suggesting high levels of stress.

Rumination Symptoms Questionnaire. The Response Rumination Scale (RRS) is a 22-item questionnaire which measures subjective levels of rumination tendencies [52]. Responses are summed, ranging from 0-88 with higher scores indicating more ruminative tendencies.

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) [53].

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 coach, 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?

Procedure

At the baseline assessment (Time 1), participants were first randomized to one of the four virtual agent conditions that varied in conversational and animation features. After providing written informed consent, they completed the mental health questionnaires (PHQ-8, PSS-10, RRS). Users then downloaded and piloted the AirHeart app using TestFlight, a beta-testing app required for Apple iPhones due to their additional security measures, while Android users could install the app directly. Next, they created an account, followed a tutorial to personalize their virtual agent, and completed the first CBT module. They used the app every other day for two weeks, which included 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.

After the two-week intervention, participants were contacted via email to complete post-intervention questionnaires. At this assessment (Time 2), participants completed the mental health (PHQ-8, PSS-10, RRS) questionnaires again as well as the user experience questionnaire (MAU-Q) and open-ended user experience questions.

Data Analysis

To investigate H1, separate 2 (Conversational Status: Present vs. Absent) X 2 (Animation Status: Present vs. Absent) X 2 (Time: Baseline vs. Post-Intervention Symptoms) mixed effects ANOVAs was used to analyze changes in depressive, stress, and rumination symptoms, respectively. Conversational Status and Animation Status were between-subjects factors; Time was a within-subjects repeated measures factor. Sensitivity analyses were conducted that focused only on participants who reported experiencing depressive symptoms (PHQ-8 scores >4).

To assess H2 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) multi-factorial ANOVAs were performed for each of the three MAUQ subscales. Using the validated cutoff scores established in prior work [48,49], 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 quantitative data, we conducted parametric ANOVA analyses after verifying that the data were normally distributed and error variances were equivalent [54-55]. Box's test confirmed equality of covariance matrices, and Levene's test verified homogeneity of variance. Mauchly's test ensured sphericity. When appropriate, post-hoc pairwise tests were conducted using Tukey's HSD for between-subjects variables and the Bonferroni adjusted alpha method for within-subjects variables. These methods are widely used in user studies and human factors research in computing [6,39,56-

58]. For the open-ended qualitative questions, a reflexive thematic analysis was performed in order with the procedure specified by Braun & Clarke [59] which has been used in numerous user studies evaluating virtual agents [60-64]. Two researchers independently reviewed de-identified responses, manually created initial codes, and then grouped codes into categories. For each of the three qualitative questions, percentage agreement for categories between researchers was >85%. Researchers then reviewed the independently-generated categories, consolidated duplicates, and refined and labeled themes. Next, 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 ($P_s>.05$). Multimedia Appendix 3 shows the full ANOVA results.

We note that 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.

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 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, ($P_s>.05$) 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 ($P_s>.05$; Multimedia Appendix 5).

User Experience Results

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 2 displays the full ANOVA results. 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 ($M=23.14, SD=9.61$), $P<.001$.

Table 2

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

mHealth App Usability Questionnaire-System Information Arrangement (MAUQ-S). As presented in Table 3, 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 3

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
<i>*Animation Main Effect</i>	<i>123.12</i>	<i><0.001</i>	<i>.38</i>
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

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 4).

Table 4
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 (η^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

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% (n=175) chose a female virtual agent, and 16% (n=34) 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 agent. 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. The 170 responses resulted in four different themes: 1) Robotic Voice/Interaction, 2) Lack of Personalization/ Customization, 2) More Engagement/Realism, 3) Technical Issues, and 4) Dislike for the Virtual Agent. 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 5).

Table 5

Visualization of Qualitative Data: Suggestions for Virtual Agent Improvement

Themes	Animation vs. Non-Animation	Conversation vs. Non-Conversation	Examples	Depressive vs. Non-Depressive
Robotic Voice/	Animated:	Conversational:	"Make it less robotic" (p.	Depressive:

Interaction	44/102 Non-Animated: 58/102	53/102 Non-Conversational: 49/102	80) “make the voice left stiff- sounds like a robot.” (p. 101) “Possibly make the voice more realistic and not as robotic” (p. 177)	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. 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 [65], 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 partially support H1. The animation feature did enhance user experience, while the conversation feature had no significant impact.

