5-2024

An Exploration Of The Efficacy Of Virtual Agents Within Smartphone Mental Health Applications

Stephanie Six
ssix@g.clemson.edu

AN EXPLORATION OF THE EFFICACY OF VIRTUAL AGENTS WITHIN SMARTPHONE MENTAL HEALTH APPLICATIONS

A Dissertation Defense
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Human Factors Psychology

---

by
Stephanie Six
May 2024

Accepted by:
Dr. Kaileigh Byrne, Committee Chair
Dr. Dawn Sarno
Dr. Cindy Pury
Dr. Sab Babu

ABSTRACT

The prevalence of depression in the U.S. has increased over the past decade, leading to an all-time-high during the COVID-19 pandemic (WHO, 2022). With this increase, the number of mental health applications (MHealth apps) on virtual e-stores increases in tandem. Despite this expanding number of MHealth apps, few demonstrate a foundation in empirical research. One design factor that may influence usability and effectiveness is the inclusion of virtual agents. Three studies were conducted to investigate users' preferences for a variety of characteristics associated with virtual agents. In Experiment 1, users completed a single-session, three-stage CBT-based interaction with virtual characters and were asked about preferences for customization versus evolution. Results demonstrated that participants preferred customization, as it involves more active inclusion in the application. In Experiment 2, users completed four CBT-based modules; two modules with an interactive design and two with a passive design. Results demonstrated a preference for the interactive virtual agent along with higher levels of trust, satisfaction, and comfort. The goal of Experiment 3 was to determine the effect of conversation and animation within a CBT-based MHealth app on depressive symptoms and user experience using a longitudinal experimental design. Results demonstrated a significant decrease in symptoms of depression; however, no significant effect of conversation or animation was observed. This collection of experiments provides insight into user preferences regarding virtual characters. While the addition of virtual agents to MHealth applications holds promise, more research and

refinement is necessary to achieve a seamless incorporation into the mental health domain.

DEDICATION

First, this dissertation is dedicated to all those who seek to help others, whether that is through education, encouragement, or lending an ear.

Second, this dissertation is dedicated to those who are brave enough to seek help. Please know you are never alone.

Lastly, this dissertation is dedicated to anyone who was ever told that they couldn’t do something simply because they were not smart enough.

Go out there and prove them wrong.

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge and thank the advisor who has inspired and guided my academic career for nearly seven years, Dr. Kaileigh Byrne. Thank you for taking your time to mentor a lost nineteen-year-old who didn’t really know if this whole psychology thing was going to work out in the end. Thank you for the last-minute phone calls and reading hundreds of pages and dozens of drafts of countless documents over the years. Lastly, thank you for finally convincing the nervous graduate student who questioned whether she was going to make it through each milestone that everything was going to be okay.

I would like to thank my family for listening to countless rants and updates about my research, classes, internship hunting, interviews, and overall graduate school experience. Thank you, mom and dad, for listening to the graduate student voice all of her anxieties and imposter syndrome over the phone at all hours, always to respond with “it’s all going to be okay”. Thank you to 1st Lt. Austin Russell for making time in his busy day to act as a reassuring sounding board as I vocally walked through whatever problem I was currently facing. Thank you to my niece and nephew for providing some much-needed comic relief and the constant reminder “it’s not scary Fifi”.

I would like to thank the “psych and friends” community here at Clemson who have supported me and encouraged me to better myself socially, academically, and personally over the years. Graduate school is a most certainly a large tidal wave, but y’all reminded me that your community will always here there to act as your anchor.

Lastly, I would like to thank the psychology department at Clemson University. While I had no intention of ever joining the department, it must have been fate, as I have enjoyed seven wonderful years as a part of your community. Thank you for teaching me core principles of psychology and taking the time to turn me into the professional I am today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Depression is a debilitating mental disorder that can significantly impact an individual's physical health, relationships, motivation, and work performance (APA, 2022). To address depressive symptoms, mental health smartphone applications (MHealth apps) have emerged to offer assistance and therapeutic techniques to the public. However, these apps face multiple issues, such as poor usability and features which lack the support of evidence-based psychological practices and empirical research. Principles of human factors may help improve the features included in these apps to become more effective and enhance usability. Usability has been defined in numerous ways, but most famously by Jakob Nielson as either the ease of use of a product or the approach towards improving the ease of use of a product (Nielson, 2012). Additionally, it contains five different components: learnability, efficiency, memorability, error tolerance, and user-satisfaction (Nielson, 2012). The present research will examine how different designs of one feature—virtual characters—influence usability, specifically learnability, user-satisfaction and ease of user, user preferences, and a change in depressive symptoms amongst users experiencing recent depressive symptoms. The following sections will describe current research regarding the relationship between depression, cognitive behavioral therapy (CBT), and MHealth apps and expound upon the potential theories and benefits behind including virtual characters into MHealth apps.

Cognitive Behavioral Therapy (CBT) as an Evidence-Based Practice for Addressing Depressive Symptoms

Symptoms of depression include apathy, anhedonia (the lack of interest and passion received from hobbies and activities that were once enjoyable), prolonged sadness experienced almost daily, insomnia or hypersomnia, irritability, fatigue, feelings of worthlessness or guilt, indecisiveness, difficulty concentrating, and thoughts of death or suicide (APA, 2022). Symptoms may manifest through a subtle decline in extracurricular activities or social interaction, sleeping for extended periods of time, difficulty performing basic tasks such as making meals, going shopping, or completing tasks at work, or self-harm (APA, 2022). Depressive symptoms are nondiscriminatory and may be experienced by different genders, ethnicities, varying levels of economic status, ages, and backgrounds (APA, 2022; Lee et al., 2014; Orgilés et al., 2021; Verhagen et al., 2010; Wang et al., 2016).

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 (NIMH, 2022). Young adults and women have been disproportionately affected (World Health Organization, 2022). Despite the prevalence and negative consequences of depressive symptoms, less than half of individuals with a depression diagnosis receive treatment (either counseling or pharmacotherapy) (Pujat et al., 2016). Treatment barriers include lack of access to care (e.g., overbooked or understaffed mental health clinics, long distances to clinics, lack of transportation), lack of time, lack of knowledge about where to go for help, privacy and confidentiality concerns (Alqahtani & Orji, 2020), negative stigma attached towards the need to seek assistance, a lack of confidence in therapeutic options (Conroy, Lin, and Ghaness, 2020), and high therapy costs (Hayes, 2021). Specifically, treatment costs in the U.S. increased 12.9% from $236.6 billion in 2010 to $326.2 billion per year in 2018 (Greenberg et al., 2021). In terms of access to care, 72% of therapists specializing in depression reported an increase in patient demand since 2020. Approximately 41% of counselors disclosed being unable to take on new patients due to current demand, which is an 11% increase from 2019 (Bethune, 2021). With higher levels of demand superseding the current supply and high economic costs of depression, supplemental ways of providing mental health assistance to a wide variety of individuals need to be created.

Cognitive Behavioral Therapy (CBT) as an Evidence-Based Therapeutic Option

One of the most frequently utilized types of therapy for reducing depressive symptoms is cognitive behavioral therapy (CBT). This evidence-based form of therapy has successfully demonstrated improvement in the quality of life for individuals of varying ages and clinical severity levels (Chaves et al., 2017; Gould et al., 2012; Grosse Holtforth et al., 2019; Oud et al., 2019; Tandon et al., 2014). While commonly used to help treat individuals with depression, CBT is also frequently employed for individuals with anxiety, post-traumatic stress disorder (PTSD), or mental rumination, as these disorders are co-morbid with depression and share similar symptoms (Otte, 2022; Querstret et al., 2016, Simon et al., 2019). CBT aimed towards combating negative moods and symptomology related to depression includes four main components: 1) identifying maladaptive thought patterns, 2) challenging and overcoming these patterns,

3) problem-solving skills, and 4) behavioral activation; two additional components are often used as supplemental material: 1) psychoeducation and 2) mindfulness and meditation (Miner et al., 2016). Below, each of the CBT components are described in detail:

Identifying & Challenging Maladaptive Thought Patterns

The main concepts of CBT include identifying maladaptive thinking patterns and beliefs, challenging these ideas, and creating new and healthy behaviors. Identifying negative thoughts and beliefs encourages the individual to examine their automatic thoughts in positive and negative situations (Cully & Teten, 2008). After identifying these thoughts and beliefs, challenging these ideas includes learning about cognitive distortions. Cognitive distortions are an exaggeration of negative patterns or beliefs (Ackerman, 2021). A total of fifteen cognitive distortions have been identified, such as filtering (ignoring the positive and focusing on the negative), overgeneralization (using one sole incidence to make an overarching, general conclusion), jumping to conclusions (tendency to be sure of an outcome without proper evidence), and catastrophizing (expecting the worst to happen) (Ackerman, 2021). After learning the various types of cognitive distortions, users are instructed to identify the cognitive distortions in their automatic thoughts. Once these are identified and challenged, new cognitive thought processes and behaviors can be created.

Problem-Solving Skills

Problem-solving is an important element of CBT that includes problem orientation and problem-solving style (Bell & D’Zurilla, 2009). Problem orientation is the metacognitive process of identifying problems in an individual's life and a realistic evaluation of an individual's problem-solving skills (Bell & D'Zurilla, 2009). Problem-solving style involves both cognitive and behavioral elements; for cognition, problem-solving style encourages introducing positive problem-solving. This involves changing one's mental mindset from viewing problems or stressors as negative barriers to opportunities for growth and believing that one can overcome these challenges. Additionally, users generate a variety of solutions, decide which solution best fits their current scenario, and take the necessary steps to enact their plan (Bell & D'Zurilla, 2009).

Behavioral Activation

Behavioral Activation involves active participation in activities that provide the individual with positive feelings (Martell, Dimidjian, Hermann-Dunn, 2021; Sturmey, 2009), such as planning activities and setting goals (Martell, Dimidjian, Hermann-Dunn, 2021; Sturmey, 2009). The objective of behavioral activation is to re-introduce positive emotion-inducing activities. Oftentimes, depressive individuals may self-isolate or cease active participation in their hobbies if it no longer provides enjoyment. Encouraging individuals to maintain an active lifestyle and slowly increasing the frequency of these activities helps to reduce depressive symptoms over time (Martell, Dimidjian, Hermann-Dunn, 2021; Sturmey, 2009).

Supplemental Material: Psychoeducation & Mindfulness and Meditation

Psychoeducation for depression includes information regarding depressive symptomology, Beck's cognitive triad (Beckham et al., 1986), the depressive spiral, and ways to address these symptoms (Cully & Teten, 2008; Miner et al., 2016). This form of learning can appear in the form of modules, activities, and explanations (Donker et al., 2009). Psychoeducation works to combat one of the barriers to seeking treatment: disbelief that it will be effective. Through psychoeducation, the individual becomes more confident in their ability to return to an enjoyable quality of life (Tursi et al., 2013).

Mindfulness and meditation introduce an individual to elements of relaxation, acceptance, and detachment. This can involve breathing exercises, stretching, walking, or becoming more aware of the positive moments during the day (Kladnitski et al., 2020). By prioritizing time for meditation, individuals can practice a variety of approaches, such as the RAIN technique. This technique involves Recognizing and Accepting negative emotions, Investigating the resulting thoughts and feelings, and non-identifying or separating oneself from the negative thoughts and emotions to gain acceptance (Brach, 2016).

This group of studies will incorporate these main and supplemental components of CBT as the primary system for delivering therapy to the user. In addition, there will be an investigation into whether these components provided through a virtual, computerized platform could help reduce depressive symptoms.

The Use of CBT & Mindfulness in Mental Health Applications (MHealth Apps) as a Means of Reducing Depressive Symptoms

Incorporating components of CBT and mindfulness into MHealth apps may be beneficial in managing or mitigating depressive symptoms. CBT-based MHealth apps may be particularly beneficial as a supplemental tool to assist individuals when they are waiting for appointments, in-between appointments, or to reinforce and practice concepts introduced in counseling sessions. Indeed, with over three-fourths of the American population owning or having regular access to smartphone technology, MHealth apps represent a viable option to improve access to mental health resources (Pew Research Center, 2019). An estimated 20,000 - 30,000 MHealth apps are available for download on various platforms, such as the Google Play Store or Apple App Store (Clay, 2021). These apps advertise an assortment of services directed at improving specific or overarching negative emotions, moods, or symptoms (Cole et al., 2017; Miner et al., 2016; Schueller, Tomasino, & Mohr, 2017; Six et al., 2022).

Some apps, like MoodMission (Marshall, Dunstan, & Bartik, 2020a), provide activities in the form of text-based modules, videos, or stories to help users understand and manage their symptoms. Additionally, some mindfulness-only apps, like Smiling Mind (Marshall et al., 2020a), encourage users to practice 10-minutes of mindfulness and meditation techniques a day to promote better well-being and awareness. MHealth apps, such as What's Up, link users with a therapist or counselor (Eftychiou & El Morr, 2017). Several CBT-based apps, such as MoodMission, MoodKit (Bakker et al., 2018), iCouch (Dahne et al., 2019), Pacifica (Moberg et al., 2019), and SuperBetter (Roepke et al., 2015), have demonstrated effectiveness in reducing depressive symptoms. Despite their effectiveness, MHealth apps' usability is often ignored or viewed as secondary (Neary & Schueller, 2018). Human-centered solutions can help improve MHealth apps to become more user-friendly, enjoyable, and effective.

The Role of Human Factors in MHealth Apps

While MHealth apps may help mitigate barriers to mental health access (Kenny, Dooley, & Fitzgerald, 2016; Marshall, Dunstan, & Bartik, 2020b), there are notable limitations, including privacy concerns, poor usability, and features that lack evidence of effectiveness (Alqahtani & Orji, 2020). Without a user-focused design, MHealth apps could negate the potential benefits of utilizing the app. One main concern with smartphone technology is the lack of privacy, specifically which data is collected and reported. Individuals who struggle with negative symptoms of mental health often report feelings of apprehension due to the possibility of a breach of personal privacy and confidentiality (Koh, Tng, & Hartanto, 2022). These levels of mistrust include concerns that personal data might be supplied to third parties. Additionally, the absence or inclusion of an ambiguous privacy policy can further deter the user (Koh et al., 2022). The lack of protection of the user's information can stem from a dearth in MHealth app regulation as the apps are not bound by law to adhere to healthcare privacy legislation (Torous et al., 2018; Torous, Kshavan, Gutheil, 2014). The inability to choose or even become aware of the information sold to external sources could deter users who want to improve their mental health. Ultimately, the need for visible and verifiable guarantees of privacy is one of many crucial elements which users expect when utilizing MHealth applications.

Another common issue with MHealth apps includes poor usability. Prior research has worked to categorize poor usability into seven main elements: poor aesthetics, navigation issues, high resource utilization, lack of customization, app instability, and performance issues (Oyebode, Alqahtani, & Orji, 2020). For aesthetics, prior research has demonstrated that users perceive poor or gaudy aesthetics as less usable and user-friendly (Chang, Kaasinen, & Kaipainen, 2012; Moran, 2017; Oyebode et al., 2020). Examples of this can include words and letters that are either too big or too small, unhelpful color combinations that reduces accessibility (i.e. red and green), or large amounts of clutter. Aesthetics have been suggested as connecting a product, such as an MHealth app, with the user's emotions, feelings, and moods (David & Glore, 2010; Zhang, 2009). This finding could help inform designers and developers when programming MHealth apps specific for individuals who often experience prolonged periods of negative emotions.

Navigation issues can affect one of Nielson's components in the definition of usability: ease of use (Neilson, 2012; Oyebode et al., 2020). Examples can include trapping the user in a feature by not supplying a return or back button, unclear directions, misleading icons, and missing or broken links (Oyebode et al., 2020). For many MHealth apps, navigation is vital for progression, understanding, and overall use of an app. If the navigation features are difficult to use and require additional mental effort or resources to understand, large groups of people may cease to use the application. When designing an MHealth application, developers should implement natural mapping in terms of linear progression and icon controls to reduce the amount of required mental resources; for example, an arrow pointing to the right could indicate progression, while an arrow pointing to the left could indicate regression (Oyebode et al., 2020).

High resource utilization is a major component to consider when designing MHealth apps, as the intended population may experience higher levels of fatigue and apathy, lower available mental resources, short attention spans, etc. As mentioned previously, requiring high levels of mental resources could impact productivity, efficiency, and willingness to use the application. Prior research has demonstrated that prolonged tasks which require high levels of mental resources and cognitive demand will encourage inattention and lower accuracy and efficiency (Bioulac et al., 2012; Dekkers et al., 2017; Harstad & Levy, 2014; Slobodin et al., 2020). The subdiscipline of human factors has taken a direct interest in finding ways to reduce the need for high levels of mental resources during a variety of tasks to further increase efficiency, satisfaction, and safety for users. Potential solutions may include the addition of speech-to-text options, which allow personalized choice for users and reduces the mental resources required to maintain focus during mundane tasks, such as typing into a virtual journal (Bakken, 2022; Kumaresan et al., 2022). In addition, speech-to-text allows the user to speak their thoughts without typing on a virtual or physical keyboard. Including the option of reading aloud with text-to-speech could help reduce the mental load from reading multiple lines of text. The user can select a preferred method that will be better suited to their personal needs.

Customization is the user's ability to change a product or interface through various options that fit the user's personality, lifestyle, or needs (Cambridge University Press, n.d.; Six et al., 2022). It has been used in a variety of virtual contexts, such as alarms, cell phone ringtones, and cookie preferences on a website or app. Prior research has shown that the ability to customize a variety of features was seen as a crucial and important element to different kinds of users (Stawarz, Cox, & Blandford, 2014; Zhang et al., 2021). Restricting customization options has been shown to negatively affect motivation and adherence (Oyebode et al., 2020). If users must alter their daily patterns or habits, this could lead to frustration and eventual discontinued use of the application.

Lastly, performance issues, such as time lag between selecting an option and the application registering the selection, & app instability, such as glitches or bugs within the software, could impact usability (Oyebode et al., 2020). Frustration can occur if the app takes too long to load, lags during critical portions, such as the therapeutic modules, or freezes and spontaneously closes the app. Software bugs and performance issues should be eliminated during the testing phase, or users should be given the option to report glitches to the developer to instill a sense of control. In addition to overarching topics of privacy and usability, human factors also plays a role in improving the efficiency and overall design of specific MHealth app features, such as the type of therapy used, journaling, mood tracking, privacy policies, and customization.

Relevance of Human Factors to Specific MHealth App Features

Therapeutic Intervention - Brief CBT: As previously stated, individuals experiencing symptoms of depression or other comorbid disorders could experience higher levels of fatigue (APA, 2022). Therefore, shortened versions of therapy combined with technology have been suggested as a means of maintaining attention while not requiring large amounts of the user's time. Most CBT-based MHealth apps implement *brief* CBT (bCBT) as a means of delivering therapeutic interventions in a time-efficient manner – around 4 –16 brief sessions (US Department of Veteran Affairs, 2013). This form of CBT has demonstrated success in reducing symptoms of depression in a wide variety of age ranges and levels of depression (sub-clinical: Richards & Richards, 2012;

Six et al., 2022; clinical: Smith et al., 2015). This form of CBT is utilized in MHealth applications as it is known for condensing the educational opportunities of in-person therapy into fewer sessions (Cully & Teten, 2008). Additionally, the treatment is flexible, which gives the user more autonomy and control over their treatment, thus potentially improving enjoyment and motivation (Rojas et al., 2022).

Journaling. Prior research suggests that journaling reduces symptoms of depression and improves an individual's quality of life in varying clinical levels of depression (Asbury et al., 2018; Krpan et al., 2013; Miller, 2014). This act of self-reflection allows the individual to assess their feeling, emotions, moods, thoughts, and behaviors privately (Maclsaac, Mushquash, & Wekerle, 2022). Users are often encouraged to write down their negative thoughts to help stop rumination and strengthen feelings of acceptance (Mims, 2015). In accordance with CBT, journaling can provide individuals a place to work through their problems, form solutions, plan social events, and identify maladaptive thoughts and beliefs.

Mobile journaling provides the user with a similar feeling of emotional release to physical journaling, but with the capability to take the journal wherever the user goes (Alqahtani & Orji, 2020). The availability of carrying your journal in your pocket for the majority of the day helps to reduce unintentional lapses of memory. Additional features, like speech-to-text, could potentially reduce time commitments and improve accessibility. Normally, individuals talk more quickly when compared to typing speed on a smartphone (Ruan et al., 2018). By implementing speech-to-text, the requirement of needing to view the smartphone screen to input information becomes irrelevant (Luo, 2021). This also allows for the user's attention to be directed elsewhere. Prior research suggests that individuals with disabilities may be at higher risk for developing symptoms of depression (Kang et al., 2015; Noh et al., 2016). Providing this hands-free option could promote use to a larger group of users, such as blind or physically disabled individuals (Li et al., 2018; Luo, 2021). However, the utilization of speech-to-text within an MHealth application could potentially negatively affect the overall experience. Some users could be uncomfortable with speaking their thoughts aloud, especially sensitive information, so they would need to find a private or quiet place, which reduces the MHealth app's claim to be a form of "on the go" therapy. MHealth apps should include the option but not require its utilization in all aspects.

