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 09, 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 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 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 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±4.86 at follow-up vs. M = 6.35±4.71 at baseline), stress (P = .005; M = 15.91±7.67 at follow-up vs. M = 17.02±6.81 at baseline), and rumination (P = .028; M = 40.42±12.96 at follow-up vs. M = 41.92±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.

Keywords: Depression, mental health applications, virtual agents, virtual therapeutic agent, 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

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 [3,4]. 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 [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 potential avenue that may enhance MHealth user experience, as virtual agents can be leveraged to mimic realistic human interactions and model social connection. Virtual agent designs include chatbots, chatbots which incorporate speech capabilities, and embodied conversational agents (ECAs) [13]. 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 bCBT-based MHealth apps, including Woebot [14], Wysa [15], and Tess [16]. Compared to chatbots or chatbots with speech capabilities, ECAs are even more synonymous to human face-to-face conversation [17]. ECAs provide a more natural social presence than traditional chatbot displays, due to the inclusion of both verbal and nonverbal communication.

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. 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 [17,18]. 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 [19]. Animations have shown to elicit greater emotional responses and sense of co-presence when compared to static conditions [20]. 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. 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 [21]. 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 [17]. Despite the potential benefits of incorporating these features into evidence-based MHealth apps, no studies have directly compared how these 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 [22], these comforting traits could help to improve the therapeutic experience.

Contribution of Current Study and Hypotheses

This 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, 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 coaches in terms of gender and similarity to people they know through qualitative methods.

METHOD

Participants

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

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+) [23,24].

***Stress Symptoms Questionnaire*.** The Perceived Stress Scale -10 (PSS-10) is a subjective assessment of the user's stress symptoms during the past month [25,26]. 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 [27]. 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) [28].

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 [6] but included new features, such as a help section, additional customization options for the virtual agent, and an additional resources section. Technical details regarding the CBT modules and virtual agent can be found in Multimedia Appendix 1.

TestFlight. While the AirHeart 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 of new and unreleased applications. The app can be found on the Apple App Store for free and requires 5.7 MB of storage prior to download [29].

Experimental Conditions

The current experiment included four experimental conditions differing based on the level of animation (animated vs. non-animated) and conversation (conversational vs. non-conversational) between the virtual agent and the participant. 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. The non-animated condition displayed a static, non-moving virtual agent with a blank facial expression.

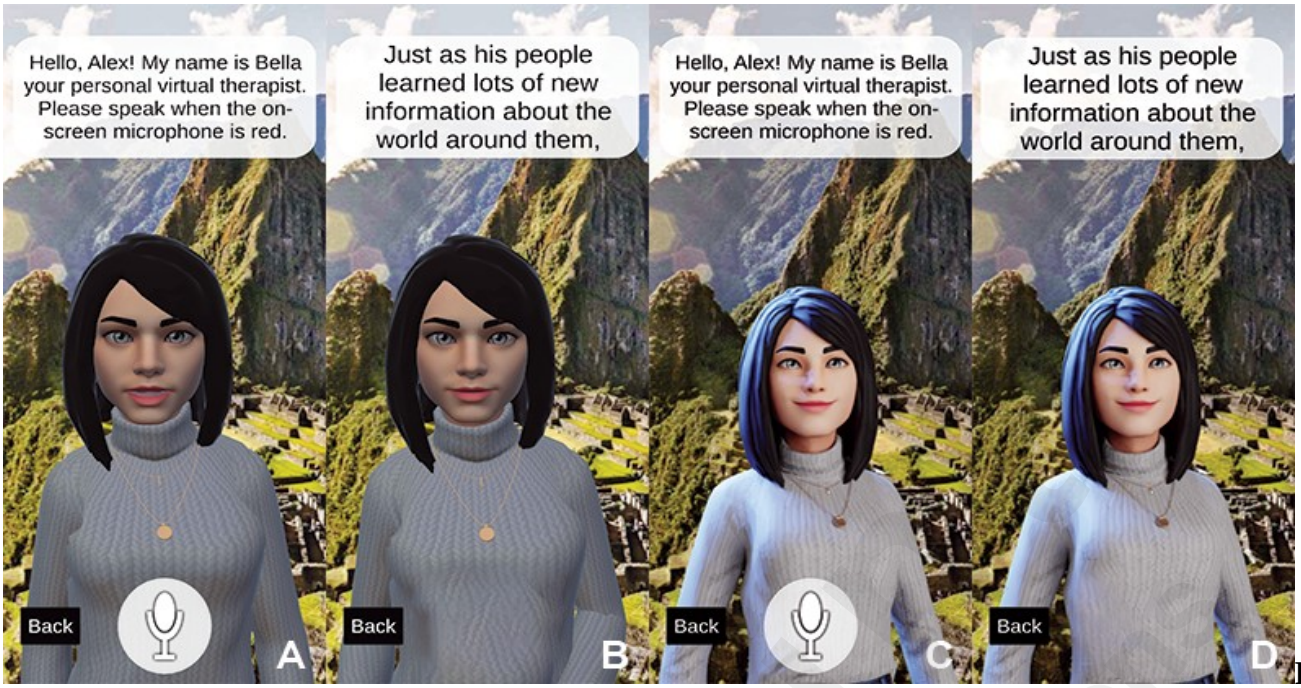

Figure 1. Example of a customized virtual agent in the (A) conversational, animated condition, (B) non-conversational, animated condition, (C) conversational, non-animated condition, and (D) non-conversational, non-animated condition.