While several empirically-evaluated bCBT-based MHealth apps like Woebot, Wysa, Tess, and Fido [20-22, 66] include virtual agents, these existing MHealth apps leverage a text-based chatbot design. Such design does not allow for animation features and certain conversational feature components like natural speech dynamics and nonverbal behaviors. Additionally, while Tess displays a static picture of a smiling Caucasian female in the text-based chat dialogue box [22], Woebot, Wysa, and Fido do not feature a human-like graphic and instead use animals or robots [20-21,66]. Furthermore, none of these apps feature a customizable virtual agent. In contrast, the AirHeart MHealth app included a human-like, customizable virtual agent, and the conversational animated app condition featured both speech and text-based verbal capabilities, nonverbal behaviors, and dynamic animations.

Small pilot studies on virtual agent-based self-monitoring technologies have shown promise in demonstrating the feasibility and preliminary efficacy in reducing depressive symptoms [67-69]. The current study advances this work by demonstrating that virtual agent-based bCBT technology can effectively reduce depressive symptoms through a moderate-size randomized controlled trial. While conversational and animation 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 ECA-style virtual agents that mimic human-human interactions may enhance perceived empathy and working alliance with the user [45,70]. 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 be sufficient.

Study results indicated that users in the animated agent conditions reported higher ratings for system information arrangement (MAUQ-S), ease of use (MAUQ-E), and usefulness (MAUQ-U) compared to those in non-animated conditions. There was no significant difference in conversational versus non-conversational conditions; therefore, H2 is partially supported. These results suggest that animation can enhance the user experience in MHealth interventions, which aligns with prior research showing that the inclusion of both nonverbal behaviors can create more human-like interactions and improve user impressions in mental health contexts [34,35]. Additionally, the inclusion of such animation design has previously demonstrated a strong connection to higher levels of agent acceptance, trust, credibility, and task appropriateness [38]. These findings are crucial for developers of MHealth interventions, as they underscore the importance of integrating virtual agents with natural animations—such as body, mouth, and gesture movements—to enhance user satisfaction and foster human-like interactions.

Consistent with the similarity-attraction effect [24-25], most participants (>90%) selected an agent of the same gender as themselves and designed it to resemble themselves or someone familiar, such as a friend, family member, or therapist. This preference aligns with research showing that familiarity provides comfort, particularly during vulnerability [71]. Many participants reported feeling more comfortable discussing mental health with females, citing greater relatability on emotional matters. This increased relatability may be more attributable to similarity than stereotypes of females as more emotionally aware and empathetic than men [72-74]. Indeed, while females often self-report higher empathy, a meta-analysis showed no objective gender differences in empathy [75]. The findings support research demonstrating stronger therapeutic alliances when clients and counselors share the same gender, particularly among female clients [76], and users' preference for same-gender virtual agents [77-80]. In mental health contexts, gender synchrony has been shown to enhance trust in virtual agents, especially when paired with similar age [80]. These results highlight the importance of virtual agent gender customization for relatability in MHealth app design.

However, past research suggests that developers often rely on stereotypical binary gender cues, which can reinforce societal gender expectations [23]. Thus, MHealth app developers and researchers should be cognizant of the limitations of stereotypical binary gender cues and enhance features that support diverse gender representation, especially in verbal and nonverbal animations.

Limitations and Future Directions

The qualitative analysis revealed that most participants found the virtual agent's voice robotic and suggested improvements to voice quality. It is possible that the quality of the virtual agent's voice may have impacted the results of the conversational feature. The app used AWS Polly Standard Voice (iOS) and RTVoice Native (Android) for text-to-speech (TTS), both of which can sound synthetic, similar to Siri or Google Assistant. Prior research has shown that synthetic, artificial voices induce an eerie feeling [81,82], and similar results were found using a TTS agent for CBT-based emotional regulation, where participants also noted the robotic speech [83]. Future studies should explore higher quality TTS or pre-recorded human voices to enhance user interactions with the virtual agent.

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 conversational or animation 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.