Mood Tracking. Mood tracking is a popular feature in MHealth apps with an inclusion rate of around 66-86% (Bubolz et al., 2020; Caldeira et al., 2017; Qu et al., 2020). This feature allows users to input information about their current and past moods and emotions. CBT-based MHealth apps that included mood tracking demonstrated significant reductions in depressive symptoms over time (Everitt et al., 2021; Firth et al., 2017; Fitzpatrick, Darcy, & Vierhile, 2017; Young et al., 2021). Utilizing this feature can help individuals self-regulate by identifying patterns of negative emotions, such as rumination, low-energy, apathy, and stress, and positive emotions, such as happiness, confidence, and peace. By understanding this pattern, users can be more mindful of the potential causes or catalysts of their negative emotions. Through understanding the patterns and identifying potential sources, users can work to further prevent negative emotions and stressors.

In addition, mood tracking could induce a sense of control and confidence in one's abilities (Caldeira et al., 2017). Another potential inclusion to improve mood tracking could be introducing an artificial intelligence agent to help explain mood patterns (Chu & Shen, 2022; Fitzpatrick, Darcy, & Vierhile, 2017). Current MHealth applications, such as Woebot, include an AI which informs the user the number of times an emotion was selected throughout the week or month (Fitzpatrick et al., 2017). Additionally, programs could explain reoccurring patterns during the week and prepare suggestions for how to improve or maintain their mood. The addition of AI assistants reduces the amount of effort required by the user regarding pattern identification. Additionally, corresponding with the AI about the user's mood could require less time than answering a lengthy questionnaire. Very little research has been conducted regarding the use of AI technology to identify mood patterns within an MHealth mobile application in a longitudinal experiment. Further research should be conducted prior to solidifying claims of MHealth app improvement.

Privacy. While providing a transparent privacy policy has not been shown to reduce symptoms of depression, it does help address the general concern of confidentiality within smartphone apps. Individuals often refuse to seek out mental health assistance due to fear that they will be stigmatized (Alqahtani & Orji, 2020). To help alleviate this privacy anxiety, MHealth applications should allow users access to a comprehensive and clear-cut privacy policy, so they can easily understand how, and which data is recorded to gauge the security measures taken to protect their data (Doherty et al., 2010). This transparency should also allow them the freedom to reject supplementary data collection.

Regarding privacy policies, prior work has shown that on average, only 24-29% of MHealth apps focusing on bipolar depression (Nicholas, Larsen, & Proudfoot, 2015) and suicide prevention (Larsen, Nicholas, & Christensen, 2016) provided a privacy policy (Torous et al., 2018). Additionally, privacy concerns are greater when the user is providing data relating to topics with strong associations of stigma, such as mental health (Kenny et al., 2016; Zhou et al., 2019). The fear of having one's thoughts, such as suicidal ideation, and behaviors, such as self-harm or acts of paranoia, shared or exposed could deter users (Alqahtani & Orji, 2020). Ultimately, ensuring the privacy and confidentiality of the user is vital for any MHealth application. Should users feel a sense of unease or distrust for the MHealth application, this could inhibit and discourage their initiative to seek out assistance.

Customization: In MHealth apps, customization is relevant in the form of aesthetics, different gendered and accented audio voices, the ability to choose the date and time for reminders, designing a virtual character, etc. Introducing customization within an MHealth can increase feelings of autonomy and choice (Zhang et al., 2021). Negative symptoms associated with depression or comorbid disorders and the subsequent recovery and symptom management can vary amongst individuals. Allowing an individual a choice in treatment is critical for symptom improvement: this can allow users to select topics and treatments most closely related to their needs without providing irrelevant topics. Additionally, customization can allow the user to self-select a steady pace that fits with their own schedule (Ntoumanis et al., 2021; Ryan & Deci, 2017). Prior research demonstrates that choice or customization within MHealth apps reports higher levels of engagement and likeability or preference with users (Alqahtani & Orji, 2020). A subcategory of the user design element of customization is the customization of virtual characters. Outside of user preference, connection, and motivation, the relationship between customization and depressive symptoms is a relatively underexplored area of research, specifically the use of customized virtual characters.

The Use of Virtual Characters within MHealth Applications

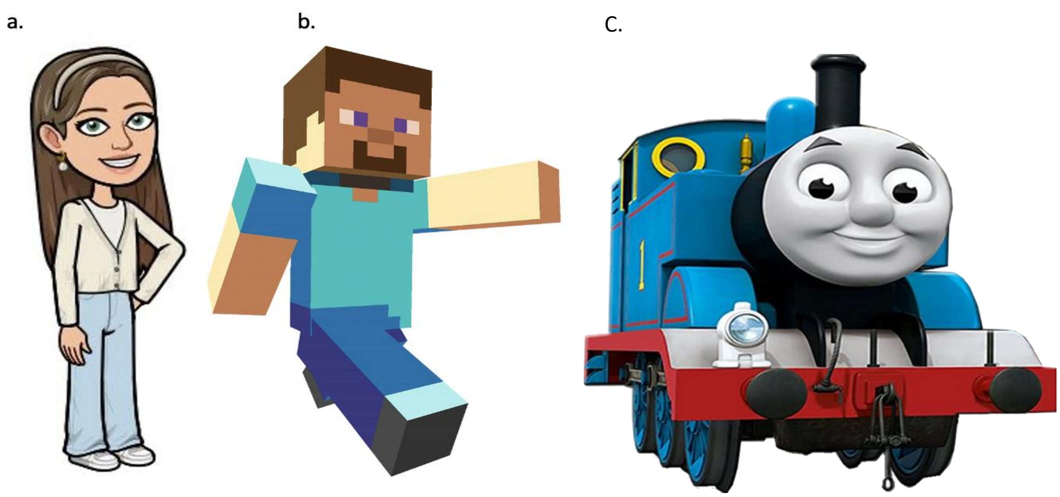
A "virtual character" is an umbrella term meant to represent a variety of different virtual entities, such as embodied conversational agents, avatars, and virtual assistants (Von der Pütton et al., 2010). As opposed to the previous AirHeart experiment which utilized the term "avatar", which represents a human controlled virtual character, this set of studies will be using the terms "virtual character" and "virtual agents", specifically "embodied conversational agents" to describe the various virtual entities (Von der Pütton et al., 2010). Virtual agents refer to a non-controllable virtual entity which includes a form of artificial intelligence (AI) (Balakrishnan and Honavar, 2001; Erickson, 1997; Von der Pütton et al., 2010). Embodied conversational agents (ECAs) are similar as they are non-controllable, computerized entities with a form of AI that exists within a specific virtual environment and mimic realistic human behavior utilizing both verbal conversation and non-verbal body language (Provoost et al., 2017).

A form of virtual characters (virtual assistants) has been commercialized through Amazon's Alexa, the Google Home, or Apple's Siri. They are often employed to help the user with small tasks, like setting a reminder or a timer, playing music, or providing an update on the status of a shipment. Virtual characters have also been used in a story-telling environment, with the virtual character helping to instruct the user or progress the story, such as in a narrating or instructing role. In a healthcare setting, a smartphone application named iHeartU was developed for heart failure patients with an ECA that engaged the user through questions asking about medications, activities, behaviors, and their general progress (Zhang et al., 2019). In addition, virtual nurse agents have been shown to discharge patients and review care plans and various medications for users (Zhou et al., 2014). The use of virtual characters within the context of mental health applications is relatively under-explored; however, the degree of anthropomorphism or level of human-like characteristics of the agent has been shown to affect the user's positive perception of the agent (Salles, Evers, & Farisco, 2020).

Virtual agents can be self-representative (e.g., share the same physical features as the user) or non-representative (e.g., does not share similar features to the user). They may also differ regarding their humanoid qualities, such that the agent may be human-like, such as Snapchat's Bitmoji, less anthropomorphic, such as the block villagers in the *Minecraft* franchise, or non-human-like, such Thomas the Tank Engine (see Figure 1.1). Anthropomorphism can be broken down into four categories: structural (relating to shape or form), gestural (relating to motions of action or human-like behaviors), character (relating to human-like traits or functions), and aware (relating to human-like thought or social qualities) (DiSalvo, Gemperie, Forlizzi, 2005). Virtual agents with higher levels of anthropomorphism may be designed with human-like faces or display various humanoid actions, such as using five fingers to pick up objects, walking on two legs, or using hands or words to communicate (DiSalvo et al., 2005). Virtual agents with higher ratings or anthropomorphism are often viewed as more credible (Nowak & Rauh, 2008) and trustworthy (Touré-Tillery & McGill, 2015) when compared to their less anthropomorphic counterparts. In addition, virtual agents with higher levels of human-like qualities have shown to improve the quality and enjoyment of the experience (Van Pinxteren et al., 2019). These findings suggest that a more humanoid body structure and facial features may provide further identification, as the agent is more relatable to the user. The inclusion of more anthropomorphic virtual agents could help overall satisfaction to MHealth applications.

Figure 1.1

Examples of Varying Levels of Anthropomorphism



Note. A. Bitmoji (Left), B. Minecraft character Steve (middle), & C. Thomas the Tank Engine (Right)

While high levels of anthropomorphism demonstrate potential benefits for improving application technology, there is an upper limit to the level of realism deemed acceptable by users. In particular, the Uncanny Valley refers to the phenomenon whereas the agent's level of anthropomorphism (i.e. human resemblance and believability) increases, as do the feelings of unease towards the agent (Mori, MacDorman, & Kageki, 2012). This effect impacts more than the induction of eerie or unwanted emotions. Prior research has demonstrated that highly realistic agents can negatively impact information processing, judgment, decisions, accuracy (Shin, Kim, & Biocca, 2019), trust towards the virtual agent (Song & Shin, 2022), motivation to perform quick judgments (Clayton & Leshner, 2015), mental resources and efficiency (De Borst & De Gelder, 2015; Yamada, Kawabe, & Ihaya, 2013). While research has shown that individuals tend to verbally respond similarly to varying levels of anthropomorphism in agents, those which are too realistic tend to elicit abnormal non-verbal interactions and behavior (Fink, 2012). Improvements in areas like identification, acceptance, and immersion could in turn impact other areas like experience, trust, time spent within the application, social response, performance, and competence. Ultimately, more research needs to be conducted regarding a variety of individuals' opinions on different agents, specifically within MHealth applications.

Theories and Prior Research Supporting Customizable Virtual Agents in Mental Health Applications

Two different theories can help explain the rationale as to the reason customization of a virtual agent could improve motivation, enjoyment, connection, and overall satisfaction within MHealth apps. Firstly, Self-Determination Theory is an approach to motivation and personality that states that competence, engagement, and autonomy are three interrelated predictors of motivation (Ryan & Deci, 2000). If a user feels competent in their abilities, engaged in the task, and if they are given a choice or opportunity to pursue their goals, the user will feel motivated (Ryan & Deci, 2000). In addition, prior research has discovered that the feeling of competence for a task only motivates an individual when combined with feelings of autonomy (Ryan & Deci, 2000). However, providing easily operated and accessible tasks with strict guidelines could potentially deter motivation over time. To maintain high levels of motivation and enjoyment within MHealth apps, developers must find a balance between a lack of guidelines which can cause confusion, and inflexible pathways that restrict or erase the user's choice. The addition of customization increases the sense of autonomy, which in turn increases motivation and enjoyment (Birk & Mandryk, 2018).

Secondly, Mood Management Theory states that users seek out a specific kind of stimulus to regulate their mood and experience a certain emotion, such as watching a comedy when wanting to feel happy (Li et al., 2011). Individuals may seek out immersion and escapism into a virtual world or MHealth application as a way of detaching from their negative symptoms, such as extreme sadness, apathy, or stress (Li et al., 2011). Customizable agents can act as vessels for immersion into a virtual world, potentially relieving themselves of those negative emotions (Rehm et al., 2016). In the virtual world, this agent can become the user's self-representation, which introduces the potential of providing therapeutic relief, education, self-compassion, and selfempowerment through the agent, thus increasing enjoyment and connection (Yee & Bailenson, 2007).

Prior research has demonstrated that users hold a strong preference for customizable agents over agents supplied by a program due to the increased relatability of the customized agents, which could relate to engagement (Cheek et al., 2014). This preference was demonstrated in three studies investigating customizable characters, motivation, and enjoyment. In the first experiment, researchers were investigating whether creating a customizable character would affect identification and motivation to complete an executive functioning training game (Birk et al., 2016). Ultimately, motivation and identification were significantly higher in the customization intervention compared to the control group (Birk et al., 2016). This suggests that this increase in autonomy through the alteration of the agent in the customization condition improved motivation and enjoyment, thus encouraging players to return to the training game and put more effort into the task. In the second experiment, participants were randomized to a customization condition, where they were instructed to customize their agent or a control condition which was assigned an agent at random, then completed an attention bias modification training or a no-training control activity (Birk & Mandryk, 2019). This experiment discovered that individuals who customized their agent demonstrated higher levels of engagement, resilience to negative moods, and identification with the agent (Birk & Mandryk, 2019).

Customization seems to be favorable amongst users (Cheek et al., 2014; Schwebel & Larimer, 2018); however, these options should remain optional as to not overtax users

(Zhang et al., 2021). Prior work shows that individuals with cognitive or emotional dysregulation, such as those with anxiety, depression, or higher levels of stress, may experience reductions in motivation and energy (APA, 2022). Thus, a multitude of options may deplete their energy levels before reaching the modules or intervention element of the MHealth app (Zhang et al., 2021). Additionally, a lack of a sufficient number of customization options could produce a negative effect. This lack of options could impede the identification and motivation to utilize the application. Further research should investigate what kinds of virtual agent features are preferred by users with varying levels of depressive symptoms and whether this has an impact on trust, connection, and overall satisfaction.

Purpose

Prior research indicates that utilizing bCBT in a MHealth application holds the potential to reduce symptoms of depression over a short period of time (Six et al., 2022), however, one of the main deterrents is poor usability. This set of studies will seek to investigate and improve the current usability of the AirHeart system by utilizing principles of human factors and user experience (UX). Through system improvement, user trust, autonomy, and enjoyment of the system will increase, thus encouraging frequent use. With more re-occurring use, participants will experience more interaction with therapeutic lessons, thus improving their chances to lower their negative systems and improve their quality of life.

The overarching purpose of this dissertation is to identify how specific features of a virtual character within a CBT-based mental health app influence efficacy (symptom reduction) and user experience. Prior research demonstrates success with improving the participant's quality of life and reducing symptoms of depression using a chatbots and conversational agents (Burton et al., 2016; Fitzpatrick et al., 2017; Fulmer et al., 2018; Gaffney et al., 2019; Inkster et al., 2018). These virtual entities potentially provide a sense of social interaction, empathy, understanding, connection, and interaction, which may be missing or reduced in individuals experiencing depressive symptoms. However, to our knowledge, no study has investigated the use of ECAs utilizing CBT for depression without a clinician intervention in a short-term and longitudinal format. Experiment 1 will empirically investigate preference between non-human, low anthropomorphic virtual characters as they evolve or unlock additional customizable elements for users with and without depressive symptoms. While the customization of virtual characters is used in a multitude of virtual environments, the use of a character's evolution to demonstrate progression, learning, and connection in a mental health context is an under-explored area of research. Secondly, Experiment 2 will investigate user preference between active involvement in CBT-based modules through auditory conversation with a virtual human compared to passive learning through CBT-based videos. This experiment will simulate a virtual therapy session with an ECA. Lastly, Experiment 3 will be a longitudinal extension of Experiment 2 utilizing the new version of the AirHeart app (Six et al., 2022). This experiment will investigate the effect of conversation and animation of a virtual therapeutic coach on symptoms of depression, anxiety, stress, and rumination, usability, and participant perception (trust, bonding, and experience with the agent and AirHeart app).

CHAPTER TWO

EXPERIMENT 1: CUSTOMIZATION VS. EVOLUTON

As previously mentioned, customization is a feature commonly found in a variety of virtual applications, however, the use of a customizable virtual character is not as frequently utilized or investigated within the context of MHealth apps. Similarly, growth or evolution has been used in a variety of gamified applications to demonstrate progression, such as Plant Nanny, a smartphone application that encouraged users to drink cups of water throughout the day to help a virtual plant grow and bloom (Rebedew, 2018). However, the use of the evolution of a virtual character within an MHealth application remains relatively uninvestigated. Theoretical support for this idea derives from the Proteus Effect (Yee, Bailenson & Ducheneaut, 2009). The Proteus Effect suggests that users are influenced by their virtual characters, such that their personalities and behaviors tend to mimic their character's persona (Yee, Bailenson & Ducheneaut, 2009). Prior research has demonstrated that this effect prompts changes in the user's level of aggression during negotiation (Yee & Bailenson, 2007; Yee, Bailenson, Ducheneaut, 2009), antisocial behavior (Yoon & Vargas, 2014), body dissatisfaction (Fox, Bailenson, & Tricase, 2013; Sylvia, King, & Morse, 2014), risk-taking behaviors (Hershfield et al., 2011), and motivation (Ratan et al., 2016). By including virtual characters in MHealth applications, users can watch and experience their virtual characters learn about mental health in a positive context, solve problems, and overcome their negative thoughts. Should the character display positive and confident emotions and non-verbal body language as a result of learning new skills and coping techniques, the user may mimic these emotions and behaviors. Visualizing their character's growth throughout their journey could instill feelings of dedication, happiness, and confidence within the user.

The purpose of Experiment 1 was to gain a better understanding of user preference regarding customization and evolution of virtual characters. Within this experiment, participants completed a total of three different CBT text-based modules. As past work indicated a potential preference for individuals with higher levels of depression to want a virtual character which does not resemble themselves (Six et al., 2022), potentially due to feelings of self-hatred (APA, 2022; Mullarkey, Marchetti, & Beevers, 2019), the current experiment utilized animal character. The user-chosen animal virtual character accompanied the participant throughout the modules. The experiment randomized participants to one of two different conditions: customization or evolution. The customization condition offered participants a greater form of autonomy or choice, while evolution provided a more passive play style. It is possible that customization and choice will allow a greater form of connection, likeability, and enjoyment when compared to a passive experience. Ultimately, this experiment inquired about the user's preference for a variety of smartphone application attributes, such as customization, evolution, and virtual character's appearance.

Experimental Aims

Aim 1: The primary aim of this experiment was to investigate the preference of individuals with varying levels of depressive symptoms regarding customization and evolution or growth of a virtual character within the context of an MHealth application.

Aim 2: The secondary aim of this experiment was to investigate user preference regarding different options for virtual character, such as appearance, type of character, connection, and motivation, as well as the user's gaming tendencies and overall opinion of the usability of the system.

Hypotheses

Aim 1 Hypothesis:

H1: Customization will score higher in preference when compared to evolution for users with higher levels of depressive symptoms when compared to users with lower levels of depressive symptoms.

Aim 2 Hypothesis:

H2: Users with higher levels of depressive symptoms will prefer a virtual character that does not resemble themselves when compared to individuals with lower levels of depression.

H3: Individuals in the customization condition will show higher levels of connection, likeability towards their character, and app enjoyment than the participants in the evolution condition.

H4: Individuals with higher levels of depressive symptoms will exhibit higher levels of connection with their character.

Exploratory Hypotheses

H5: Investigate user preference relating to the statement, "I feel more connected when I know the character's background and story".

H6: Investigate user preference in regarding to five different kinds of virtual characters.

H7: Investigate whether a difference between the low-level depressive and high-level depressive groups occurred for usability of the AirHeart system.

METHOD

Participants

Prior to starting the experiment, a G*power analysis was conducted to determine the number of participants needed to maintain an 80% power level to detect an effect at the level of significance (p = .05). A total of 120 participants were needed to maintain the desired power level, however, a total of 168 participants were recruited to account for potential data loss. Participants were incentivized to participate in this experiment with compensation in the form of course credit or extra credit for a class. Half of the participants were allocated to the customization condition, and the remaining half were allocated to the evolution condition.

Exclusion Criteria: Participants were excluded from participating in the experiment for three reasons: 1) the individual was under the age of 18 and classified as a minor, 2) the individual was not fluent in English, and 3) the individual was not enrolled as a student at Clemson University. During data cleaning, data were excluded for three reasons: 1) the participant did not complete the experiment, 2) the participant finished the experiment in less than 20 minutes or took longer than 50 minutes, and 3) an individual failed one or more attention checks.

Measures

This experiment consisted of two main surveys: a pre-experimental survey and a post-experimental survey. All surveys were created using the online survey maker *Qualtrics*.