Conversational: 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 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.

Procedure

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 mental health questionnaires (PHQ-8, PSS-10, RRS) with attention checks embedded. Afterward, participants downloaded the AirHeart MHealth app via Test Flight (iPhones) or Google Drive (Android), created an account, and followed a tutorial to personalize their virtual agent and complete 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 mental health and user experience questionnaires as well as open-ended questions regarding user experience.

Data Analysis

To investigate H1, 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 >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) ANOVAs were performed for each of the three MAUQ subscales. Using the validated cutoff scores established in prior work [23,24], 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, a thematic analysis was performed similar to the proceed outlined in past work [30]. 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 (*P*s>.05)[1]. Multimedia Appendix 2 shows the 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 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 3).

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 4).[2]

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

[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.

[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.

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

*Significant using alpha = 0.05

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, ηp^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
**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 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 (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	2.70	.10	.013

Depressive Status Interaction Effect			

*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 5).

Table 5

Visualization of Qualitative Data: Suggestions for Virtual Therapeutic Agent 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. 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 [31], 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.

Small pilot studies on virtual agent-based self-monitoring technologies have shown promise in demonstrating the feasibility and preliminary efficacy in reducing depressive symptoms [32-34]. 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 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 [35-36], which may, in turn, improve intervention outcomes [21, 37]. 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.

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, H2 is partially supported. 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 [38,39]. Additionally, the inclusion of such animation design has previously demonstrated a strong connection to higher levels of agent acceptance, trust, credibility, and task appropriateness [19]. 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 agent. Over 80% of participants chose a female agent, often designing it to resemble themselves or someone familiar, like a friend, family member, or therapist. Many participants felt more comfortable discussing mental health with females, citing greater relatability on emotional matters. Social role theory proposes that women are stereotyped as more nurturing, emotionally aware, and empathetic than men [40-42], which may influence therapist choice, both in-person and virtually. However, a meta-analysis found no objective differences in empathy between genders, despite females self-reporting higher empathy [43], highlighting the need for further research on whether these preferences affect therapy quality.

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' tendency to design a coach resembling someone they knew aligns with research showing familiarity provides

comfort, especially when feeling vulnerable [44]. This study extends previous work, indicating that users identify with their virtual coach by customizing it to reflect familiar, empathetic figures.

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 [45,46], and similar results were found using a TTS agent for CBT-based emotional regulation, where participants also noted the robotic speech [47]. Future studies should explore higher quality TTS or pre-recorded human voices to enhance user interactions with the virtual therapeutic 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 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. 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. Pew Research Center. Mobile Fact Sheet. Pew Research Center. January 31. 2024. Accessed April 25, 2024. https://www.pewresearch.org/internet/fact-sheet/mobile/
4. 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
5. 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.
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. 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
14. 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
15. 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
16. 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

17. 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
18. 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
19. 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
20. 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
21. 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
22. 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
23. 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
24. 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

25. Cohen S, Kamarck T, Mermelstein R. Perceived stress scale. *Measuring stress: A guide for health and social scientists*. 1994;10(2):1-2.
26. 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
27. 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
28. 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
29. Test Flight, App Store. Date unknown. https://apps.apple.com/us/app/testflight/id899247664
30. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa
31. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Fifth Edition. American Psychiatric Association; 2013. doi:10.1176/appi.books.9780890425596
32. 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
33. 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
34. 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

35. 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
36. 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
37. 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
38. Salles A, Evers K, Farisco M. Anthropomorphism in AI. *AJOB Neuroscience*. 2020;11(2):88-95. doi:10.1080/21507740.2020.1740350
39. 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
40. 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
41. 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.
42. Ellemers N. Gender Stereotypes. *Annu Rev Psychol*. 2018;69(1):275-298. doi:10.1146/annurev-psych-122216-011719
43. 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

44. 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
45. 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
46. 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
47. 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

Detailed Description of the Virtual Agent and Cognitive Behavioral Therapy Modules.
URL: http://asset.jmir.pub/assets/f7b93c3e723eca733e1388a135c10529.docx

Mixed ANOVA Results for Change in Depressive Symptoms.
URL: http://asset.jmir.pub/assets/1afafa039cb48d77f70da1805b928a05.docx

Mixed ANOVA Results for Change in Stress.
URL: http://asset.jmir.pub/assets/adb096dc76ac763ef63892c83d369690.docx

Mixed ANOVA Results for Change in Rumination Symptoms.
URL: http://asset.jmir.pub/assets/47308fe77b1e624285dce3978b2e9cd5.docx

TOC/Feature image for homepages

Example of the virtual agent in the mental health app.