Appendices

Multimedia Appendix 1: Study Power Analysis Details

Multimedia Appendix 2: Detailed Description of the AirHeart App Development, Virtual Agent Features, and Cognitive Behavioral Therapy Modules

Multimedia Appendix 3: Mixed ANOVA Results for Change in Depressive Symptoms

Multimedia Appendix 4: Mixed ANOVA Results for Change in Self-Reported

Multimedia Appendix 5: Mixed ANOVA Results for Change in Rumination Symptoms

References

1. 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>.
2. 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>
3. 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
4. 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>.
5. Turner A, Hambridge J, Baker A, Bowman J, McElduff P. Randomised controlled trial of group cognitive behaviour therapy versus brief intervention for depression in cardiac patients. *Aust N Z J Psychiatry*. 2013;47(3):235-243. doi:10.1177/0004867412460592
6. 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
7. 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
8. 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
 9. 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
 10. 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
 11. 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
 12. 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
 13. Darcy A, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. *JMIR Form Res*. 2021;5(5):e27868. doi: 10.2196/27868.
 14. Bickmore T, Cassell J. Social Dialogue 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
 15. Cassell J, Bickmore T, Campbell L, Vilhjálmsón 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
16. 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.
 17. Bertrand J, Babu SV, Polgreen P, Segre A. Virtual agents-based simulation for training healthcare workers in hand hygiene procedures. In: *Intelligent Virtual Agents: 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings*. Springer Berlin Heidelberg; 2010:125-131.
 18. 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
 19. Weizenbaum J. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman and Company; 1976.
 20. 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
 21. 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
 22. 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

23. Ghosh R, Feijóo-García PG, Stuart J, Wrenn C, Lok B. Evaluating face gender cues in virtual humans within and beyond the gender binary. *Front Virtual Real.* 2023;4:1251420. doi:10.3389/frvir.2023.1251420
24. Bernier EP, Scassellati B. The similarity-attraction effect in human-robot interaction. In: *Proceedings of the IEEE 9th International Conference on Development and Learning.* IEEE; 2010:286-290. doi:10.1109/DEVLRN.2010.5578828
25. Feijóo-García PG, Wrenn C, Stuart J, De Siqueira AG, Lok B. Participatory Design of Virtual Humans for Mental Health Support Among North American Computer Science Students: Voice, Appearance, and the Similarity-Attraction Effect. *ACM Trans Appl Percept.* 2023;20(3):1-27. doi:10.1145/3613961
26. Feijóo-García PG, Wrenn C, Gomes de Siqueira A, Ghosh R, Stuart J, Yao H, et al. Exploring the effects of user-agent and user-designer similarity in virtual human design to promote mental health intentions for college students. *ACM Trans Appl Percept.* 2024;22(1):1-41.. doi: 10.1145/3689822
27. Babu S, Schmugge S, Inugala R, Rao S, Barnes T, Hodges LF. Marve: A Prototype Virtual Human Interface Framework for Studying Human-Virtual Human Interaction. In: *Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005.* Proceedings. Springer Berlin Heidelberg; 2005:120-133. doi: 10.1007/11550617_11
28. Babu S, Schmugge S, Barnes T, Hodges LF. "What Would You Like to Talk About?" An Evaluation of Social Conversations with a Virtual Receptionist. In: *Intelligent Virtual Agents: 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21-23, 2006.* Proceedings. Springer Berlin Heidelberg; 2006:169-180. doi: 10.1007/11821830_14
29. Cassell J, Bickmore T. External Manifestations of Trustworthiness in the Interface. *Communications of the ACM.* 2000;43(12):50-56. doi:10.1145/355112.355123.
30. Li Q, Luximon Y, Zhang J. The Influence of Anthropomorphic Cues on Patients' Perceived

- Anthropomorphism, Social Presence, Trust Building, and Acceptance of Health Care Conversational Agents: Within-Subject Web-Based Experiment. *J Med Internet Res.* 2023;25:e44479. doi:10.2196/44479.
31. Rheu M, Shin JY, Peng W, Huh-Yoo J. Systematic review: Trust-building factors and implications for conversational agent design. *Int J Hum-Comput Interact.* 2021;37(1):81-96. doi: 10.1080/10447318.2020.1807710
32. 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
33. Frischen A, Bayliss AP, Tipper SP. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol Bull.* 2007;133(4):694. doi:10.1037/0033-2909.133.4.694
34. Gratch J, Lucas G. Rapport between humans and socially interactive agents. In: *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, Volume 1: Methods, Behavior, Cognition*; 2021:433-462. doi:10.1145/3477322.3477335
35. DeVault D, Artstein R, Benn G, et al. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*; 2014:1061-1068. International Foundation for Autonomous Agents and Multiagent Systems. Available from: <http://dl.acm.org/citation.cfm?id=2617388.2617415>
36. Ekman P. Facial expression and emotion. *American Psychol.* 1993;48(4):384-392. doi:10.1037/0003-066X.48.4.384
37. Hyde J, Carter EJ, Kiesler S, Hodgins JK. Evaluating animated characters: Facial motion magnitude influences personality perceptions. *ACM Trans Appl Percept.* 2016;13(2):1-1. doi:10.1145/2851499