Pre-Experimental Survey: This survey consisted of two demographic questions (age and gender), the 20-item Center for Epidemiological Studies Depression Scale (CES-D), and an 18-item character preference questionnaire. The CES-D questionnaire measures levels of depression over the previous week. Participants are asked a variety of questions regarding their experience with common symptoms of depression, such as fatigue, overwhelming sadness, and loneliness. With scores ranging from 0-60, individuals are asked to rate their response to the questions on a 4-point Likert scale from 0 (rarely or none of the time) to 3 (most or almost all the time. Scores greater than 16 are considered indications of higher levels of depression (Radloff, 1977). The CES-D has maintained high levels of internal validity in a variety of populations, such as the elderly (α = .83) (O'Halloran, Kenny, & King-Kallimanis, 2014), middle aged adults (α = .90) (Cosco et al., 2017), and adolescents (χ^2 = 74.74) (Cheng, Chan, & Fung, 2006).

The 18-item virtual character preference questionnaire includes questions regarding the user's opinion on their favorite type of character, what attributes make a character more likely to be selected above others, appearance, customization, evolution, motivations, and rewards. For fourteen of the questions, users indicated their response on a 5-point Likert scale ranging from 0 (Strongly Disagree) to 5 (Strongly Agree). For the remaining four questions, users either typed their response into a word box, ranked the desirability of certain virtual characters on a scale from 1 (most desirable) to 5 (least desirable), or selected one or multi-answers in multiple-choice format.

Post-Experimental Survey: This survey consisted of a follow-up 11-item virtual character preference questionnaire and the System Usability Scale (SUS). The virtual character preference questionnaire inquired about the user's connection and identification with their animal character, the likeability of the character, the character's positive effects on the app, and the user's gaming habits. Users responded to the questions on a 5-point Likert scale ranging from 0 (Strongly Disagree) to 5 (Strongly Agree). Additionally, users were asked about ways to improve their animal character with the MHealth app. The SUS was used to determine the user's satisfaction with the simulated MHealth app. This included questions regarding the likeability, complexity, functions, consistency, fluidity, and directions provided within the app. Users responded to the questions on a 5-point Likert scale ranging from 0 (Strongly Disagree) to 5 (Strongly Agree). Scores were then tallied ranging from 0-100, with scores below 68 deemed below average with poor usability (Bangor, Kortum, & Miller, 2008; Lewis, 2018). The SUS is one of the most used usability scales for a wide variety of products due to its short nature and easily understood questions; this scale maintains high internal validity with a Cronbach's alpha of around 0.9-0.92 (Lewis, Brown, & Mayes, 2015; Lewis & Sauro, 2009).

Materials

Participants completed this experiment in lab on desktop computers utilizing a standard monitor, keyboard, and mouse.

AirHeart V2: AirHeart V2 is a simplified version of the original CBT-based AirHeart application (Six et al., 2022). The application included 2D cartoon-style aesthetics. This application took users on a journey to three of the wonders of the modern world (Petra, the Roman Colosseum, and the Taj Mahal) before returning them to Clemson University. Each stop along the way taught users about a principle of CBT (identifying and challenging maladaptive thoughts and behaviors and problem-solving techniques) and encouraged them to practice these new therapeutic tricks through activities. At the final stop along the journey, the user received a summary of their new skills and techniques and was encouraged to integrate them into their everyday lives.

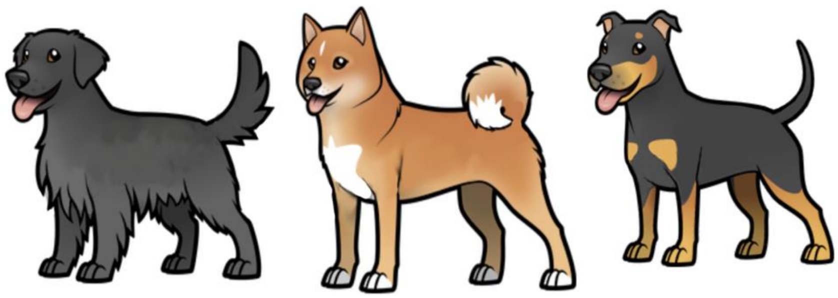
While the original AirHeart application utilized virtual humans, AirHeart V2 integrated animal characters in the form of dogs and cats, see Figures 2.1 and 2.2. A total of three different cat breeds (a white Himalayan, a black Shorthair, and a black and orange Maine Coon) and three different dog breeds (a black Lab, a tan Shiba Inu, and a black and tan Rottweiler) selections were available to the user. A variety of options were included as a means of increasing autonomy and potential connection with the virtual character. Similar to the original AirHeart application (Six et al., 2022), AirHeart V2 included the virtual character in the CBT-based modules to provide companionship along their journey. The two different conditions impacted how the virtual character's appearance would change as the user progressed through their journey: customization and evolution.

Figure 2.1

Three Adult Virtual Cat Characters Used within AirHeart V2

Figure 2.2

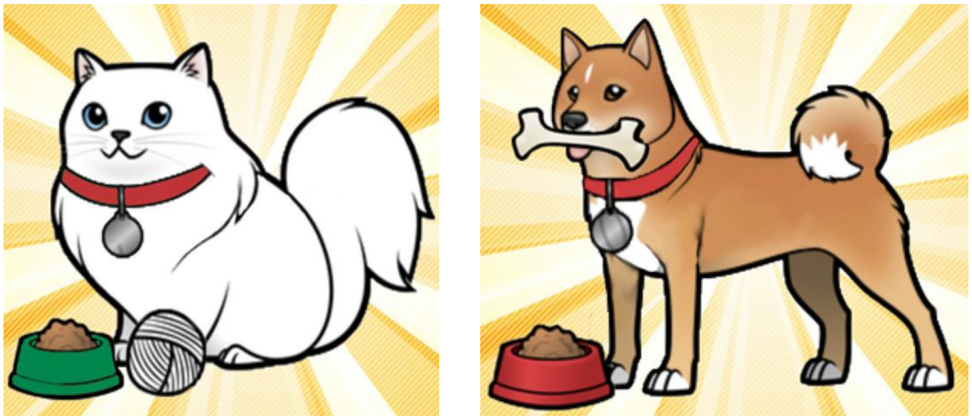
Three Adult Virtual Dog Characters Used within AirHeart V2



Experimental Condition 1: Customization. Within the customization condition, the users would collect new accessories for their virtual characters. Users would earn one new accessory after completing a module, for a total of three accessories. They would be given the choice of a red or green collar, a red or green bowl, and a toy for their virtual character (i.e. ball of yarn or toy mouse for cat characters and bone or tennis ball for dog characters). After selection, the user's chosen accessory could be seen on or around their virtual character in the following module. See Figure 2.3 for examples.

Figure 2.3

White Himalayan Cat (Left) & Shibu Inu (Right) Fully Customized



Experimental Condition 2: Evolution. Within the evolution condition, the user's virtual character would begin their journey as a young puppy or kitten. Throughout the modules, the user would watch their virtual character grow in size and maturity. Once the user started the final module, their character would be fully grown. See Figures 2.4 and 2.5 for an example of evolution and Figure 2.6 for an example of the modules.

Figure 2.4

Black Shorthair Cat Evolution

Figure 2.5

Tan Shibu Inu Evolution

Figure 2.6

Example of a CBT Module within AirHeart V2

Experimental Design

To investigate H1 & H2, an experimental between-subjects design was utilized for this experiment. Participants were randomized to either the evolution or customization condition.

Procedure

Prior to commencing data collection, approval for this experiment was given by Clemson University's Institutional Review Board.

Upon arrival to the lab, participants were instructed to sit down at a desk with a computer and read, review, and sign a consent form provided by the researcher. Once verbal and written consent was received, the participant was randomized to either the customization or evolution condition via a random number generator (i.e., even numbers = customization & odd numbers = evolution) and asked whether they would prefer a dog or cat character. The user was then directed to the computer to complete the pre-experimental survey consisting of demographics, the CES-D, and the virtual character preference questionnaire. A total of five attention check questions were embedded amongst the three questionnaires to ensure sustained attention. After the user completed the pre-experimental survey, they selected their dog or cat breed to start the modules. As the user completed the modules, the customization condition was given the option of adding an accessory to their virtual character, while the evolution condition saw a screen that said "Look how much your character has grown" with the next phase of the evolution of their character present. Upon completion of the four CBT-based modules, the user started the post-experimental survey consisting of the follow-up 11-item virtual character preference questionnaire and the System Usability Scale (SUS). Upon completion, they were debriefed about the purpose of the experiment, given confirmation about their partial course credit or extra credit for participating, and exited the lab.

Data Analysis

To investigate H1, an independent samples t-test was utilized to discover whether a preference between evolution and customization emerged between the depressive and non-depressive samples.

To investigate H2, an independent samples t-test was conducted to explore whether a difference between depressive and non-depressive samples emerged regarding a preference for their virtual character to resemble themselves.

To investigate H3, three independent samples t-tests were used to inquire as to whether the condition (customization or evolution) impacted the user's level of connection, likeability towards their character, and overall app enjoyment.

To investigate H4, an independent samples t-test was conducted to discover the relationship between perceptions of connection with the virtual character and levels of depressive symptoms.

For the exploratory hypothesis H5, an independent samples t-test was conducted to explore the connection between depressive symptoms and levels of connection due to provided background information.

The exploratory hypothesis H6 utilized an independent samples t-test and an observation of the means for both groups to determine whether any difference between levels of depressive symptoms and ranked order of preference for types of virtual characters existed.

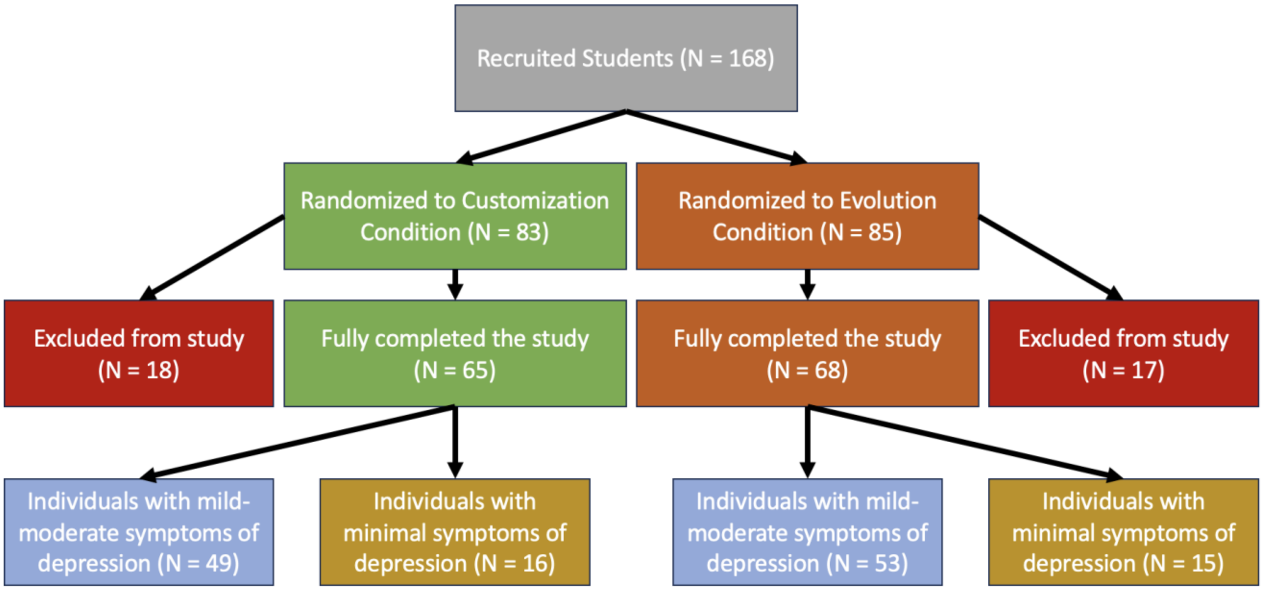
Lastly, exploratory hypothesis, H7, utilized an independent sample's t-test to determine whether any difference in usability was reported between the depressive and non-depressive groups.

RESULTS

A total of 168 undergraduate students were recruited through the SONA system at Clemson University (Evolution N = 85; Customization N = 85). Thirty-three students were excluded due to meeting at least one of the three exclusionary criteria (Evolution N = 18; Customization N = 15). A total of 133 undergraduate students fully completed the experiment (Evolution N = 68; Customization N = 65; M_{age} = 18.88, SD_{age} = 1.18; Females = 94 (70.68%), Male = 39 (29.32%)). The majority of individuals in both the evolution condition (N_{dep} = 49, $N_{non\text{-}dep}$ = 16) and the customization condition (N_{dep} = 53, $N_{non\text{-}dep}$ = 15) exhibited at least mild to moderate symptoms of depression according to the CES-D. See Figure 2.7 for a visualization of this information.

Figure 2.7

Customization and Evolution Conditions Sample Sizes



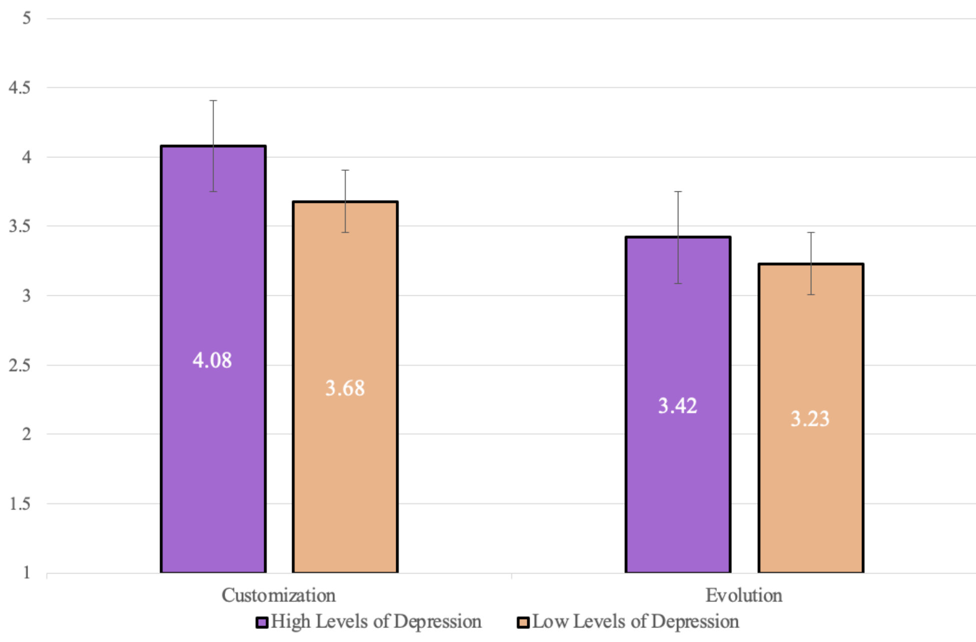
Primary Hypotheses

Results of the independent samples t-test to test H1 showed a significant difference between groups for customization preferences. For the first statement ('I would like the ability to customize my avatar'), (t (131) = -2.31, p = .023), individuals with higher levels of depressive symptoms more strongly agreed with this statement (M = 4.08, SD = 0.79) when compared to non-depressive individuals (M = 3.68, SD = 1.01). No significant difference was found between differing levels of depressive symptoms for evolution for the statement 'I want to see my character grow' (t (131) = -.957, p = .340; individuals with higher levels of depressive symptoms (M = 3.42, SD = 0.98); non-depressive individuals

(*M* = 3.23, *SD* = 1.06)). Thus, H1 was supported. See Figure 2.8 for a visualization of these results.

Figure 2.8

Virtual Character Preferences of Users with Low and High Levels of Depressive Symptoms



The results of the independent samples t-test to test H2 were non-significant ($t(131) = -.30$, $p = .765$). Thus, no difference in reported desire for self-representative virtual characters between those with and without high levels of depressive symptoms was found; H2 was not supported.

For H3, three independent samples t-tests were conducted, one for each dependent variable: 1) level of connection with the virtual character, 2) level of likeability of the virtual character, and 3) virtual character made the application more enjoyable. Findings from all three t-tests were non-significant (1. $t(131) = .1.309$, $p = .193$; 2. $T(131) = .800$, $p = .425$; 3. $T(131) = .539$, $p = .591$). Thus, H3 was not supported.

The results of the independent samples t-test to test H4 were non-significant ($t(131) = -.383$, $p = .702$). Thus, no difference in reported connection with the virtual character between users with lower and high levels of depressive symptoms; H4 was not supported.

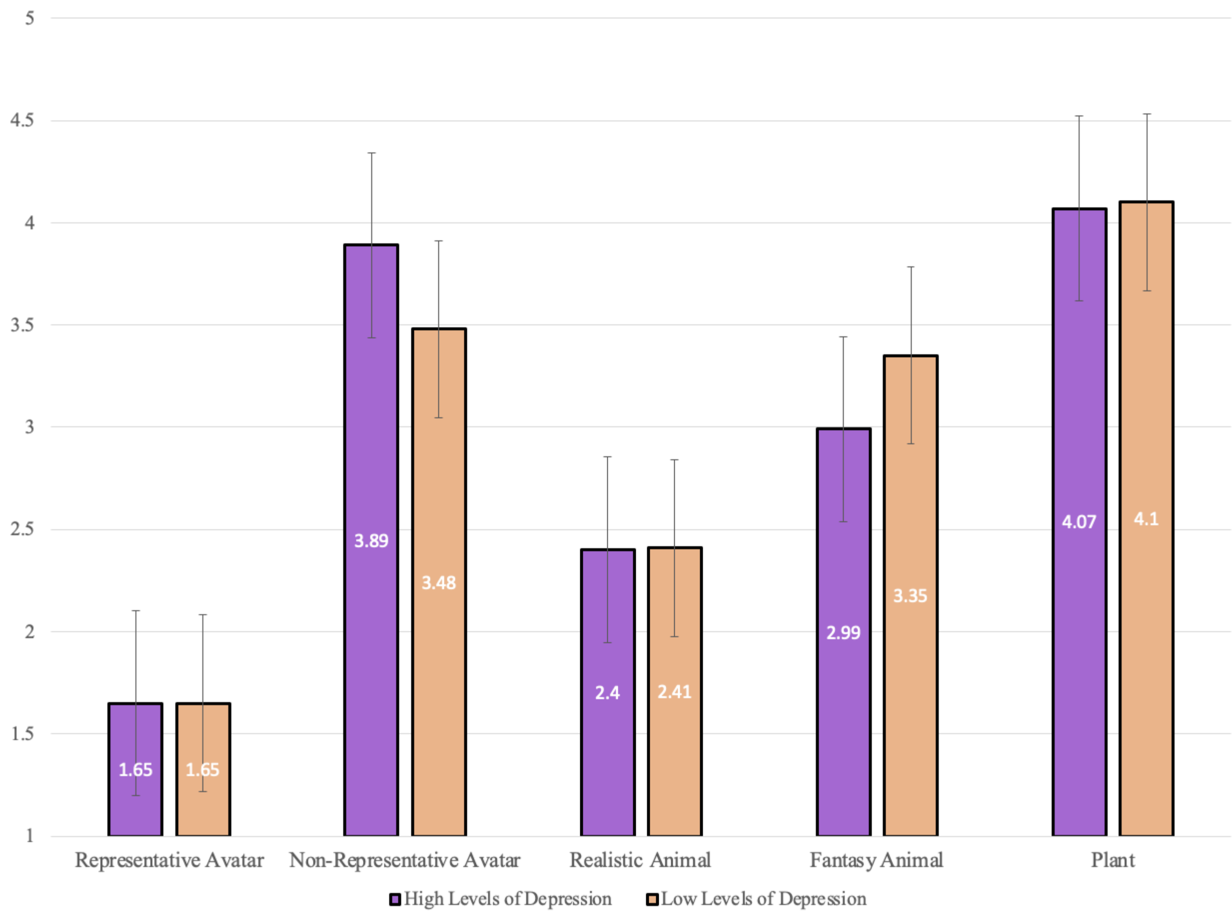
Exploratory Hypotheses

For the exploratory hypothesis H5, users responded with how much they agreed with the following phrase: "I feel more connected when I know the character's background and story". While there was no significant difference between the two groups ($t(131) = -1.134$, $p = .261$), both individuals with higher levels of depressive symptoms ($M = 4.29$, $SD = .63$) and individuals with lower levels of depressive symptoms ($M = 4.13$, $SD = .80$) reported strong agreement to the statement.