38. 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
39. 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
40. Volonte M, Robb A, Duchowski AT, Babu SV. Empirical evaluation of virtual human conversational and affective animations on visual attention in inter-personal simulations. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*; 2018 Mar; 25-32. IEEE. doi:10.1109/VR.2018.8446364
41. Wang I, Ruiz J. Examining the use of nonverbal communication in virtual agents. *Int J Hum Comput Interact.* 2021;37(17):1648-1673. doi:10.1080/10447318.2021.1898851
42. Malin AJ, Pos AE. The impact of early empathy on alliance building, emotional processing, and outcome during experiential treatment of depression. *Psychother Res.* 2014;25(4):445–459. doi: 10.1080/10503307.2014.901572
43. 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
44. Pos AE, Greenberg LS, Goldman RN, Korman LM. Emotional processing during experiential treatment of depression. *J Consult Clin Psychol.* 2003;71:1007–1016. doi:10.1037/0022-006X.71.6.1007
45. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, O'Donnell J. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput.* 2010;22(4):289–298. doi:10.1016/j.intcom.2009.12.001
46. 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
47. Six SG, Byrne KA, Tibbett TP, Pericot-Valverde I. Examining the effectiveness of gamification in mental health apps for depression: systematic review and meta-analysis. *JMIR Ment Health*. 2021;8(11):e32199. doi:10.2196/32199
48. 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
49. 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
50. Cohen S, Kamarck T, Mermelstein R. Perceived stress scale. *Measuring stress: A guide for health and social scientists*. 1994;10(2):1-2.
51. 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
52. 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
53. 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
54. Cohen BH, Lea RB. *Essentials of statistics for the social and behavioral sciences*. Hoboken, NJ: John Wiley & Sons; 2004.
55. Wilcox R. *Modern statistics for the social and behavioral sciences: A practical introduction*. Boca Raton, FL: Chapman and Hall/CRC; 2017.
56. Babu SV, Grechkin TY, Chihak B, Ziemer C, Kearney JK, Cremer JF, Plumert JM. An immersive virtual peer for studying social influences on child cyclists' road-crossing behavior.

- IEEE Trans Vis Comput Graph.* 2010;17(1):14-25. doi:10.1109/TVCG.2010.39.
57. Volante M, Babu SV, Chaturvedi H, Newsome N, Ebrahimi E, Roy T, Daily SB, Fasolino T. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Trans Vis Comput Graph.* 2016;22(4):1326-1335. doi:10.1109/TVCG.2016.2518158.
58. Bhargava A, Bertrand JW, Gramopadhye AK, Madathil KC, Babu SV. Evaluating multiple levels of an interaction fidelity continuum on performance and learning in near-field training simulations. *IEEE Trans Vis Comput Graph.* 2018;24(4):1418-1427. doi:10.1109/TVCG.2017.2657238.
59. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology.* 2006;3(2):77-101. doi:10.1191/1478088706qp063oa
60. Karhiy M, Sagar M, Antoni M, Loveys K, Broadbent E. Can a virtual human increase mindfulness and reduce stress? A randomized trial. *Comput Hum Behav Artif Humans.* 2024;2(1):100069. doi:10.1016/j.chbah.2023.100069.
61. Loveys K, Sagar M, Pickering I, Broadbent E. A digital human for delivering a remote loneliness and stress intervention to at-risk younger and older adults during the COVID-19 pandemic: randomized pilot trial. *JMIR Ment Health.* 2021;8(11):e31586. doi:10.2196/31586.
62. Loveys K, Antoni M, Donkin L, Sagar M, Broadbent E. Comparing the feasibility and acceptability of a virtual human, teletherapy, and an e-manual in delivering a stress management intervention to distressed adult women: Pilot study. *JMIR Formativ Res.* 2023;7:e42390. doi:10.2196/42390.
63. Feijóo-García PG, Wrenn C, Gomes de Siqueira A, Ghosh R, Stuart J, Yao H, Lok B. Exploring the effects of user-agent and user-designer similarity in virtual human design to promote mental health intentions for college students. *ACM Trans Appl Percept.* 2024;22(1):1-41. doi:10.1145/1234567.

64. Kruse L, Hertel J, Mostajeran F, Schmidt S, Steinicke F. Would you go to a virtual doctor? A systematic literature review on user preferences for embodied virtual agents in healthcare. In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*; 2023 Oct 672-682. IEEE. doi: 10.1109/ISMAR59233.2023.00082
65. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Fifth Edition. American Psychiatric Association; 2013. doi:10.1176/appi.books.9780890425596
66. Karkosz, S., Szymański, R., Sanna, K., & Michałowski, J. (2024). Effectiveness of a web-based and mobile therapy chatbot on anxiety and depressive symptoms in subclinical young adults: Randomized controlled trial. *JMIR Form Res*, 8, e4796. doi: 10.2196/47960
67. 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
68. 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
69. 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
70. Bickmore T, Gruber A. Relational agents in clinical psychiatry. *Harv Rev Psychiatry*. 2010;18(2):119–130. doi: 10.3109/10673221003707538
71. 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

72. Christov-Moore L, Simpson EA, Coudé G, Grigaityte K, Iacoboni M, Ferrari PF. Empathy: Gender effects in brain and behavior. *Neurosci Biobehav Rev.* 2014;46:604-627. doi:10.1016/j.neubiorev.2014.09.001
73. Eagly AH, Woo W, Diekmann AB. Social Role Theory of Sex Differences and Similarities: A Current Appraisal. In: Eckes T, Trautner HM, eds. *The Developmental Social Psychology of Gender.* Psychology Press; 2012:123-174.
74. Ellemers N. Gender Stereotypes. *Annu Rev Psychol.* 2018;69(1):275-298. doi:10.1146/annurev-psych-122216-011719
75. Eisenberg N, Lennon R. Sex differences in empathy and related capacities. *Psychol. Bull.* 1983;94(1):100-131. doi:10.1037/0033-2909.94.1.100
76. Bhati KS. Effect of client-therapist gender match on the therapeutic relationship: An exploratory analysis. *Psychol Rep.* 2014;115(2):565–583. doi: 10.2466/21.02.PR0.115c23
77. Guadagno RE, Swinth KR, Blascovich J. Social evaluations of embodied agents and avatars. *Comput Hum Behav.* 2011;27(6):2380–2385.. doi:10.1016/j.chb.2011.07.017
78. Kim Y, Baylor AL, Shen E. Pedagogical agents as learning companions: The impact of agent emotion and gender. *J Comput Assist Learn.* 2007;23(3):220–234. doi:10.1111/j.1365-2729.2006.00210.x
79. Lee EJ, Nass C, Brave S. Can computer-generated speech have gender? An experimental test of gender stereotype. In *Proceedings of the CHI'00 Extended Abstracts on Human Factors in Computing Systems*; 2000; 289–290. doi:10.1145/633292.633461
80. Feijóo-García PG, Wrenn C, Kalogeras S, Payne C, Lok B, Omojokun O. Effects of gender synchrony in user-agent interactions: Integrating the designer as a product cue in virtual human design for mental health support. In *Proceedings of the 12th International Conference on Human-Agent Interaction*; Nov 2024; 123-131.
81. Abdulrahman A, Richards D. Is Natural Necessary? Human Voice versus Synthetic Voice for