For the exploratory hypothesis H6, additional data was collected regarding virtual character type preference. Users were instructed to rank five types of virtual characters from most desirable (1) to least desirable (5). The ranked order remained the same across varying levels of depressive symptoms: 1) a human character that looks like the user ($M_D = 1.647$; $M_{ND} = 1.645$), 2) an animal ($M_D = 2.40$; $M_{ND} = 2.41$), 3) a fantasy animal ($M_D = 2.99$; $M_{ND} = 3.35$), 4) a human character that does not look like the user ($M_D = 3.89$; $M_{ND} = 3.48$), and 5) a plant ($M_D = 4.07$; $M_{ND} = 4.10$). No significant differences were discovered between the two groups (ps > .05), see Figure 2.9.

Figure 2.9

Character Preference amongst Individuals with Low and High Levels of Depressive Symptoms



Lastly, for exploratory H7, no significant difference was reported between the low-level depressive group (*M* = 74.10; *SD* = 11.71) and high-level depressive group (*M* = 77.47; *SD* = 1.29) regarding SUS scores ($t(133) = 1.39$, $p = .17$).

DISCUSSION

This experiment investigated the relationship between user preference regarding various attributes of virtual characters and differing levels of depressive symptoms. Individuals with depressive symptoms often experience higher levels of fatigue, apathy, and anhedonia, which can affect their inclination to utilize MHealth applications (APA, 2022). The current experiment demonstrated that users with higher levels of depressive symptoms more avidly agree that they would prefer customization to evolution in MHealth apps when compared to users with lower levels of depressive symptoms. This preference for customization supports prior research indicating a strong user predilection toward customization (Birk, Mandryk, & Atkins, 2016; Cheek et al., 2014; Hudson et al., 2022; Six et al., 2022). The ability to alter a device, application, or feature allows for an increase in autonomy and perceived usability, both of which are features suggested to improve user satisfaction. Additionally, the inclusion of customization could allow for a strong level of immersion for the user into an application. This improved sense of captivation could invite stronger sensations of escapism, motivation, enjoyment, and simply desire to utilize the application.

This experiment did not support previous suggestions that individuals with higher levels of depression want a character that does not resemble themselves (Six et al., 2022). Additionally, when asked to rank a selection of virtual characters, individuals with all ranges of depressive symptoms stated that their primary preference was a character who resembled themselves. Should an application include customization features for virtual characters, a multitude of options should be included. An expectation exists that certain attributes reminiscent of the user, such as hair color, hairstyles, clothes, different facial features, etc. would be included in the application. Should applications not include enough options or features to allow a user to create a character to resemble themselves, this could induce negative feelings of poor usability and exclusion. Prior research supports this notion with higher levels of satisfaction being attributed to applications with a multitude of customization options (Alqahtani & Orji, 2020; Oyebode et al., 2020; Stawarz, Cox, & Blandford, 2014; Zhang et al., 2021x). App developers and designers should concentrate resources on supplying an arsenal of customization options.

Prior research has suggested that customization of virtual characters helps to improve the connection between the user and the application (Birk & Mandryk, 2018); however, utilizing evolution or progression to improve connection is relatively under-investigated. The current experiment discovered that no difference in connection existed between customization and evolution, however, individuals overall seem to prefer customization to evolution. These findings present two main interpretations: preference for autonomy and lack of necessary time to form the connection. Regarding a preference for autonomy, prior research suggests that users enjoy autonomy and the option of choice (Birk et al., 2016; Ryan & Deci, 2000). Presenting the user with the option of choice invites them to actively engage in the application as opposed to the system performing a function unprompted for them. Future research should investigate whether active engagement in MHealth apps improves the connection in customization conditions. An alternative explanation for the lack of connection between the individual and the virtual character could be the limited 45-minute time frame of the experiment. Future studies should investigate the effect of customization on connection in a longitudinal experiment.

In addition to customization, the current experiment investigated the difference between levels of connection between individuals with low and high levels of depressive symptoms. Prior research suggests that a depressive diagnosis is highly correlated with escapism and connection with characters in virtual environments (Fernandes et al., 2020; Li et al., 2011). This experiment did not corroborate these findings with both depressive and non-depressive individuals somewhat agreeing to a strong connection with their virtual character (M_{dep} = 3.43, *SD* = .88; $M_{non\text{-}dep}$ = 3.37, *SD* = .89). Similar to the preceding paragraph, the lack of time the user interacted with the virtual character could have potentially resulted in lower levels of connection than predicted.

Regarding the experimental hypotheses, results demonstrated that individuals with higher levels of depressive symptoms demonstrated a larger preference for the idea that they would feel more connected with a character or character if the character's background was presented. This could indicate a stronger preference for immersion into an application facilitated by knowledge of the character's history or other informal characteristics, such as family members or current aspirations. Prior research suggests that social connection is of vital importance to a healthy mindset and overall well-being; learning more about the virtual character could provide a form of social connection (Saeri et al., 2018; Seppala, Rossomando, & Doty, 2013). Further research is needed to help ascertain the rationale for this difference between individuals with higher and lower levels of depressive symptoms in this regard. Additionally, research should include the implementation of additional social information to virtual characters within an MHealth app setting.

Both users with low and higher levels of depressive symptoms reported a similar preference regarding the type of virtual character. Users reported wanting a virtual human character that resembles themselves followed by an animal or a fantasy animal. As previously mentioned, users could desire a character which resembles themselves as a stronger means of immersing their personality into the virtual world. Additionally, users tend to create identities for their virtual characters which match representations of themselves or their perceptions of themselves and their world (Kang & Yang, 2006; O'Brien & Murnane, 2009). Prior research supports this idea as users tend to report stronger positive emotions and connectedness towards characters which resemble themselves (Cheek et al., 2014; Suh, Kim, & Suh, 2011). These results also suggest that users would prefer an animal- based virtual character instead of a virtual human which does not resemble themselves. This could be indicative of the user's connection with a current or past pet.

Lastly, the SUS scores for both the low-level depressive and high-level depressive groups demonstrated that this version of the AirHeart system was in the 70-80th percentile, suggesting good usability.

Limitations

Two limitations permeate the methods and data analysis. Primarily, the findings between individuals with low and high levels of depressive symptoms should be interpreted lightly. The sample sizes strongly favored individuals with higher levels of depression, thus potentially resulting in type 1 and type 2 errors. Future studies should strengthen the claims made within this paper by implementing a 50/50 ratio of low and high levels of depressive symptoms. Additionally, one of the questionnaires did not originally appear in the dog customization and evolution conditions, thus introducing data loss into the experiment. While proper statistical tests were utilized to reduce the impact of the loss, it remains a possibility that this could have impacted and skewed the results. Additional research should be conducted with this questionnaire utilizing an appropriate sample size to achieve the necessary effect size.

Overall, this introductory experiment investigated the difference in preference for numerous virtual character characteristics and customization versus evolution. While the experiment faced limitations, it also produced numerous questions to be answered in a multitude of follow-up experiments. As individuals with higher levels of depressive symptoms demonstrated a stronger preference for customization, further exploration into additional elements of therapy modules within MHealth applications, specifically active engagement from a virtual character compared to a more passive approach, will be explored.

CHAPTER 3

EXPERIMENT 2: PASSIVE VS. ACTIVE INVOLVEMENT

Prior research has indicated a difference in user preference between passive and active involvement within therapy (Uzoka, 1983). These results demonstrated that the active involvement of the therapist encouraged higher levels of diction and self-disclosure from patients. Patients also reported higher levels of adherence (e.g., returning to therapy sessions) (Uzoka, 1983). In parallel, patients who actively engage in therapy sessions often experience stronger reductions in negative symptoms when compared to individuals who passively absorb psycho-information and therapeutic techniques provided by a clinician or therapist (Kertes et al., 2011). With a steady increase in the demand for mental health assistance, MHealth apps have begun to implement virtual agents in therapist roles. As such, these virtual agents have shown to be effective as conversationalists in a mental health counselor role (Burton et al., 2016; Fitzpatrick et al., 2017; Fulmer et al., 2018; Gaffney et al., 2019). This could be due to high levels of social presence emanating from the agent and experienced by the users (Lee, Kavya, & Lasser, 2021). Social presence theory builds on ideas of interpersonal communication or interactions within human-computer interactions (Short, Williams, & Christie, 1976). In addition to providing various types of therapy, such as CBT, chatbots, such as Woebot (Fitzpatrick et al., 2017), Wysa (Inkster, Sarda, & Subramanian, 2018), or Tess (Fulmer et al., 2018), virtual agents can provide social interaction which mimics connection and understanding experienced through interactions with other humans (D’Alfonso, 2020).

Additionally, a systematic review regarding the role and effectiveness of conversational agents for therapeutic use, such as assistance with symptoms of depression, anxiety, loneliness, and distress, supported all findings mentioned above (Gaffney, Mansell, & Tai, 2019). Specifically, significant improvements were reported by users with negative symptoms of depression (Burton et al., 2016; Fitzpatrick et al., 2017; Fulmer et al., 2018; Gaffney et al., 2019; Inkster et al., 2018; Pinto et al., 2016; Pinto et al., 2013). Qualitative research revealed that individuals valued the virtual agent's conversational approach to therapy, high levels of empathy, human-like personality, and interactivity (Gaffney et al., 2019). In contrast, if the agents became repetitive, confused, or showed a lack of understanding regarding the user's experiences or emotions, the level of connection between the user and the agent decreased and lead to feelings of frustration (Gaffney et al., 2019). Additionally, prior research reported that individuals felt more trust towards an agent if the voice matched the user's gender (Lee et al., 2021). Lastly, users were more likely to recommend active agents to their friends and family over the agents who had lower levels of engagement (Bird et al., 2018; Gaffney et al., 2019; Moher et al., 2009). However, many of the studies mentioned in the systematic review involved the assessment of chatbots, which utilized messaging to distribute conversational therapy. Future research should investigate if conversational therapy paired with realistic human-like animations would achieve similar results utilizing CBT.

This experiment will also include a new CBT-based therapeutic module to help reduce symptoms of depression: episodic future thinking. Episodic future thinking (EFT) involves combining prospective imagery, a part of CBT, with prompts asking about participants' details for future enjoyable events (Hallford, Sharma, & Austin, 2020; Renner et al., 2021). This activity has been shown to increase anticipatory pleasure and joy regarding the upcoming event (Hallford et al., 2020). By providing these prompts on an MHealth smartphone application, individuals with depressive symptoms or a depression diagnosis would have access to the activity without having to make an appointment or leave their homes. By performing this activity when the user needed motivation, anticipatory pleasure could increase, as shown in past research (Hallford et al., 2020). Furthermore, this could help fortify individuals with depression to gain motivation in the more mundane and everyday areas of their lives.

The purpose of the current experiment is to provide support for prior research and to investigate whether individuals with higher levels of depression will prefer a module with an interactive agent when compared to a module with a passive therapeutic experience within a short period. Additionally, this experiment will seek to explore the user's preferences, impressions, and judgments regarding the 3D, life-like virtual agents. All participants will complete four cCBT modules, two active modules with a conversational virtual agent and two passive modules in video format.

Experimental Aims

Aim 1: The first aim of this experiment is to investigate whether the CBT modules will significantly reduce levels of negative mood over the course of the experiment.

<u>Aim 2:</u> The secondary aim of this experiment is to investigate the preference of the depressive group regarding active vs. passive therapeutic experiences with a virtual agent.

<u>Aim 3</u>: The tertiary aim of this experiment is to investigate user preference regarding their level of connection, trust, likeability, and overall experience with the virtual agent.

Exploratory Aims

<u>Aim 4</u>: The first exploratory aim is to determine the reason why participants would prefer an active or passive condition.

<u>Aim 5:</u> The second exploratory aim of this experiment is to ascertain the rationale as to why a participant would choose one virtual agent over the other.

Hypotheses

Aim 1 Hypothesis:

H1: Users, regardless of depressive group, will report lower ratings of state worry, sadness, and rumination after completing the four modules.

Please note, condition type (Active vs. Passive) is not relevant for these hypotheses because, by the completion of the four modules, participants will have completed ***both*** conditions; only pre- and post-measures measures will be given.

Aim 2 Hypotheses:

H2: Participants will prefer talking to the agent (the active condition) over simply listening to the agent (the passive condition). This preference will be magnified among those with depressive symptoms compared to the non- depressive group.

Aim 3 Hypothesis:

H3: Users will experience higher levels of comfort with the agent in the active condition when compared to the passive condition; this increase will be magnified for the depressive symptom group compared to non-depressive group.

H4: Users will experience higher levels of trust with the agent in the active condition when compared to the passive condition; no differences between the depressive symptom and non-depressive group are expected.

H5: Users will express higher levels of likeability of the agent in the active condition when compared to the passive condition; this increase will be magnified for the depressive symptom group compared to non-depressive group.

H6: Users will express a higher overall satisfaction with the agent in the active condition when compared to the passive condition; this increase will be magnified for the depressive symptom group compared to non-depressive group.

Exploratory Hypotheses:

H7: Participants will prefer the active condition primarily due to the high levels of engagement.

H8: The primary reason participants selected their agent is because the agent matched the user's gender identity.

METHOD

Participants

Prior to starting the experiment , a G*power analysis was conducted to determine the number of participants needed to maintain an 80% power level to detect an effect at the level of significance (p = .05). A total of 56 participants are needed to maintain the desired power level, however, a total of 70 participants will be recruited to account for potential data loss or attention check failure. The G*power analysis was conducted according to H1 for the current experiment. Participants will be incentivized to participate in this experiment with compensation in the form of course credit or extra credit for a class.

Exclusion Criteria: Participants were excluded from participating in the experiment for four reasons: 1) the individual was under the age of 18 and classified as a minor, 2) the individual was not fluent in English, and 3) the individual was not enrolled as a student at Clemson University. During data cleaning, data was excluded for three reasons: 1) the participant did not complete the experiment, and 2) an individual failed one or more attention checks.

Measures

Depression & Personality Measures

<u>Patient Health Questionnaire -8</u>*:* The PHQ-8 was prior to commencing the full experiment to screen participants for depression. This questionnaire is utilized to estimate depressive symptom severity over the past two weeks ranging from mild (score of less than four) to severe (score of twenty or higher) (Kroenke, Sptizer, & Williams, 2001; Wu

et al., 2020). Users responded to a total of eight questions regarding negative emotions or moods, feelings of hopelessness, self-hatred, anhedonia, fatigue, inability to concentrate, appetite, and restlessness on a Likert scale from zero (not at all) to three (nearly every day) (Kroenke et al., 2001; Wu et al., 2020).

Subjective Mood Assessment: The SMA is an 8-item assessment that was used to investigate the subjective change in mood throughout experiment. Participants responded to eight statements, such as "I feel happy" or "I feel worried", on a 5-item Likert scale ranging from "Strongly Disagree" to "Strongly Agree".

User Preference Measures

The Working Alliance Inventory (WAI): The WAI consists of three different subscales investigating the alliance between tasks, bonds between the participant and an agent, and their goals (Horvath & Greenberg, 1989). Each subscale consists of twelve questions with a seven-point response system ranging from 'Never' (1) to 'Always' (7) (Horvath & Greenberg, 1989). While this inventory is customarily utilized to analyze the bond between the participant and a therapist, clinician, or medical expert, their goals, and their tasks, this experiment will use the bond subscale to investigate the relationship between the participant and the virtual agent (Leo or Val). Example questions include "I am confident in Leo's/Val's ability to help me" and "I feel that Leo/Val appreciates me".

The Character Rating Questionnaire (CRQ): This 56-item questionnaire was used to investigate the user's preferences and relationship with the virtual agent. The first fifty questions will be answered on a 7-point Likert scale ranging from 'Strongly Disagree' (1) to 'Strongly Agree' (7). Example questions include: "Leo's/Val's movements seemed natural”, “Leo/Val was stiff”, and “I had fun interacting with Leo/Val”. The remaining four questions were qualitative & short answer questions allowing the participant the ability to type in their answers using their own words. Example questions included: “I would best describe Leo’s/Val’s appearance as…”, “How would you best describe Leo’s/Val’s personality?”, “How would you best describe Leo’s/Val’s emotional state?”, and “How would you describe the overall behavior of Leo/Val?”.

Virtual Agent Impressions Questionnaire (VAI): The VAI is an 8-item questionnaire that analyzes user preferences regarding their satisfaction, willingness to continue to work with the virtual agent, level of trust of the agent, level of similarity and naturality, and perception of likeability, knowledge, and relationship with the agent. These questions have pre-supplied answers in a 5-item Likert-scale format ranging from ‘Not at all’ to ‘Very much so’ depending on the topic of the question. Example questions include: “How much do you like Leo/Val?”, “How natural was your conversation with Leo/Val?”, and “How would you characterize your relationship with Leo/Val?”.

Questions Regarding User Preference: Additional questions were asked regarding the user’s preference for the agent and overall experiment. The first ten statements will be answered via a 5-point Likert scale ranging from ‘Strongly Disagree’ (1) to ‘Strongly Agree’ (5), and these questions include: “I liked being able to talk to Leo or Val.”, “The app would be useful for my health and well-being.”, and “I prefer more passive engagement with Leo and Val.”. Next, two preference questions asked the participants to indicate their inclination towards more active modules or passive video options and explain their rationale. Secondly, they were asked for their reason for selecting their virtual agent of choice. Lastly, one ranking and two short answer questions were used to discover the user's preference regarding their favorite and least favorite modules completed in the experiment.

Materials

This experiment required the use of the AirHeart V3 application, a computer, monitor, keyboard, mouse, a microphone, a pair of noise-canceling headphones, and two rooms conjoined by a door. In addition, the experiment consisted of three main surveys: a pre-experimental survey, a mid-point survey, and a post-experimental survey. All surveys were created using the online survey maker *Qualtrics*.

Pre-Experimental Survey: This survey consisted of four questionnaires: 1) demographics, 2) the Big-5 Inventory Short Version (BFI-10), 3) the PHQ-8, and 4) the SMA. The demographics questionnaire consists of questions regarding the participant's age, gender, race, highest level of education, major or profession, mental health diagnosis, video games and phone usage, vision, and fluency in English.

Mid-Point Survey: This survey consisted of three questionnaires: 1) the Working Alliance Inventory (WAI), 2) a character rating questionnaire (CRQ), and 3) the Virtual Agent Impressions questionnaire (VAI).

Post-Experimental Survey: This survey consisted of six questionnaires: 1) the PHQ-8, 2) the WAI, 3) the CRQ, 4) the VAI questionnaire, 5) the SMA, and 6) additional questions about the agent and overall experiment.

AirHeart V3: The newest iteration of AirHeart included the use of a male (Leo) and female (Val) virtual characters and different forms of CBT modules. ECAs were used in the active condition and a virtual character in the passive condition. See Figure 3.1 for a visualization of Leo and Val. The user selected the specific gender of their agent to allow for autonomy and comfort. Additionally, personalization was added to the experiment through the inclusion of the user's name which will be utilized by the agent of their choice in the CBT modules. A total of four modules, psychoeducation EFT, identifying and challenging maladaptive thoughts, and problem-solving, were included. The first three modules were identical to three of the modules utilized in the original AirHeart pilot experiment (Six et al., 2022). The new EFT module was created with the original premises and guidance from prior research studies which investigated EFT in relation to depression (Hallford et al., 2020; Renner et al., 2021). It encouraged participants to explore the mundane activities of their daily life, as well as note the potentially exciting events occurring in both the near and distant future. The modules had two different forms: interactive (active condition) and video-based (passive condition). The participant completed two active and two passive modules in a randomized order; the conditions were run together to prevent user error (i.e. the user will complete two active modules and then two passive modules, or vice versa).

Figure 3.1

The Embodied Conversational Agents Leo (Left) & Val (Right)

Active Condition. In the active condition, the ECA, Leo or Val, was utilized as a virtual therapeutic coach engaging the participants in conversation and leading them through the modules and activities. The agent implements speech-to-text and text-to-speech dialogue structures to emulate human-like conversational dialogue. As the virtual agent speaks, their words were presented on the computer screen for the user to read along in addition to the audio dialogue. The agent supplied the participant with information and asked questions to engage the user and ensure clarification of the information. Activities were completed in a speech-to-text and text-to-speech format. Users were informed to verbally respond to the agent when a red microphone appears on the screen.

Passive Condition. In the passive condition, the speech-to-text dialogue feature was not utilized. This condition consisted of a video that contains the agent teaching the user about the information for the specified module; however, the agent did not ask the user questions nor expect a response. The conversational aspect within the active condition was replaced by a passive observation in this condition. Additionally, the ECA form of Leo or Val was replaced by a static, non-animated 2D image of Leo or Val.

Experimental Design

To investigate the hypotheses mentioned above, a mixed effects experimental design was utilized for this experiment. All participants completed four CBT modules split evenly between the two conditions: active and passive; the between element for the experiment is the two samples: individuals with high or low depressive symptoms. A repeated effects ANOVA was used to assess the primary hypothesis of the experiment.

Procedure

The user entered the lab and was seated at a desktop computer with a monitor, keyboard, microphone, and mouse. The user was instructed to read through a consent form and sign at their leisure. After verbal and written consent, the individual then completed the pre-survey questionnaire.