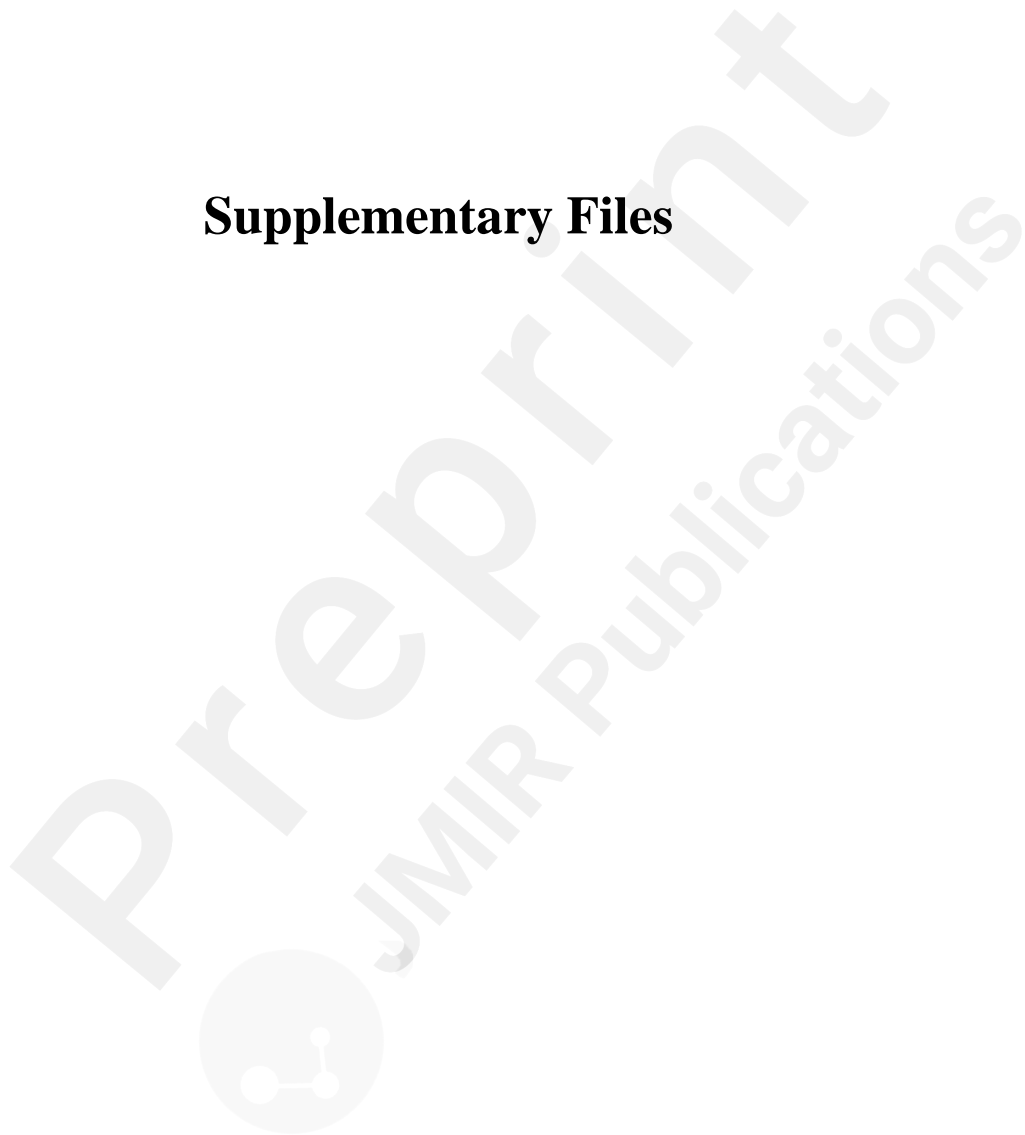
Intelligent Virtual Agents. *MTI*. 2022;6(7):51. doi:10.3390/mti6070051

82. 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](https://doi.org/10.3390/electronics11193069)

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



Supplementary Files



Multimedia Appendixes

Study Power Analysis Details.

URL: <http://asset.jmir.pub/assets/48667f6dc285eeeacb1f10441b838a8b.docx>

Detailed Description of the AirHeart App Development, Virtual Agent Features, and Cognitive Behavioral Therapy Modules.

URL: <http://asset.jmir.pub/assets/119d562713ca614a6ef0a7c443918a8f.docx>

Mixed ANOVA Results for Change in Depressive Symptoms.

URL: <http://asset.jmir.pub/assets/5aebb8211c219152bb08fd3999b1cc7f.docx>

Mixed ANOVA Results for Change in Stress.

URL: <http://asset.jmir.pub/assets/55bc664ebf150d8f8334bd06ead34106.docx>

Mixed ANOVA Results for Change in Rumination Symptoms.