After completion, the user was informed about the speech-to-text features within the experiment and ensured of the privacy and confidentiality of their information:

"*You are about to complete a total of four different modules. For two of them, you will be asked to verbally respond to the agent by speaking into the microphone. The privacy and confidentiality of your information are of the utmost importance. To ensure privacy, I will be in the other room with the door between us shut. Additionally, I will be wearing noise-canceling headphones.*"

After a verbal confirmation of understanding, the researcher exited the room, shut the door, put on the noise canceling headphones, and the participant began the modules. Upon completion of the first two modules, the participant completed the middle survey and took a short 30-second fatigue break. Afterward, the participant completed the remaining two modules before being asked to complete the post-survey questionnaire. After the post-survey questionnaire was submitted, the participant opened the door to inform the researcher that they are finished with their tasks and left the lab.

Data Analysis

For H1, a 2 (Depressed vs. Non-Depressed) X 2 (Pre vs. Post) mixed ANOVA was conducted to explore the change in subjective state feelings (worry, sadness, rumination) before and after the completion of the four CBT modules.

For H2, a 2 (Active vs. Passive) X 2 (Depressed vs. Non-depressed) chi-squared test was utilized to investigate the preference for the active condition or the passive condition for individuals in the depressive condition.

For H3 a 2 (Active vs. Passive) X 2 (Depressed vs. Non-depressed) ANOVA was used to compare the effect of active and passive conditions as well as level of depressive symptoms on the perceived comfort with the agent. The level of comfort was defined by the question “Leo’s/Val’s behavior was comfortable”.

To investigate H4, a One-Way ANOVA was conducted to compare the effect of condition on levels of trust with the agent.

To investigate H5, a 2 (active vs. Passive) X 2 (depressed vs. non-depressed) ANOVA was conducted to compare the effect of depressive symptoms and condition on level of likeability of the agent.

For H6, a 2 (Active vs. Passive) X 2 (depressed vs. non-depressed) ANOVA was used to investigate the effect of depressive symptoms and condition on overall satisfaction with the virtual agent.

For H7 & H8, two thematic qualitative analyses were completed to assess if most participants preferred the active condition due to higher levels of engagement (H7) and if participants overall prefer an agent that matches their gender identity (H8). The analyses followed the six-stage method outlined by Braun and Clarke (Braun & Clarke, 2006). To avoid biases, participant IDs and identifiers, condition (passive vs. active interaction with the agent), and information regarding depressive group were removed. Researchers reviewed the theme suggestions and made final selections once a consensus had been reached. Upon finalization, all qualitative data was assigned to one or more of the themes.

RESULTS

A total of 76 participants have been recruited, but only the data from 68 individuals is included in the current analysis (M_{age} = 22.92, SD_{age} = 3.40; female = 80.9%; N_{Dep} = 37; $N_{Non\text{-}Dep}$ = 31). Eight participants have been excluded: seven due to failing one or more attention checks and one for lacking fluency in English.

For H1, the 2 (Depressed vs. Non-Depressed) X 2 (Pre vs. Post) mixed ANOVA was conducted to explore the change in negative state feelings before and after the completion of the four CBT modules. A statistically significant difference was discovered for the main effect of time and the interaction effect between worry and depressive condition ($F(1, 66) = 4.33$, $p = .04$; $\eta p^2 = .06$). No significant effects were ascertained for depressive condition regarding state worry. Furthermore, no significant main effects nor interaction effects were discovered for state sadness. However, two significant main effects for depressive condition and state rumination were also determined (Depressive Condition: $F(1,66) = 10.60$, $p = .002$, $\eta p^2 = .88$; Time: $F(1,66) = 3.73$, $p = .03$, $\eta p^2 = .07$). No significant effects were discovered for the interaction between pre- and post- state rumination and depressive condition. Thus, H1 is partially supported. Full results and demographics for this hypothesis can be seen in Table 3.1.

Table 3.1

Results of the 2x2 Mixed ANOVA for Sadness, Worry, & Rumination

Sadness			
Means (*M*)		**Standard Deviation (*SD*)**	
Average	PRE: 2.27 POST: 2.37	Average	PRE: .98 POST: 1.13
Depressive	PRE: 2.13 POST: 2.09	Depressive	PRE: .81 POST: 1.00
Non-Depressive	PRE: 2.34 POST: 2.52	Non-Depressive	PRE: 1.06 POST: 1.17
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp^2)**
Depressive Main Effect	1.87	.18	.028
Time Main Effect	.28	.60	.004
Interaction Effect	.74	.39	.011

Worry			
Means (*M*)		**Standard Deviation (*SD*)**	
Average	PRE: 2.88 POST: 2.7	Average	PRE: 1.33 POST: 1.17
Depressive	PRE: 2.91 POST: 2.39	Depressive	PRE: 1.24 POST: 0.99
Non-Depressive	PRE: 2.86 POST: 2.86	Non-Depressive	PRE: 1.39 POST: 1.23
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp^2)**
Depressive Main Effect	.51	.48	.008
Time Main Effect	*4.33*	*.04*	*.06*
Interaction Effect	*4.33*	*.04*	*.06*

Rumination			
Means (*M*)		**Standard Deviation (*SD*)**	
Average	PRE: 3.12 POST: 2.82	Average	PRE: 1.39 POST: 1.25
Depressive	PRE: 2.65 POST: 2.13	Depressive	PRE: 1.30 POST: 1.25
Non-Depressive	PRE: 3.36 POST: 3.18	Non-Depressive	PRE: 1.38 POST: 1.21

Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp²)**
Depressive Main Effect	*10.60*	*.002*	*.09*
Time Main Effect	*3.73*	*.03*	*.07*
Interaction Effect	1.18	.28	.02

H2 investigated whether individuals in the depressive or non-depressive group more avidly preferred an active interaction with the virtual agent to passively listening to the agent. Results from the 2 X 2 chi squared test failed to reveal a significant preference for the active condition (X^2 = .013, *p* =.91). This indicates no strong difference in preference for active vs. passive conditions, however, a strong preference emerged for the active condition in both depressive (N_{active} = 27/37, $N_{passive}$ = 10/37) and non-depressive (N_{active} = 23/31, $N_{passive}$ = 8/31) conditions. An additional Chi square analysis was conducted on this matter; results demonstrated that participants in both the depressive and non-depressive conditions significantly preferred talking to the agent (Depressive: X^2 = 10.0, *p* =.002, Talking: 30/40, Listening: 10/40; Non-Depressive: X^2 = 5.14, *p* =.023, Talking: 20/28, Listening: 8/28). Thus, these findings partially support H2.

For H3, the 2 (Active vs. Passive) x 2 (Depressive vs. Non-Depressive) ANOVA for comfort revealed statistically significant main effect of Condition (*F* (1,135) = 7.10, *p* = .009) such that participants in the active condition (*M* = 4.49, *SD* = 1.44) perceived the agent's behavior as more comfortable than the passive (*M* = 3.74, *SD* = 1.76) condition. However, there was no significant main effect of depressive symptoms in the depressive and non-depressive groups (*F*(1, 135) = .006, *p* = .94; $M_{depressive}$ = 4.11, $SD_{depressive}$= 1.66 vs. $M_{non\text{-}depressive}$ = 4.13, $SD_{non\text{-}depressive}$ = 1.61). The interaction between depressive groups and condition was also non-significant, ($F(1, 135) = .90$, $p = .34$). These results support H2 for expected differences between active and passive conditions; however, given the nonsignificant interaction, this difference was not magnified in the depressive symptom group.

For H4, the one-way ANOVA was conducted to assess whether a significant difference regarding trust could be found between the active and passive conditions for the total population. A significant difference was discovered, F(1,66) = 29.42, $p < 0.001$, with participants reporting higher levels of trust during the active condition ($M = 2.75$, $SD = 1.15$) when compared to the passive condition ($M = 2.49$, $SD = 1.18$). Therefore, H4 was supported.

For H5, the 2 (Active vs. Passive) x 2 (Depressive vs. Non-Depressive) ANOVA was used to investigate the effect of likeability, failed to detect a significant difference was found between the two conditions (F(1, 135) = .33, $p = .57$; $M_{\text{active}} = 4.62$, $SD_{\text{active}} = 1.51$ vs. $M_{\text{passive}} = 4.47$, $SD_{\text{passive}} = 1.56$), nor between the depressive groups ($F(1, 135) = .95$, $p = .33$; $M_{\text{depressive}} = 4.66$, $SD_{\text{depressive}} = 1.41$ vs. $M_{\text{non-depressive}} = 4.40$, $SD_{\text{non-depressive}} = 1.67$). Additionally, no significance was found for interaction effect ($F(1, 135) = .07$, $p = .79$). Therefore, H5 is not supported.

For H6, the 2(Active vs. Passive) x 2 (Depressive vs. Non-Depressive) ANOVA used to investigate the effect of overall satisfaction did reveal statistically significant results for the two conditions ($F(1,135) = 5.17$, $p = .025$; $M_{active} = 4.25$, $SD_{\text{active}} = 0.99$ vs. $M_{\text{passive}} = 3.84$, $SD_{\text{passive}} = 1.08$). However, no significance was found between depressive groups ($F(1, 135) = 1.17$, $p = 28$; $M_{\text{depressive}} = 3.96$, $SD_{\text{depressive}} = 1.01$ vs. $M_{\text{non-}}$

$_{\text{depressive}} = 4.15$, $SD_{\text{non-depressive}} = 1.11$), nor for the interaction between depressive groups and condition ($F(1, 135) = .09$, $p = .76$). Therefore, H6 is partially supported.

Results for Exploratory Qualitative Analyses

To investigate H7, two researchers conducted a thematic analysis to ascertain the participants' primary rationale for their preference for the active or passive conditions within the experiment. One participant informed the researchers that they did not prefer the active nor the passive, thus their response was excluded from this analysis. Of the 67 remaining participants, 73.1% (49/67) preferred the active condition, 26.9% (18/67) preferred the passive condition, and 1 participant did not have preference between the two. After removing the identifiers, five themes were created for the active condition: 1) Engagement/Maintains Attention, 2) Connection, 3) Self-Reflection, 4) Improved Comprehension, and 5) Novelty.

As predicted in H2, many participants attributed their preference for the active condition to the increased engagement with the virtual agent (49.3%). Participants commonly included supporting rationale in conjunction with engagement as to why they preferred the active condition, such as liking the interaction, feeling present, and inherent conversation. Participants also attributed their engagement to the agent's ability to sustain their attention during the modules. Thus, H7 is supported.

The second reason participants supplied for their preference for the active condition was the strong sense of human connection emanating for the agent (28.4%). This connection came in a variety of forms, whether as the conversational component, the feeling of having someone to listen to your thoughts or worries, or the lack of awkward interactions. Thirdly, participants cited self-reflection as a reason they selected either the active or passive condition (25.4%). Self-reflection can include connecting the learned information to their own lives and experiences as well as simply sorting through their own thoughts.

Fourthly, 10.4% of participants said that the active condition permitted a stronger sense of self-reflection by encouraging them to create examples that related to their own life. Lastly, two participants (3.0%) preferred the active condition because of its novel nature; these participants stated that chatting with the agent was favored as it was a unique and fun experience. For visualization of this information and corresponding quotes, please see Table 3.2.

Table 3.2

Visualization of Qualitative Data for H7: Reasons Why Participants Preferred the Active Condition

Active Condition: 72.1% (49/67)				
Themes	**% of Overall Sample**	**% of Active Sample**	**Examples**	**% of Overall Sample: Depressive vs. Non-Depressive**
Engagement/ Attention	49.3%	67.3%	"Because it is interactive and engaging!" "It kept my attention to the modules; I did not doze off when I interacted with Val."	D: 19.4% ND: 29.9%
Connection	28.4%	38.8%	"Felt more connected to the character when I could talk to them, and they respond to my interactions." "Feels nice to have someone listen to you…rather than just listening to some information."	D: 9.0% ND: 19.4%
Self-Reflection	25.4%	34.7%	"It was easier to understand the importance of the information in my own life/situations."	D: 7.5% ND: 17.9%

			"I sort my thoughts better when I have to voice them out loud."	
Improved Comprehension	10.4%	14.3%	"I think being able to read and listen to what was being said helped me understand everything and make it easier to envision someone real reading it." "It is more beneficial for me to practice the skills learned."	D: 3.0% ND: 7.5%
Novelty	3.0%	4.1%	"Talking to an agent makes the experience unique, compared to reading off a passage." "[It was a] new experience to test the bot, chatting with the agent was fun."	D: 1.5% ND: 1.5%

A total of 18 participants preferred the passive condition to the active. Three themes immerged for the passive condition: 1) Negative Interactions with the Agent, 2) Prefer to Listen, and 3) the Pressure to Respond. Ten participants (15.0%) stated that they favored the passive condition due to a negative interaction with the agent in the active condition. These experiences include feeling the agent was scary, unnatural, or unnecessary with an artificial or mechanical voice. Further complaints include being vocally cut off by the agent, which led to the disillusionment of a conversation with a therapist to simply talking to a screen. The other common negative response supplied by participants was that they felt a pressure to respond to the agent's questions in the active condition (4.5%). Participants often reported that they simply did not know what to say in response to the agent or that they could not think of an answer.

Lastly, eight participants (12%) supplied that they simply preferred listening to the agent as opposed to actively engaging. Rationale for this answer included the partiality towards passively absorbing the information, the opinion that listening was a quicker way of learning the information, or an individual's high level of fatigue which may skew preference towards listening. For visualization of this information and corresponding quotes, please see Table 3.3.

Table 3.3

Visualization of Qualitative Data for H7: Reasons Why Participants Preferred the Passive Condition

Passive Condition: 26.5% (18/67)				
Themes	**% of Overall Sample**	**% of Active Sample**	**Examples**	**% of Overall Sample: Depressive vs. Non-Depressive**
Negative Interactions with the Agent	15.0%	55.6%	"The talking felt unnatural and forced." "I got cut off a lot in conversation and couldn't ask questions." "The voice was very mechanical and lacked that natural touch."	D: 3.0% ND: 11.9%
Preference for Listening	12.0%	66.7%	"I prefer this option because I feel like I am a good listener and is a better way for me to absorb the information being given." "If my goal is to get the information that would help me, I prefer getting it in the fastest way possible. Listening seemed like a faster means to this end."	D: 7.5% ND: 4.5%
Pressure to Respond	4.5%	16.7%	"It felt awkward to have to respond and I felt pressured to say something within a certain time frame, so I feel like I wasn't able to actually say what I wanted to." "I don't know what to say back."	D: 3.0% ND: 1.5%

Similar to H7, for H8, two researchers conducted a thematic analysis to ascertain the participant's rationale for selecting their virtual agent. For the selection of a virtual agent, participants heavily favored Val, the female agent (57/68), to Leo, the male agent

(11/68). After removing the identifiers, 12 participants (17.6%) were excluded for citing "no reason" as the rationale for selecting their agent. Seven themes were created for the remaining reasons: 1) Gender Match/Relatability, 2) Comfort Talking to a Female, 3) Preference for Talking to a Female about Personal Matters/Understanding, 4) Habit/Prior Experience, 5) Negative Reaction to the Agent, 6) Cross Gender Match, and 7) Comfort Talking to a Male.

In accordance with H8, the primary reason participants cited for their choice of agent was a gender match or relatability to themselves (46.4%). Secondly, participants stated that they chose Val, the female agent, because they were more comfortable conversing with a female (41.1%). This could be comfort overall, or comfort specific to speaking with females about mental health or personal topics. Participants also stated rationale for this such as females having a more comforting nature or lighter, less intense voice; thus, H8 is supported. Conversely, only two participants (3.6%) cited they chose Leo because they are more comfortable speaking with a male agent. Fourth, 11 participants (19.6%) stated they chose Val because of their preference for talking to females about personal or sensitive matters. Thus, H8 is supported.

The remaining three categories were less commonly cited as the rationale for selecting their agent, ranging from around 3% - 7%. Four participants (7.1%) cited habit or prior experience with virtual agents for their rationale for selecting their agent. One participant stated that they had prior experience seeing Leo, so they selected the agent with which they had not yet interacted. Another stated that most of the AI assistance (like

Alexa or Siri) that they commonly interacted with were female, so they selected the female agent.

Both remaining categories, a negative reaction to the agent and selecting their agent based on a cross gender match, were cited by 5.4% of participants. The three participants who selected their agent based on a negative reaction to the agent selected Val due to Leo's creepy appearance (2/3) and disturbing voice (1/3). For the cross-gender match, individuals cited the desire to break down their own gender biases. For visualization of this information and corresponding quotes, please see Table 3.4.

Table 3.4

Visualization of Qualitative Data for H8: Investigating How Participants Selected Their Virtual Agent

Themes	% of Sample*	Examples	Depressive vs. Non-Depressive*
Gender Match/ Relatability	46.4%	"I chose Val because she is a female, and I am too. I figured I'd be able to relate more to her than Leo." "I am a girl, so I wanted to speak to a virtual agent of the same gender."	D: 14.3% ND: 32.1%
Comfort Talking to a Female	41.1%	"I always feel more comfortable conversing and opening up with a female due to their comforting nature." "I've always felt more comfortable with female health care providers so I assumed this would be a similar situation."	D: 14.3% ND: 26.8%
Preference for Talking to a Female about Personal Matters/ Understanding	19.6%	"I chose the female agent Val because when it comes to talking about personal situations in my life I prefer to talk to a female." "I generally prefer speaking to women about mental health and psychology."	D: 12.5% ND: 7.1%

Habit/Prior Experience	7.1%	"I think I had interacted slightly more with Leo as part of a different software in the past, so I wanted to give Val a try." "Habit, I guess, probably because my Siri is a female agent."	D: 3.6% ND: 3.6%
Negative Reaction to Agent	5.4%	"The male looked creepy." "Male AI voices are bad, in my experience"	D: 3.6% ND: 1.8%
Cross Gender Match	5.4%	"I chose Val to establish a cross-gender dynamic between the user (male) and the agent (female)." "Gender bias in a way that the opposite gender might be more engaging."	D: 0% ND: 5.4%
Comfort Talking to a Male	3.6%	"Felt more comfortable talking about mental health topics with another male as I often talk about these types of issues with my close male friends and family members." "Felt comfortable talking to male agent"	D: 1.8% ND: 1.8%

*: Sample of usable data for this analysis is 68 participants

DISCUSSION

This experiment sought to investigate whether one therapeutic CBT session with a ECA could reduce negative state emotions (H1). Additionally, the current research sought to determine user preference for active or passive engagement (H2) and establish whether these conditions impacted comfort, trust, likeability, and overall satisfaction with a ECA (H3-H6). This endeavor also explored potential rationale for why participants prefer active or passive engagement (H7) and why they selected either the male (Leo) or female ECA (Val) (H8).

While prior research suggests that individuals need to complete anywhere from four to fifteen sessions of CBT prior to seeing major differences in mood and negative symptoms (Chaves et al., 2017; Gould et al., 2012; Grosse Holtforth et al., 2019; Oud et al., 2019; Tandon et al., 2014), the current experiment demonstrated a reduction in state negative emotions during one session. It was originally hypothesized that completing the four CBT modules would reduce negative state emotions of worry, sadness, and rumination; while there was no significant change in the user's degree of sadness, rumination and worry both showed significant improvement, thus H1 was partially supported. These findings lend support to prior research that has also supported the idea that virtual agents, like chatbots, can demonstrate success in improving well-being and quality of life through CBT-based interventions (Burton et al., 2016; Fitzpatrick et al., 2017; Fulmer et al., 2018; Gaffney et al., 2019). However, as the current experiment only included around half an hour of exposure to the CBT modules and the virtual agents, the brevity of the timeframe may not accurately present findings which correlate to longer, longitudinal experiments. As such, these findings should be taken lightly. Future studies should replicate this procedure over a longer duration for comparison.

During the current experiment, a total of four CBT modules were completed by participants: two with an active and interactive agent and two with a more passive, video-like format. In accordance with prior research, it was hypothesized that individuals with higher levels of depressive symptoms would more avidly prefer the active condition compared to the passive condition (Kertes et al., 2011; Lee et al. 2021; Uzoka, 1983). This rationale comes from participants' aspiration to take an active role in their therapy as well as the desire to connect with another person (Kertes et al., 2011; Short et al., 1976; Uzoka, 1983). However, the current experiment did not lend support to this hypothesis (H2). The directionality of the data suggested that individuals, regardless of depressive symptoms, supplied a stronger preference for the active condition (50/68; 73.5%)

compared to the passive condition (18/68; 26.5%). While these findings do not lend support to H2, it does suggest that active involvement should be readily supplied to all participants regardless of symptom severity. Furthermore, a large portion of participants cited their rationale to the engagement or attention maintaining factors presented by the active condition (supporting H7). As for the passive condition, many individuals who preferred this option did so due to the negative interaction with the virtual agent. Further developments within the AirHeart applications to remedy these negative interactions could potentially change this outcome. Future studies should replicate the current experiment in a longitudinal format to ascertain whether the current findings maintain their validity.