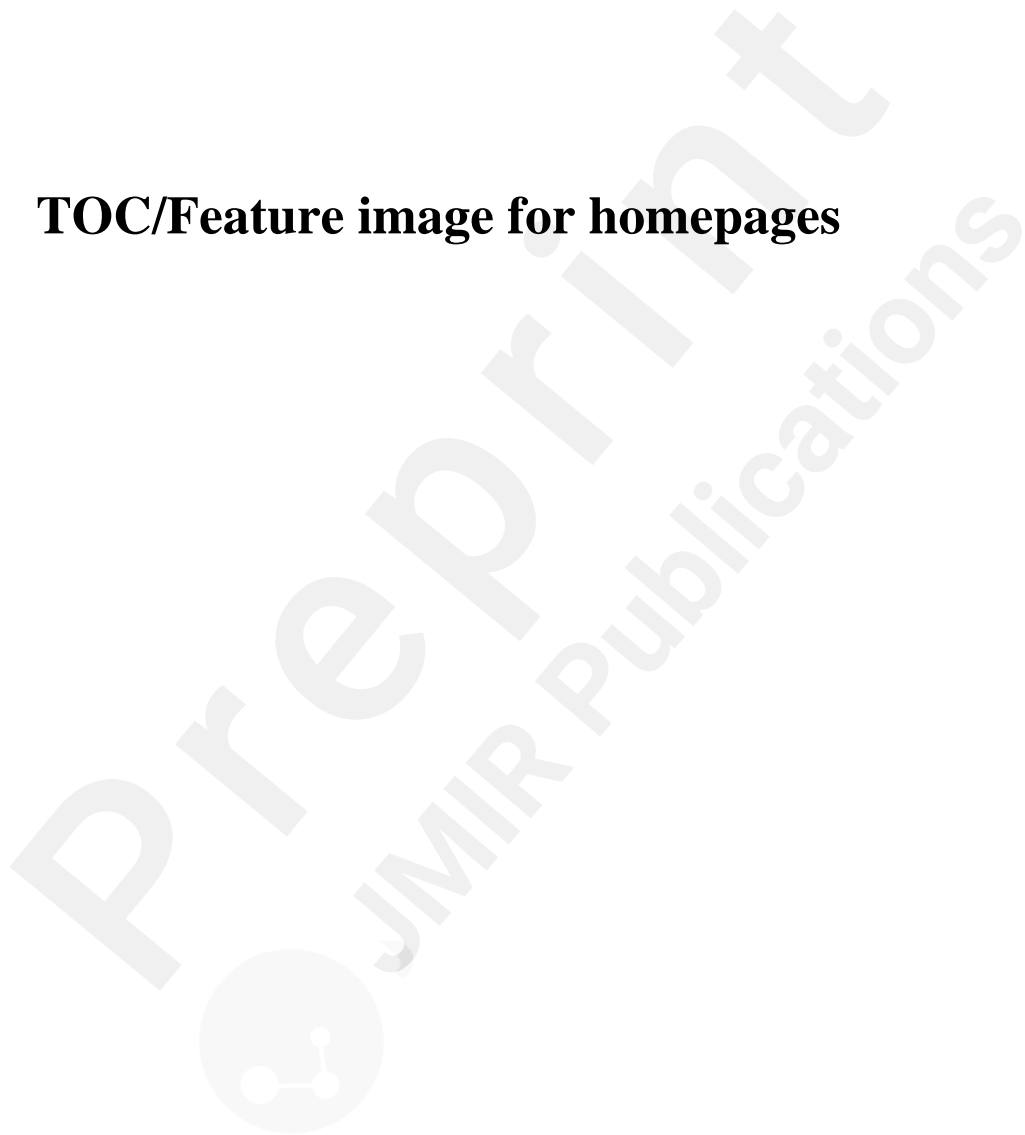
URL: <http://asset.jmir.pub/assets/0d3793dc306391efdc5d821830925793.docx>

CONSORT (or other) checklists

CONSORT Checklist from JMIR website.

URL: <http://asset.jmir.pub/assets/39a23e73cadb4419b2c541c8c7220ec6.pdf>

TOC/Feature image for homepages



Example of the virtual agent in the mental health app.