Throughout the experiment, participants accumulated around half an hour of screen time learning about different topics relating to mental health and practicing the new skills learned from Leo and Val, the virtual agents. The findings from the current experiment demonstrate that Leo and Val’s behavior was viewed as more natural and comforting during the active condition as opposed to the passive condition, thus supporting H3. Additionally, higher levels of trust in the agent were expressed regarding the active condition, thus supporting H4. Participants reported feelings of awkwardness during the passive condition due to the lack of movement from the virtual agent. The active condition felt more realistic and conversive than simply listening to a computer program speak. Furthermore, the current experiment also concluded that the participants were more satisfied with the agent overall in the active condition as opposed to the passive condition, thus supporting H6. These findings also support prior research stating that humans tend to feel more comfortable and trusting of interactive entities when compared to stationary images of a being (Mori et al., 2012; Shin et al., 2019; Touré-Tillery & McGill, 2015; Van Pinxteren et al., 2019). Future studies should implement an experimental design with an active virtual agent who utilizes body movement, facial expression, and non-verbal body language in both conversational and solely educational settings to determine if a preference emerges.

The sole finding which did not appear to significantly differ between the two conditions was that of the likeability of the agent, thus failing to support H5. There appears to be some impartiality regarding this question, as most participants rated both conditions between "neither agree nor disagree" (4) and "somewhat agree" (5) (M_{active} = 4.62, SD_{active}= 1.51 vs. M_{passive}= 4.47, SD_{passive}= 1.56). This could indicate that the short amount of time spent with both agents was not enough to indicate a difference in preference, thus resulting in the semi-positive neutrality of the response. Further research on this topic with additional follow-up questions regarding appearance, voice, and other elements which could encompass the concept of "likeability" should be conducted.

Lastly, this experiment investigated the participant's rationale as how why they selected the male or female virtual agent. The knowledge of how an individual makes the selection or what factors are included within that decision could help MHealth app developers to determine specific options to prioritize and include in their applications. For applications which include mental health themes, modules, or lessons, participants seem to prefer an agent which matches their own gender (supporting H8). According to participants, this preference stems from a better understanding of what the individual is experiencing, their experience with having a medical provider of the same gender, or the fact that they prefer to turn to their friends of the same gender to talk about mental health related topics. However, the second most common option was the desire to talk to a female agent (Val) because participants feel more comfortable opening up to them. These findings suggest that users should be able to select the gender and voice of their virtual agent to ensure all users feel comfortable learning and communicating with their agent.

Limitations

The current experiment presents two main limitations. Firstly, the design of the current experiment set to investigate two popular formats for mental health applications: an interactive, conversational condition and a passive, video-like condition. In creating the different formats, two unintentional changes between conditions occurred in the form of conversation and animation design. Within the active condition, the virtual agent design included body and mouth animations as well as the ability to converse with the participant. For the passive condition, no animations, movements, or conversational elements were included in the design for the virtual agent. This drastic difference could have unintentionally created a design which allows for statistical significance while failing to include the middle ground between the two designs.

Secondly, the facial design for both virtual agents utilized a neutral expression. Prior research has posited that individuals with major depressive disorder or specific depressive symptoms which make them interpret their surroundings as hostile or negative, tend to perceive neutral facial expression as negative (Bodenschatz et al., 2021). While facial expression was not intentionally designed into the experiment, the potential that depressive users perceived the neutral expression as negative could have impacted the current experiment's results. For future iterations for the AirHeart application, virtual agents will be programmed to display a soft smile while listening or waiting for the user.

Overall, the current experiment adds to the literature investigating the effectiveness of virtual agents in presenting therapeutic teachings and techniques to a wide variety of individuals with diverse backgrounds and needs. In addition to providing auxiliary support to this idea, these findings suggest that participants are receptive to the idea of using virtual applications. This could help supplement therapy sessions with a licensed therapist or clinician by providing an interactive agent with whom individuals can practice their therapeutic techniques or exercises. While the fact that virtual agents could fully replace human therapists remains unlikely, providing users with an application where they can receive virtual assistance from a trained agent could help reduce the frequency of therapy sessions. This in turn would allow medical professionals to accept more patients, and patients from a wide variety of backgrounds, financial standings, and geographical locations could access care.

CHAPTER 4

EXPERIMENT 3: AirHeart V4 – The Longitudinal Experiment

The purpose of this experiment was to implement the findings of the previous two studies into a newly enhanced smartphone mental health application format utilized over two weeks. To accomplish this goal, the previous smartphone version of AirHeart was re-evaluated and upgraded. Previously, the 83 participants who tested AirHeart V1 over a two-week duration rated the usability as below-average on the SUS scale (54.35) (Six et al., 2022). To remedy this below average score, seven steps were taken to create a new and improved AirHeart: 1) user persona creation, 2) journey map creation, 3) competitor analysis creation, 4) prototype creation, 5) user testing, 6) heuristic evaluation, and 7) final changes to the prototype design.

User Personas

The first action for reconstructing the AirHeart app was to create personas to act as a constant reminder of the type of users who would utilize this type of application. User personas act as fictitious individuals who may be interested in using a product in a traditional or atypical way (Cooper, 1999; Miaskiewicz & Kozar, 2011; Pruitt & Adlin, 2010). Personas have demonstrated numerous benefits towards product development including improving a focus on a user-centric approach, determining which features should be included in the product, simplifying decision-making, and promoting a transition to user-friendly products (Cooper, 1999; Cooper & Reimann, 2015; Long 2009; Miaskiewicz & Kozar, 2011).

A total of six personas were created to represent the user demographic. These personas ranged from a freshman suffering from homesickness to a transgender student attempting to dissolve their negative self-view to graduate students who compare themselves to their more successful peers. These personas were chosen to represent a diverse population of people who may experience depressive symptoms on a college campus. The user personas were utilized in all remaining six steps to ensure a user-centric design. The graduate student, Amit's, persona was used in the journey map to help understand the thought process behind deciding to use AirHeart. Additionally, the personas were utilized in the competitor analysis to help guide decisions regarding which features would help fulfill their needs. During the final changes following the heuristic evaluation, the team revisited the user personas' needs to ensure fulfillment. The six personas can be found in Appendix A as Figures A1-A6.

Journey Map

To aid with visualizing the user's progression from discovering the AirHeart application to deciding whether to continue using the app, a journey map was created for each of the user personas (Howard, 2014; Micheaux & Bosio, 2019). This map depicts the emotions, thought processes, needs, decisions, and overall satisfaction that the user may experience during each step of their journey with AirHeart (Howard, 2014; Micheaux & Bosio, 2019). Creating a journey map has been shown to help improve user experience by identifying users' unmet needs and pain points (Micheaux & Bosio, 2019; Oliveira et al., 2019). Furthermore, it can highlight a common rationale or path taken by the user, such as utilizing an MHealth app when experiencing high levels of stress during finals, and infrequent paths, such as using the app to learn more about a loved one's symptoms. Amit's journey map can be found in Appendix A as Figure A7.

Competitor Analysis

The first step to creating a quality product is to identify a need for that product in the current market. To identify other competing products and ascertain what novelty the AirHeart app provides, a competitor analysis was conducted. This analysis is used to assess competitors' products and performance along with user-satisfaction. It investigates both the overall design of the product and individual subcomponents, such as features within an MHealth app (Baymard Institute, 2022). Furthermore, it can identify gaps in the market, standards and expectations, commonly reported usability issues, and the strengths and weaknesses of one's product compared to other products on the market (Baymard Institute, 2022). Most importantly, a competitor analysis can provide evidence to encourage change and new development (Baymard Institute, 2022). Ultimately, the goal is to provide a competitive product at the level or exceeding expectations of the market.

To understand the features and capabilities offered by other MHealth apps currently available, a competitor analysis was conducted for five popular MHealth apps prominently advertised to help improve depressive symptoms: What's Up, Sanvello, Youper, Replika, & Woebot. A variety of features were assessed between AirHeart and the five competitors. It was discovered that while each app boasts unique features and specifications, such as mood tracking or a conversational element, AirHeart combines most of these features into one comprehensive platform. See Tables A8 & A9 in Appendix A for more information.

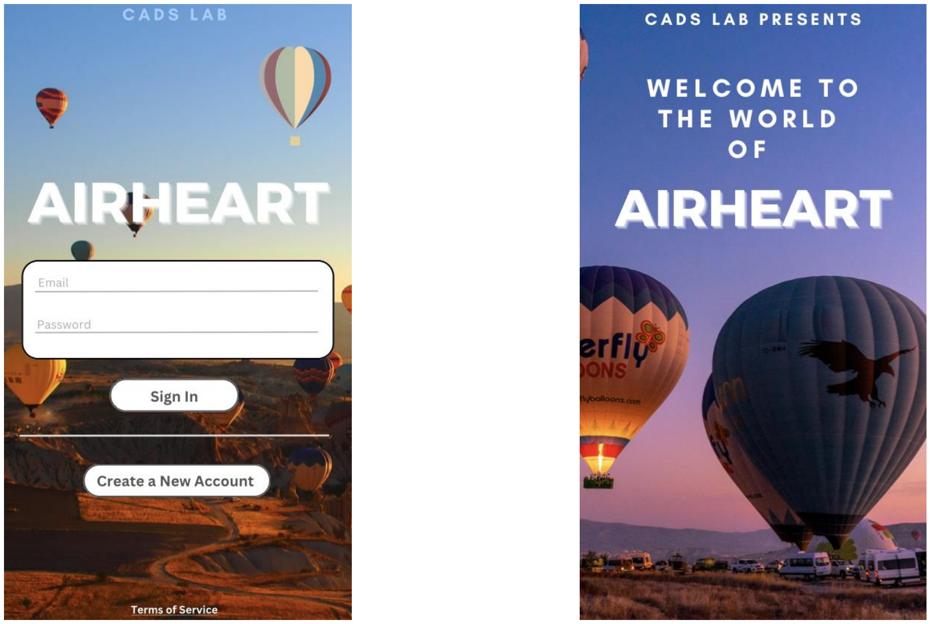
Major Prototype Changes

The results from the creation of user personas, journey maps, a competitor analysis, and feedback from the original AirHeart experiment (Six et al., 2022) were compiled to form a list of potential inclusions and updates to features in the previous version of AirHeart. A total of six alterations were chosen to be implemented in the AirHeart 2.0 prototype: 1) updated aesthetics, 2) improved usability, 3) additional virtual character customization options, 4) a redesign of the map home page, 5) new "Additional Resources" and "Help" sections, and 6) the addition of interactive and animated options for the CBT modules.

Updated Aesthetics: Individuals with symptoms of depression often use virtual environments for a sense of escapism (Li et al., 2011). Prior research also suggests that individuals with varying degrees of depressive symptoms would prefer a colorful and fun app design (Alqahtani, Winn, Orji, 2021). Therefore, the overall appearance of the AirHeart app was re-designed to represent a trip through the clouds with realistic images and bright colors to contrast the sullen topics relating to depression. These backgrounds were selected with the intent to induce feelings of wonder while maintaining a calming presence. This type of design was recommended by participants from the original AirHeart experiment (Six et al., 2022). Additionally, the competitor analysis revealed that the Sanvello app incorporated realistic calming images in its app design, and What's Up, Youper, and Woebot effortlessly included bright and saliant colors into their app design. See Figure 4.1 for an example of the new aesthetics within the AirHeart app.

Figure 4.1

AirHeart V4 Login Screen (Left) and AirHeart V4 Welcome Screen (Right)

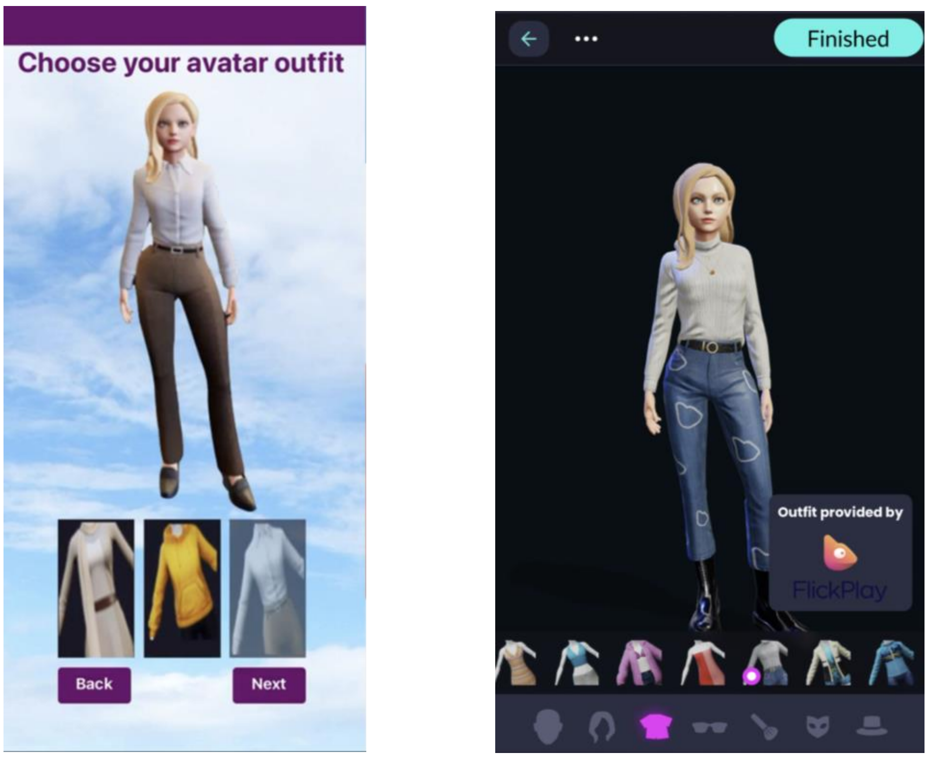


Improve Usability: During the testing phase for the previous iteration of the AirHeart app, users stated that the poor usability of the journal and the mood tracker caused frustration (Six et al., 2022). For the journal, the user's visualization was occasionally impaired by the keyboard when typing. As a remedy, the keyboard was re-designed to appear in front of the current journal entry and have the screen auto-scroll as the user types without pushing the text out of the screen. This allows the user to see their entry as they type.

Agent Customization Options: The first iteration of the AirHeart app included a limited selection of customization options (Six et al., 2022). Participants reported that this creative restriction negatively impacted the connection between the user and their virtual agent (Six et al., 2022). To remedy this negative feedback, the entire customization library from readyplayer.me was implemented into the new iteration of AirHeart. This included more starter agents, accessories (i.e. hats, makeup, facial hair, and glasses), clothing options, etc. See Figure 4.2 for a comparison of the agent customization feature from the first and updated versions of AirHeart.

Figure 4.2

AirHeart V1 (Left) & V4 (Right) Agent Customization



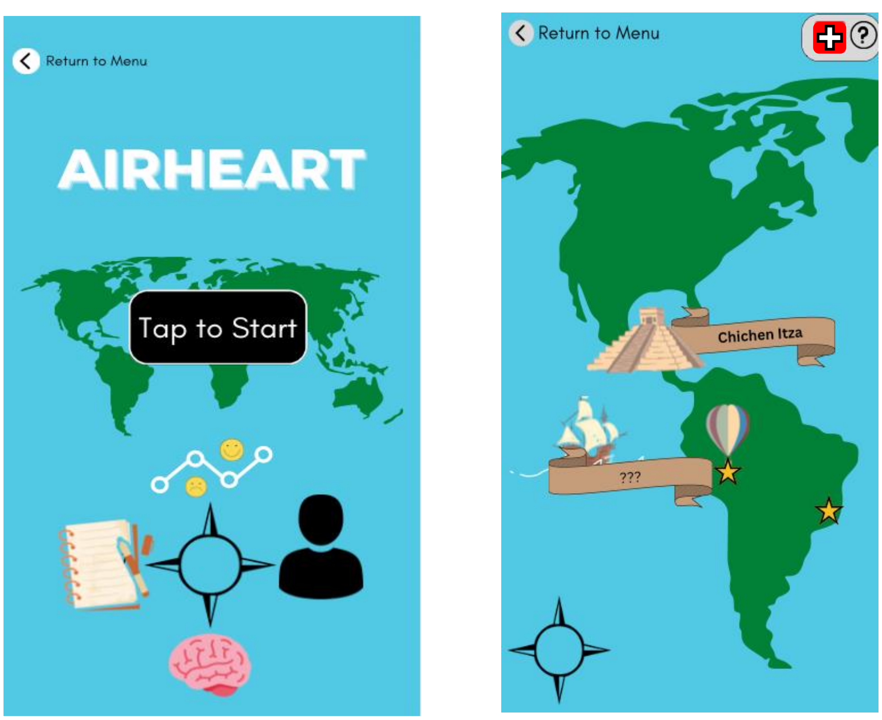
Redesign Map Home Page: The original map home page in the first iteration of AirHeart presented notable usability issues. The size of the stars used to indicate CBT modules were very small on phone screens. This required high levels of precision to access the modules, which was deemed frustrating. Additionally, the red and green coloring could present impaired visibility to color-blind individuals. Users also reported feeling trapped on the map home page, as there was no return to menu option.

Furthermore, the horizontal direction required users to rotate their phones but quickly return to vertical orientation for exercises embedded into the CBT modules. This consistent rotation was strongly disliked by users.

To address these comments, a vertical orientation was utilized throughout the application, and the sections of the map of the world were enlarged. The user can zoom out to view the full world map, but the enlarged version was standard. This allowed users the option to swipe to see different parts of the map. The size of the stars on the map was increased, and the concept of the stars changing colors to indicate completion was changed to show cartoon images of the wonder. This change was to prevent discrimination against specific groups of users. To eradicate feelings of confinement, a "return to menu" button was implemented in the top left-hand corner of the screen. Figure 4.3 shows a visualization of the new map home screen.

Figure 4.3

AirHeart 2.0 Redesigned Map Screen



Additional Resources & Help Section: The original AirHeart application did not include Additional Resources and Help sections (Six et al., 2022), however, both were recommended by participants. The Additional Resources section was designed to provide additional psychoeducation on four negative moods the user may experience: blue (depressed), isolated, overwhelmed (stressed), and panicked (anxious) as well as general mental health assistance and suicide prevention assistance. Selecting these options would take users to official pages provided by the World Health Organization (WHO), American Psychiatric Association (APA), and other reputable mental health organizations. As AirHeart is an MHealth app, the Additional Resources section was prioritized as establishing a place where users could easily and efficiently receive quality mental health information and assistance in a crisis.

The Help section included five different options: 1) frequently asked questions, 2) about AirHeart, 3) terms and conditions, 4) report an error, and 5) contact the AirHeart team. The "frequently asked questions" and "report an error" sections were highly requested as a means of self-diagnosing a problem or confusion without the need to reach out to a researcher. If the issue persisted, the "contact the AirHeart team" was provided in-app in case the participant lost the researcher's email. A few participants requested to see the terms and conditions for the AirHeart application, so that option was included. Lastly, numerous participants inquired as to the purpose of the AirHeart application, so an "about AirHeart" section was also provided.

Video-like Modules: To improve accessibility to a wider audience of users, a video-like option was created for the CBT modules. Some users may experience higher levels of fatigue, so this option may further entice them to complete the modules as opposed to longer blocks of text. While this option still includes text on the screen, this is paired with the virtual therapeutic coach speaking the text to the users.

User Studies

Once the new additions were added to the prototype, five users matching the target audience were recruited to complete ten different tasks which mirrored the daily use of the AirHeart app: 1) create an account, 2) create a virtual character, 3) complete the daily questionnaire, 4) view the weekly mood tracker, 5) create a new entry in their journal,

6) complete the first CBT module, 7) view their activities from the modules, 8) navigate to the SOS page, 9) contact the AirHeart team, and 10) report an error.

All five users successfully completed twelve of the thirteen tasks without mis-click errors, confusion, or the need for assistance; however, creating an account appeared misleading for most of the users. Three out of five users entered their email and password before clicking the “create an account” button. This button was meant to take them to a separate create an account page where they would then enter their name, email, and password of choice. The remaining two users asked whether to click the “create an account button” or to enter their information first.

To conclude the prototype testing, the users completed the System Usability Scale (SUS) to assess AirHeart’s usability and user satisfaction. The mean SUS score for this prototype iteration (*M* = 82.50, *SD* = 10.61) was significantly higher than the reported score found during the initial AirHeart experiment (*M* = 54.35, *SD* = 18.29) suggesting a large improvement; however, due to contrasting sample sizes and difference in duration of use for the app, this comparison should be interpreted lightly. Lastly, the users participated in a short post-experiment interview to ask about potential improvements.

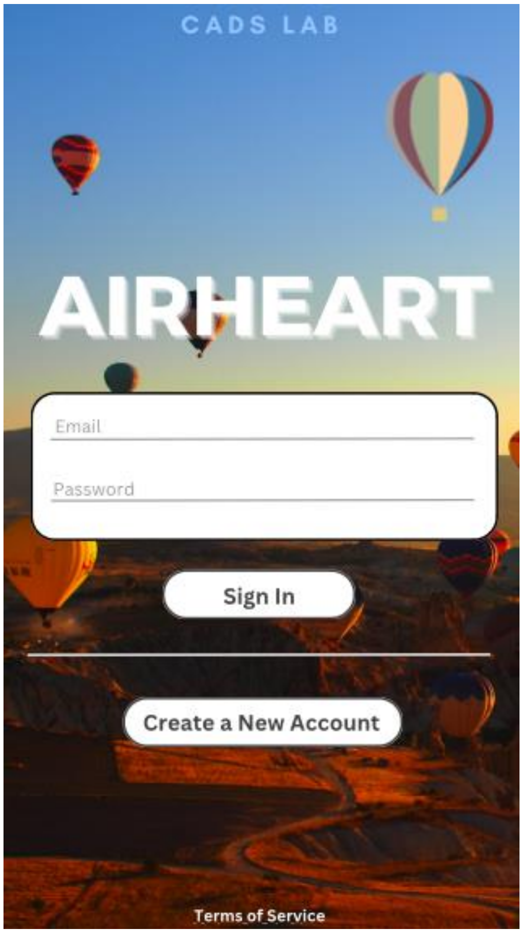
AirHeart Prototype Iteration #2

Two suggestions were collected from the five user studies: 1) redesign the login page and 2) create a mandatory agreement option to the terms of service.

Login Screen: Overall, users found AirHeart easy to use but tended to make mistakes when attempting to create an account. To prevent this error, the “create a new account” button was placed under a white vertical line away from the “sign in” button. This line was meant to indicate a separation from the email, password, and sign in options. The “create a new account” button was also enlarged to further differentiate it from the “sign in” button. Please see Figure 4.4 for a visualization.

Figure 4.4

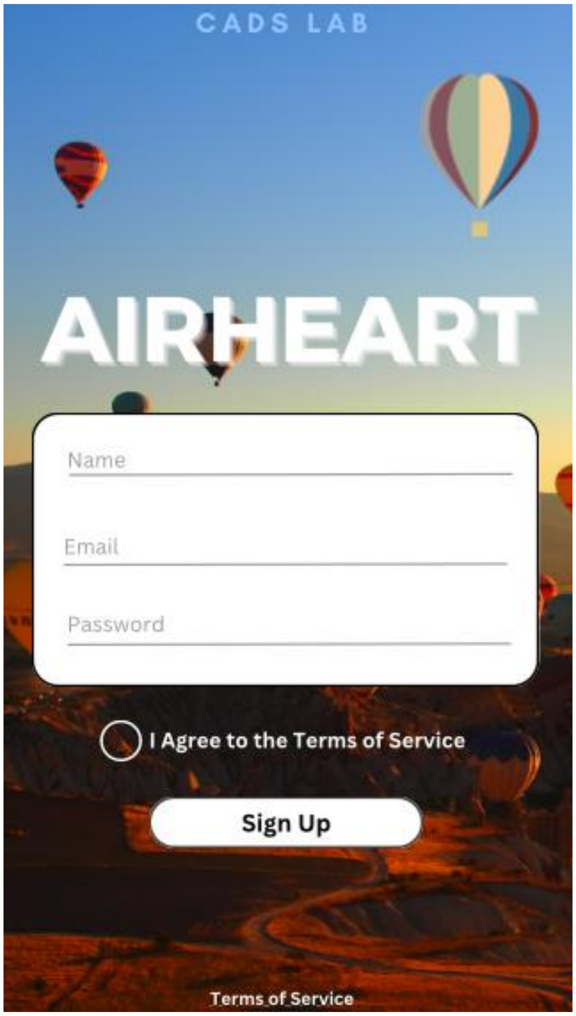
AirHeart 2.0 – New Login Screen



Terms of Service: It was suggested that the users be required to read and agree to the terms and conditions prior to being granted access to the AirHeart application. A separate page was created to allow the users to read through the terms and conditions with an agree button at the bottom of the page. When the users returned to the sign-up page, the "review the terms and conditions" option had a green check beside it to indicate this requirement is fulfilled. Figure 4.5 show the sign-up page with the requirement to select "I Agree to the Terms of Service".

Figure 4.5

AirHeart 2.0 – Account Sign Up Page with Terms of Service Agreement



Heuristic Evaluation

After implementing the changes from the feedback from the first AirHeart experiment (Six et al., 2022) and the user studies into the prototype, a heuristic evaluation was conducted utilizing Nielson's 10 Heuristics (Nielson, 1994). A total of five primary tasks were identified for evaluation: 1) create an account, 2) complete the daily questionnaire, 3) fill out the mood tracker, 4) write a journal entry, and 5) complete the first module; these tasks were chosen to mimic the daily use pattern a participant may experience. Virtual agent creation was not included in the evaluation, as the process was created by a third-party source and is not able to be altered. Each task received a ranking on a scale from 0 (no usability problem, not a priority) to 4 (completely unacceptable or unusable, top priority) (Nielson, 1994).

Two of the six tasks revealed a problem: two flaws with scores ranging from 0-3. Of these flaws, one was a cosmetic fix: making font sizes and button locations consistent in the first module. The remaining flaw was more serious. In the daily questionnaire, users could select the right or left arrows to continue or return to a previous question without answering the question. Users could also select more than one answer to a question. As the prototype was created utilizing the online web creation tool Figma, there were limited capabilities for logic. When the AirHeart application was transferred onto a smartphone application using Unity, this issue was corrected.

The AirHeart app went through considerable alterations to improve usability and user satisfaction. Prior to starting the redesign, user personas and a journey map were created to enter the mindset of the user. A competitor analysis was conducted to compare the current plans for the AirHeart redesign to five MHealth apps currently for sale or download on either the Apple App Store or Google Play Store. Based on the feedback provided by users from the original AirHeart experiment (Six et al, 2022), improvements were identified. A new iteration of AirHeart was created in Figma and given to five different users with the goal of completing thirteen different tasks within the app. Overall usability scored above 80% (M = 82.5), which was substantially higher than the previous edition of AirHeart (*M* = 54.35). Users provided three suggestions for further improvement. After these changes were implemented, a heuristic evaluation was performed to reaffirm improved usability. One minor, cosmetic issue and one more severe issue were identified and remedied for the final prototype. Overall, the AirHeart app underwent a total redesign to better deliver CBT and mental health assistance to a wide variety of users but maintained its original features. Additionally, the app gained two new features: help and additional resources sections.

Current Experiment

The current experiment utilizes the results and findings from the various user experience (UX) research methods completed and implements the findings into the newly developed AirHeart app created for smartphones using Unity. Additionally, it incorporates the findings from the previous two studies by implementing customization of virtual agents over progression (Experiment 1) and an active condition where participants can talk with their virtual therapeutic coaches (Experiment 2). The novel element under investigation for the present experiment is virtual agent animation.

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 (Parmar et al., 2022). While overly expressive facial animations were deemed unrealistic and cartoon-like, users reported natural facial expressions as signifying more authoritarian traits, like respect and competency, as well as comforting traits, like calmness and warmth (Hyde et al., 2016). Furthermore, animations have shown to elicit greater emotional responses and sense of co-presence when compared to static conditions (Wu et al., 2014). These positive traits combined, encourage natural animation to be used in collaboration with virtual agents, especially virtual therapeutic coaches. For depressive individuals who often experience negative perceptions of themselves, their situation, and the world around them (Beckham, 1986), these comforting traits could help to improve the virtual therapy experience.

This research endeavor will primarily investigate the effect of conversation and animation features of a virtual therapeutic coach. The previous experiment (Experiment 2) left multiple questions as the methods compared a conversational, animated agent to a non-conversational, non-animated agent. Thus, this final follow-up experiment will delve more in-depth as to the differences between these elements and their impacts on depressive symptoms, overall user experience, and perception of the agent.

Experimental Aims

Aim 1: The primary aim of this experiment was to investigate whether conversational vs. non-conversational and animated vs. non-animated virtual agent features within the AirHeart app can reduce symptoms of depression over two weeks.

Aim 2: The secondary aim of this experiment was to examine the effect of conversation and animation on a participant's experience with the app modules, activities, their agent, and app overall.

Aim 3: The tertiary aim of this experiment was to analyze user impressions regarding their customizable virtual agent and overall app usability.

Aim 4: The exploratory aim was to investigate whether the AirHeart app can reduce symptoms of anxiety, stress, and rumination over the course of two weeks.

Exploratory Aims

Aim 5: The secondary exploratory aim was to review participant's thematic responses to ascertain ways to improve the virtual therapeutic coach.

Hypotheses

Aim 1 Hypothesis:

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

Aim 2 Hypothesis:

H2: Participants with higher levels of depressive symptoms will show the strongest overall preference for the conversational condition compared to the non-conversational condition and the animated condition compared to the non-animated condition. This evaluation will be operationalized on a ranking scale from 1-10 for participants' experiences with the app overall, the virtual therapeutic coach, the modules, and activities.

Aim 3 Hypotheses:

H3: Participants will express higher levels of bonding and trust with their coach in animated and conversational conditions. Trust will be interpreted based on a rating system completed by the participants in the post survey. Bonding will be interpreted using the WAI bond subscale.

H4: The overall SUS score for the new version of AirHeart will be higher than the SUS score recorded for the previous version of AirHeart (Six et al., 2022).

Aim 4 Hypotheses:

H5: Individuals with higher levels of anxiety, stress, and rumination symptoms will show significantly lower symptoms of depression after two weeks. It is expected this reduction will be more pronounced in the conversational compared to non-conversation condition and in the animated compared to non-animated condition.

Exploratory Hypothesis:

In addition to the quantitative aims and hypotheses, the current experiment investigated prompts and suggestions from participants regarding ways to improve the virtual therapeutic coach. As this virtual coach has not been previously utilized by a large sample of participants, aside from alpha and beta testing, recommendations will be analyzed as a means of gaining insight and improving AirHeart for the next version.

METHOD

Participants

Prior to the experiment commencement, a G*power analysis was conducted to determine the number of participants needed to maintain an 80% power level to detect a small to medium effect at the level of significance (p = .05). A total of 128 participants were needed to maintain the desired power level for determining whether a reduction in depressive symptoms would occur overall. A total of 192 participants were needed to maintain the desired power level for determining whether a reduction in depressive symptoms would occur between the four conditions. An additional 88 participants were recruited to account for potential data loss or attention check failure. Participants were incentivized to participate in this experiment with compensation in the form of course credit, extra credit, or a $20 Amazon gift card.

Exclusion Criteria: Participants were excluded from participating in the experiment for four reasons: 1) the individual was under the age of 18 and classified as a minor or older than 30, 2) the individual was not fluent in English, 3) the user did not have daily access to an iPhone or Android phone, and 4) the user participated in the first AirHeart experiment (Six et al., 2022). During data cleaning, data was excluded for two reasons: 1) the participant did not complete the experiment, and 2) the participant failed more than one attention check.

Measures

Demographics and Usability Questionnaires

To start, a demographics questionnaire was used to collect information regarding age, gender, race, education, gaming, eyesight, and phone usage. The SUS was used to assess usability regarding the virtual therapeutic coach, the AirHeart app, and the overall experience.

Mental Health Related Questionnaires

Three different questionnaires will be utilized to assess changes in mental health between the two weeks: 1) the PHQ-8, 2) GAD-7, and 3) the Perceived Stress Scale-10 (PSS-10).

Perceived Stress Scale -10 (PSS-10): The PSS-10 is a subjective assessment of the user's stress symptoms during the past month (Cohen, Kamarck, and Mermelstein, 1994; Lee, 2012). The ten questions are assessed with a 5-point Likert scale; question topics include feeling out of control, confidence levels, and irritations (Cohen et al., 1994). This questionnaire demonstrates a high level of interval validity with Cronbach's alpha ranging from 0.60 to .91 (Lee, 2012).

Within the AirHeart application, the daily questionnaire will consist of three different questionnaires: 1) the Patient Health Questionnaire-4, 2) the Perceived Stress Scale-4, and 3) the Positive and Negative Affect Scale.

Patient Health Questionnaire-4 (PHQ-4): The PHQ-4 is a shortened version of the PHQ-9 used to assess subjective reports of depressive symptoms over the past 2 weeks (Khubchandani et al., 2016; Kroenke et al., 2009). Similar to the PHQ-9, this questionnaire utilizes a 4-point Likert scale ranging from "Not at all" to "Nearly every day" (Kroenke et al., 2009). Question topics include nervousness, feeling blue, and apathy (Kroenke et al., 2009). Scores range from 0-12 with higher scores indicating more severe depressive symptoms (Kroenke et al., 2009). This questionnaire has strong internal validity with a Cronbach's alpha of 0.85 (Kroenke et al., 2009).

The Perceived Stress Scale-4 (PSS-4): The PSS-4 is a shortened version of the PSS-10 and is used to quickly assess an individual's stress levels during the previous month (Wartting et al., 2013). Similar to the PSS-10, this shortened scale uses a 5-point Likert scale ranging from 0-4 (Cohen, Kamarck, & Mermelstein, 1983). Scores can vary from 0-16 with a higher score indicating higher levels of perceived stress (Cohen et al., 1983). This shortened questionnaire has a strong internal validity with a Cronbach alpha of 0.77 (Wartting et al., 2013).

Positive and Negative Affect Scale (PANAS): The PANAS is a subjective assessment of affect experienced during the previous week (Watson, Clark, Tellegen, 1988); positive affect can appear as interested, excited, confident, etc. while negative affect may manifest as guilty, upset, or irritable (Watson et al., 1988). This 20-item questionnaire is separated into two sections (positive and negative) and scaled from 1-5 with scores ranging from 10-50 for each section (Watson et al., 1988). This questionnaire retains a high internal validity with a Cronbach's alpha ranging from 0.84 to 0.90 (Watson et al., 1988).

Virtual Agent Questionnaires

Personality and the participant's perceptions were assessed with the two different questionnaires used in Experiment 2: 1) the WAI and 2) self-report impressions of the agent.

Virtual Agent Open-Ended Question

A question regarding how the users would improve the virtual agent was added at the end to help provide insight into the virtual agent questionnaires.

Materials

AirHeart V4

The newest version of the AirHeart app contains all themes and features of the prior version (Six et al., 2022) with new inclusions, such as a help section, additional customization options for the virtual agent, and an additional resources section for a total of nine features.

Virtual Therapeutic Coach: The virtual therapeutic coach joined the participants on their journey to help guide them through the various topics of CBT or related psychotherapies. A dialogue framework utilizing the RTVoice + AWS text-to-speech (TTS) engine was enabled to provide audio-based dialogue to the participants. The framework consisted of a custom dialogue object, which contained multiple lines of dialogue for the module scenario. A looping dialogue structure iterated through each line of text, which was converted into audio using the TTS engine. After the audio file finished playing, the loop continued onto the next line. This process continued until the end of the module's dialogue object. In the conversational agent condition, the SpeechRecognition speech-to-text (STT) engine was used to record the participant's response, then save the text to a local database. In this condition's dialogue object, line numbers that required user input were cued to "pause" the dialogue loop. At this point, the STT engine was turned on to record user audio. To effectively collect all user input, the dialogue loop waited three seconds after the user finished speaking before continuing to the next line of dialogue. Once the participant finished speaking, the agent would respond with one of five randomized backchanneling responses, such as “Okay” or “I see”.

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

The coach customization section utilized the online avatar creation service Ready Player Me. This free system allowed participants access to more customization options within the application, such as clothes, hairstyles, hair, skin, eye colors, body and face shape, etc. Participants were asked to customize an agent after the creation of their AirHeart account, but they could change their agent’s appearance at any time afterward within the app. The customizer could be accessed from the map home page for convenience. Examples of the customization page and potential agents can be found below as Figure 4.6.

Figure 4.6

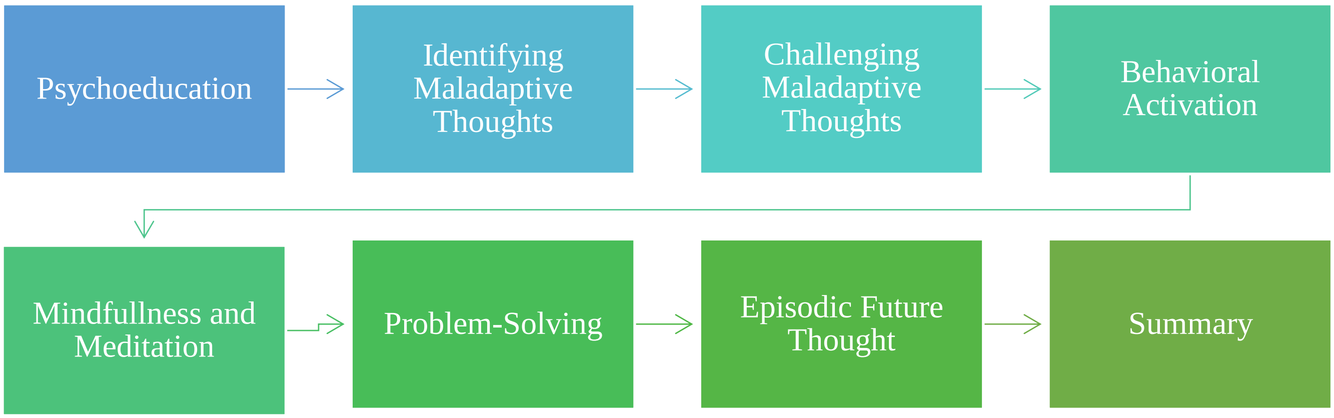
Example of the Virtual Therapeutic Coach Customization Process in AirHeart V4

CBT Modules: A total of eight CBT modules were included in the new version of the AirHeart application. The background displayed a realistic image of the specific location with the virtual therapeutic coach in front of the current wonder of the modern world. A text box with therapist's dialogue was located above their head, at the top. Topics of the modules are identical to the previous version of AirHeart (psychoeducation, identifying and combatting maladaptive thoughts, mindfulness and meditation, problem-solving, behavioral activation, and an overall summary) (Six et al., 2022) with the addition of the module on episodic future thought from Experiment 2. Similar to the previous longitudinal AirHeart experiment (Six et al., 2022), AirHeart 4.0 will contain the original 8 bCBT modules (psychoeducation, identifying and combatting maladaptive thoughts, mindfulness and meditation, problem-solving, behavioral activation, and an overall summary), with the addition of the module on episodic future thought from experiment 2 (US Department of Veteran Affairs, 2013). These modules were originally created using traditional CBT manuals and guidance books provided to therapists and clinicians (Ackerman 2021; Cully & Teten, 2008; Miner et al., 2016; Muñoz, Miranda, & Aguilar-Gaxiola, 2000).

The eight modules took place at the seven wonders of the modern world and on Bowman Field in Clemson, South Carolina. Participants were encouraged to use their new skills and techniques learned from the app during the experiment. The order of the modules can be visualized in Figure 4.7.

Figure 4.7

Order of CBT Modules within AirHeart V4



Daily Questionnaire: Similar to the previous version of AirHeart, users completed a thirteen-item questionnaire during their first login to the app for that specific day (Six et al., 2022). This questionnaire contains the PHQ-4, the PSS-4, and five positive items from the PANAS scale. The PANAS items were included as a means of introducing positivity into the questionnaire.

New Mood Tracker: The previous mood tracker utilized in the first AirHeart experiment (Six et al., 2022), displayed the user's mood over the course of the week through a line graph. Participant and observer feedback indicated that creating a feature which would potentially display the user's mood as decreasing over a period appeared disheartening. Thus, the line graph was altered to a simple message that would appear once the user completed the daily questionnaire. The three messages are below:

Mood Improved: It looks like your mood has improved from the last time you were here. I'm glad to see that you're having a better day today! Remember to take some time for yourself today to help your positive mood last all day.

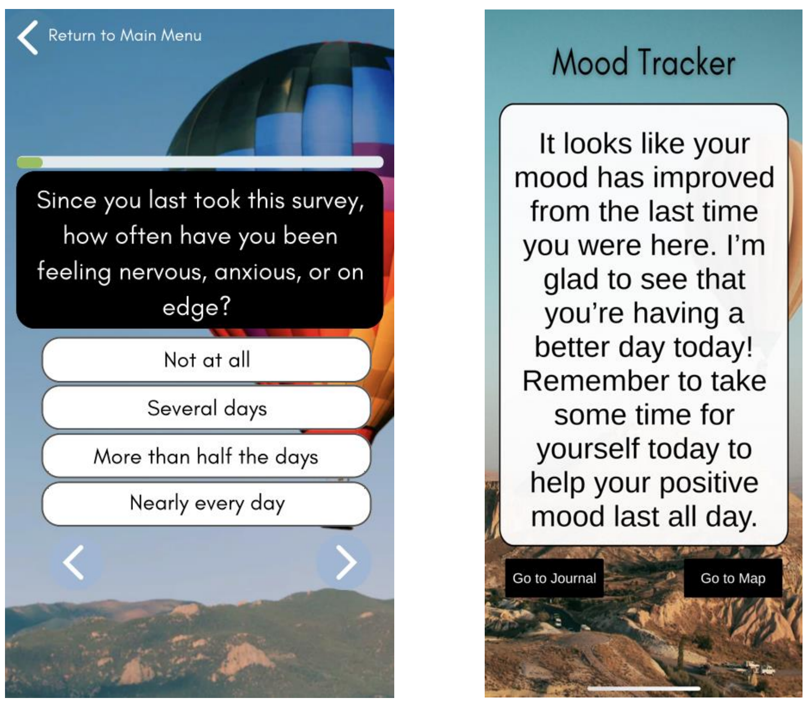
Mood Remained Similar to the Previous Day: It looks like your mood is around the same as the last time you were here. Try to do something nice for yourself today in addition to your modules to help further improve your mood.

Mood Decreased: It looks like you might not be feeling as great today. Remember to practice self-compassion and do something kind for yourself today.

See Figure 4.8 for examples of the daily questionnaire and mood tracker.

Figure 4.8

AirHeart V4 Daily Questionnaire Example (Left) & Mood Tracker (Right)



Journal: The virtual journal allowed users to write their thoughts about their day or the lessons and tips learned within the modules (Six et al., 2022). All journal entries were saved in chronological order from when they were created. The journal can be accessed from the map home page for convenience. An example of the journal page within AirHeart can be seen as Figure 4.9 below.

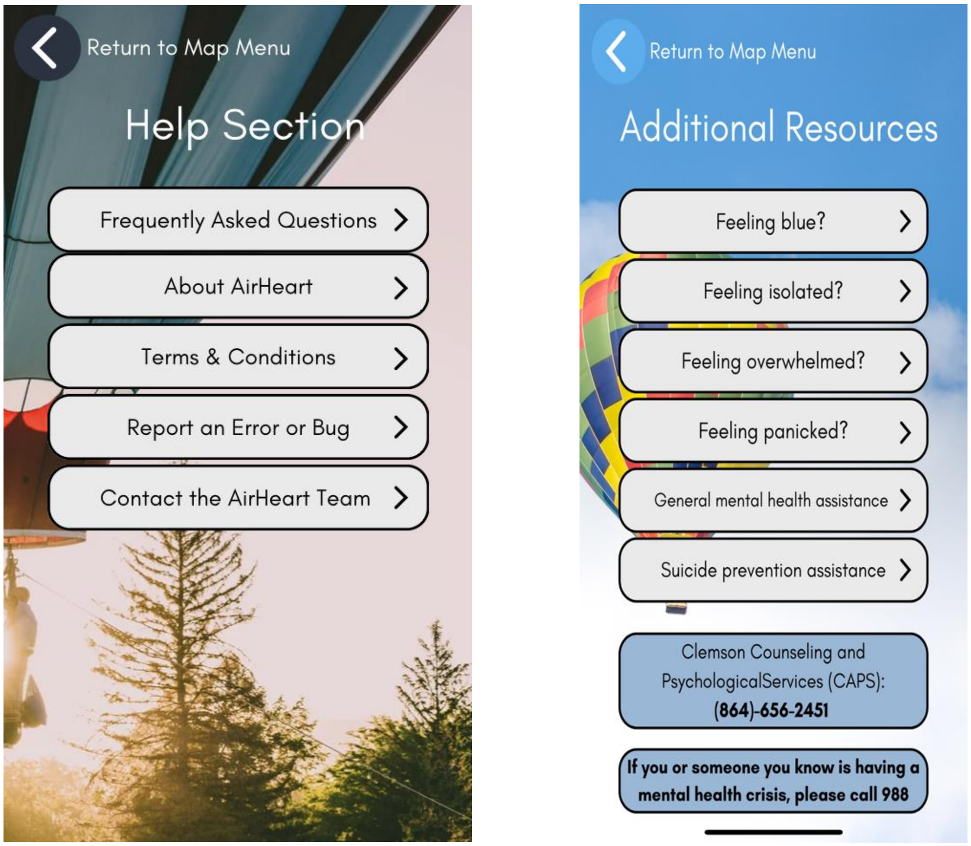
Figure 4.9

AirHeart V4 Journal

Help and Additional Resources Sections: The Help and Additional Resources sections are the two newest additions to the AirHeart app which encouraged users to ask questions, report bugs, talk to the team, and provided a place to seek mental health assistance or information if necessary. Examples of these two sections can be seen in Figure 4.10.

Figure 4.10

AirHeart V4 Help Section (Left) and Additional Resources Section (Right)



TestFlight

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

Experimental Conditions

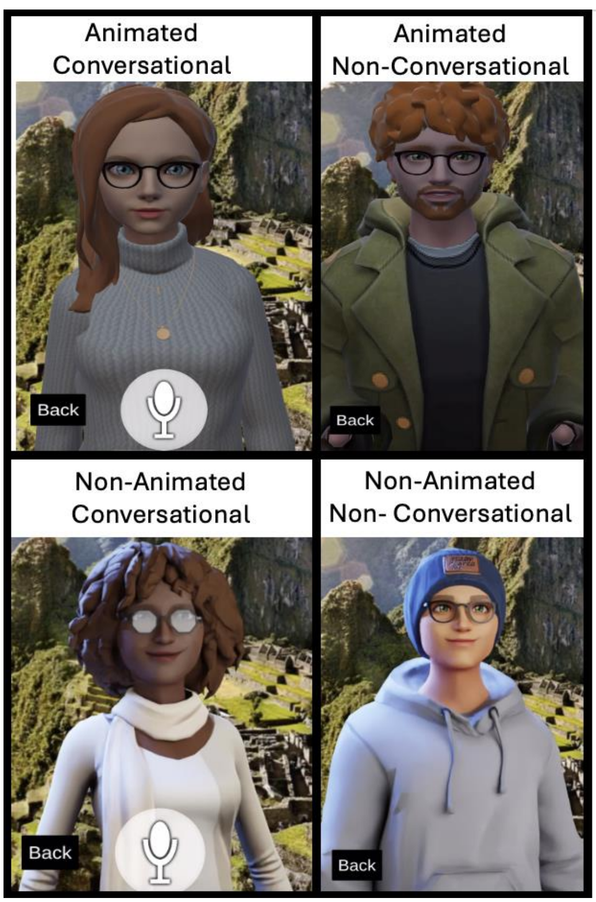
The current experiment included four different experimental conditions: 1) conversational animated virtual therapeutic coach, 2) non-conversational animated virtual therapeutic coach, 3) conversational non-animated virtual therapeutic coach, 4) non-conversational non-animated virtual therapeutic coach. The conditions differ regarding the level of animation of the virtual therapeutic coach (animated vs. non-animated) and the conversation between the virtual therapeutic coach and the participant (conversational vs. non-conversational). All conditions had access to all features of the AirHeart application (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 therapeutic coaching the CBT modules. The animated condition included human-like non-verbal body movements, mouth movements, and gestures in association with the information provided by the virtual therapeutic coach. For example, if the virtual therapeutic coach is informing the user about the process of the depressive spiral, the body language will be informative or educational. The facial expressions mimicked the tone of the information or conversation. For example, when the participant converses with the virtual therapeutic coach, they will have an interested facial expression. The non-animated condition displayed a static, non-moving virtual agent with a blank facial expression.

Conversational: The two levels of conversation within the experiment mimicked the active and passive conditions from the previous experiment. The active conversation included a question and response engagement spread throughout the educational portions of the CBT modules. The virtual therapeutic coach asked questions or instructed the participant to complete activities aloud. The passive, non-conversational condition, did not allow the user to add their input or respond to questions. See Figure 4.11 for a visualization of the virtual therapeutic coach in the four different conditions.

Figure 4.11

Virtual Therapeutic Coach in the Four Different Conditions in Module 1 - Psychoeducation



Experimental Design

For the current experiment, a 2 (Depressive Status: Depressive vs. Non-Depressive) x 2 (Conversation: Present vs. Absent) x 2 (Animation: Present vs. Absent) X 2 (Time: Pre vs. Post-Intervention) mixed experimental design was utilized for this experiment. The within element was the two-week period when the participants used the AirHeart app on their smartphones. Participants were randomized to one of the four conditions via an online randomizer.

Procedure

The procedure for this experiment is nearly identical to the longitudinal experiment performed using the first version of AirHeart (Six et al., 2022). To start, participants were randomized to one of the four conditions via a randomizer. When participants entered the laboratory, they were instructed to read and sign a consent form before starting the experiment. Once signed, the participants were reinformed of their rights as a participant and asked for verbal consent for the experiment. To start, individuals completed a pre-experimental survey consisting of demographics questions, and mental health-related questionnaires (the PHQ-8, the CES-D, the STAI, the GAD-7, the PSS-10, and the PANAS). Attention check questions were hidden throughout the survey to assess attention. Upon completion of the pre-experimental survey, individuals were instructed to download the newest version of AirHeart onto their iPhones via Test Flight or download it via Google Drive onto their Android phone. Users were instructed to create an account and complete the tutorial document consisting of creating a personalized virtual agent, answering the daily questionnaire, viewing their mood tracker statement, writing a short entry in their journal, and completing the first introductory CBT module. When the participant has finished the tutorial, they received the following instructions:

"You are being instructed to utilize the AirHeart app every other day at least once a day for the next two weeks. Today counts as day one, so two weeks from today, you will receive an email asking you to complete a follow-up survey. This survey is the last thing that you will need to complete. If you complete all your tasks, you will receive your class credits and Amazon gift card. You do not need to return to this lab after today. Do you have any questions?". After all the questions were answered, the participant left the lab.

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

End of Intervention Assessment: After fourteen days, the participant received a follow-up survey in their email inbox. This survey included mental health related questionnaires (PHQ-8, the CES-D, the GAD-7, the STAI, and the PSS-10), usability questionnaires (the SUS and the MAUQ), and questionnaires relating to the virtual therapeutic coach (the WAI, the CRQ, and self-report impressions of the agent). Attention check questions were hidden throughout the survey to assess attention. Once the user completed the survey, researchers looked through their data. If the participant completed all eight CBT modules and at least eight daily questionnaires, they would receive an email confirmation of the 10 SONA credits and Amazon gift card within five business days. If the user did not complete the seven daily questionnaires along with all eight CBT modules, they would not receive the Amazon gift card. Partial credit for SONA was granted corresponding to the number of CBT modules completed.

Data Analysis

The subsections below outline the five hypotheses with corresponding aims and statistical tests.

Aim 1 Hypotheses:

To investigate H1, a 2 (Animation Status) X 2 (Conversational Status) X 2 (Depressive vs. Non-Depressive Symptoms) mixed effects ANOVA was used to analyze pre-post depressive symptoms as the within-subjects factor and app animation status and app conversational status as the between-subjects factors.

Aim 2 Hypotheses:

For H2, two separate 2 (Animation Status) X 2 (Conversational Status) X 2 (Depressive vs. Non-Depressive Symptoms) ANOVAs were used to investigate participant's perceived experiences with their virtual therapeutic coach and the AirHeart app overall.

Aim 3 Hypotheses:

For H3, two separate 2 (Animation Status) X 2 (Conversational Status) ANOVAs for each outcome variable (bonding and trust) were conducted.

For H4, an independent samples t-test was used to compare the overall SUS score for the new version of AirHeart with the previous version (Six et al., 2022).

Aim 4 Hypotheses:

Similar to H1, to investigate H5, separate 2 (Animation Status) X 2 (Conversational Status) X 2 (Time: Pre vs. Post) mixed effects ANOVAs were employed for change in anxiety, stress, and rumination symptoms.

Exploratory Hypothesis:

A thematic analysis was performed (Braun & Clarke, 2006) commencing with the de-identification of the data to avoid unintentional bias; names, emails, ages, participant IDs, condition, and depressive group were removed. One researcher reviewed the responses and created categories using axial coding. Depressive group (Depressive vs. Non-Depressive) and condition (Animated vs. Non-Animated & Conversational vs. Non-Conversational) coding were re-attached to the responses to create a frequency data table.

RESULTS

A total of 280 participants were recruited and began the AirHeart testing process. From there, 71 participants were excluded (43 for failing to complete the post-survey, 14 for completing fewer than 3 modules, 11 for failing two or more attention checks, two for exceeding the pre-determined age range, and one for failing to complete the pre-survey); for a visualization of this information, please see Figure 4.8. The final sample size contained 209 participants between the four conditions ($M_{age} = 19.97$, $SD_{age} = 2.19$). Additional details regarding the participant sample are displayed in Table 4.1.

Table 4.1

Demographic Information for Experiment 3 by Depressive Group and Condition

AirHeart 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 Bi-Polar II: 4/209 Eating Disorder: 2/209 Adjustment Disorder: 1/209 Trichotillomania: 1/209 Mood Disorder: 1/209

Non-Depressive Group		Depressive Group	
$M = 2.15$	$SD = 1.34$	$M = 9.29$	$SD = 3.91$
$M_{age} = 20.24$	$SD_{age} = 2.49$	$M_{age} = 19.84$	$SD_{age} = 2.03$
Diagnosis: 16/86	No Diagnosis = 70/86	Diagnosis: 49/123	No Diagnosis = 74/123

Condition	Gender	Age	Diagnosis	Depressive or Non-Depressive Group
Conversation – Animated (Condition 1) N = 52	M = 11 F = 40 NB = 1	$M = 20.35$ $SD = 2.07$	Yes: 18 No: 34	D: 31 ND: 21
Non-Conversation – Animated (Condition 2) N = 54	M = 8 F = 45 NB = 1	$M = 20.93$ $SD = 3.00$	Yes: 18 No: 36	D: 28 ND: 26
Conversation – Non-Animated (Condition 3) N = 53	M = 10 F = 42 NB = 0	$M = 19.36$ $SD = 1.47$	Yes:16 No: 37	D: 36 ND: 17
Non-Conversation – Non-Animated (Condition 4) N = 50	M = 10 F = 40 NB = 0	$M = 19.22$ $SD = 1.49$	Yes: 13 No: 37	D: 28 ND: 22

Hypothesis 1 Results:

For H1, a 2 (Conversational vs. Non-Conversational) X 2 (Animated vs. Non-Animated) X 2 (Pre vs. Post) mixed ANOVA was conducted to explore the change in depressive symptoms before and after the two-week period for the different interaction designs. A statistically significant difference was discovered for the 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. No significant effects were ascertained for the main effects of animation condition ($F(1, 208) = .02$, $p = .91$; $\eta p^2 < .001$), the conversational condition (F (1, 208) =.25, $p = .62$, $\eta p^2 = .001$), nor any of the interaction effect (*p*s > .05). Thus, H1 is partially supported. See Table 4.2 for more information.

Table 4.2

Visualization of Main and Interaction Effects for Hypothesis 1 – Depressive Symptoms

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	PRE: 6.32 POST: 5.60	Animated	PRE: 4.84 POST: 5.00
Non-Animated *N* = 102	PRE: 6.38 POST: 5.40	Non-Animated	PRE: 4.59 POST: 4.73
Conversational *N* = 105	PRE: 6.49 POST: 5.67	Conversational	PRE: 4.53 POST: 4.97
Non-Conversational *N* = 104	PRE: 6.21 POST: 5.50	Non-Conversational	PRE: 4.90 POST: 4.86

Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp^2)**
**Time Main Effect*	*10.60*	*.002*	*.05*
Animated Main Effect	.015	.91	<.001
Conversational Main Effect	.25	.62	.001
Animated X Conversational Interaction Effect	.037	.85	<.001

Time X Animated Interaction Effect	.24	.62	.001
Time X Conversational Interaction Effect	.02	.89	< .001
Time X Animated X Conversational Interaction Effect	.76	.38	.004

*Significant using alpha = 0.05

Hypothesis 2 Results:

For H2, two 2 (Conversational vs. Non-Conversational) X 2 (Animated vs. Non-Animated) X 2 (Depressive vs. Non-Depressive) ANOVAs were conducted to assess participant's perceived experiences of the follow attributes on their assigned condition: overall experience with the AirHeart app and experience with the virtual therapeutic coach. For overall experience with the application, no statistically significant findings were discovered for either main effects of animation condition (F (1, 208) = 2.68, p = .10, ηp^2 = .013), conversational condition (F (1, 208) = .59, p = .44, ηp^2 = .003), or depressive group (F (1, 208) = .33, p = .57, ηp^2 = .002). The interaction effects also failed to provide a significant finding (ps > .05). See Table 4.3 for more information.

Table 4.3

Visualization of Main and Interaction Effects for Hypothesis 2 – Overall Experience with the AirHeart App

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	6.21	Animated	2.51
Non-Animated *N* = 102	5.77	Non-Animated	2.52
Conversational *N* = 105	6.10	Conversational	2.54
Non-Conversational	5.89	Non-Conversational	2.50

N = 104			
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp²)**
Animated Main Effect	2.68	.10	.013
Conversational Main Effect	.59	.44	.003
Animated X Conversational Interaction Effect	1.99	.16	.010

For experience with the virtual therapeutic coach, no statistically significant findings were discovered for either main effects of animation condition (F (1, 208) = .01, p = .91, $\eta p^2 < .001$), conversational condition (F (1, 208) = 1.30, p = .26, ηp^2 = .006), or depressive group (F (1, 208) = 3.50, p = .062, ηp^2 = .017). The interaction effects were also non-significant (*p*s > .05). Thus, H2 is not supported. See Table 4.4 for more further information.

Table 4.4

Visualization of Main and Interaction Effects for Hypothesis 2 – Experience with the Virtual Therapeutic Coach

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	5.07	Animated	3.06
Non-Animated *N* = 102	5.27	Non-Animated	3.02
Conversational *N* = 105	5.39	Conversational	3.15
Non-Conversational *N* = 104	4.94	Non-Conversational	3.03
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp²)**
Animated Main Effect	.01	.91	<.001
Conversational Main Effect	1.30	.26	.006
Animated X Conversational Interaction Effect	3.33	.07	.016

Hypothesis 3 Results

For H3, separate 2 (Conversational vs. Non-Conversational) X 2 (Animated vs. Non-Animated) ANOVAs were performed for the dependent variables bonding and trust in the agent. The main effect of animation condition demonstrated a significant finding ($F(1, 208) = 4.31$, $p = .039$, $\eta p^2 = .02$) with the animated conditions experiencing a lower average bond with the virtual therapeutic coach ($M = 15.52$, $SD = 7.52$) when compared to the non-animated conditions ($M = 17.59$, $SD = 6.60$). No statistically significant finding was discovered for conversational condition ($F(1, 208) = 1.72$, $p = .19$, $\eta p^2 = .008$) nor the interaction effect ($F(1, 208) = 2.07$, $p = .15$, $\eta p^2 = .01$). See Table 4.5 reports the detailed results.

Table 4.5

Visualization of Main Effects and Interaction Effects of Hypothesis 3 – Bonding with the Virtual Therapeutic Coach

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	15.52	Animated	7.53
Non-Animated *N* = 102	17.59	Non-Animated	6.60
Conversational *N* = 105	17.22	Conversational	6.72
Non-Conversational *N* = 104	15.84	Non-Conversational	7.53
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared** (ηp^2)
**Animated Main Effect*	*4.31*	*.039*	*.001*
Conversational Main Effect	1.71	.19	.008
Animated X Conversational Interaction Effect	2.07	.15	.010

*Significant using alpha = 0.05

Secondly, ANOVA results for trust in the virtual therapeutic coach did not show a statistically significant finding for animation condition (F (1, 208) = 1.36, p = .25, ηp^2 = .007), conversational condition (F (1, 208) = .021, p = .88, ηp^2 < .001), or the interaction (F (1, 208) = .073, p = .79, ηp^2 < .001). Therefore, H3 is not supported. See Table 4.6 for more information.

Table 4.6

Visualization of Main and Interaction Effects for Hypothesis 3 – Trust of a Virtual Therapeutic Coach

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	2.72	Animated	1.28
Non-Animated *N* = 102	2.92	Non-Animated	1.22
Conversational *N* = 105	2.81	Conversational	1.19
Non-Conversational *N* = 104	2.83	Non-Conversational	1.32

Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp^2)**
Animated Main Effect	1.36	.25	.007
Conversational Main Effect	.02	.88	< .001
Animated X Conversational Interaction Effect	.07	.79	< .001

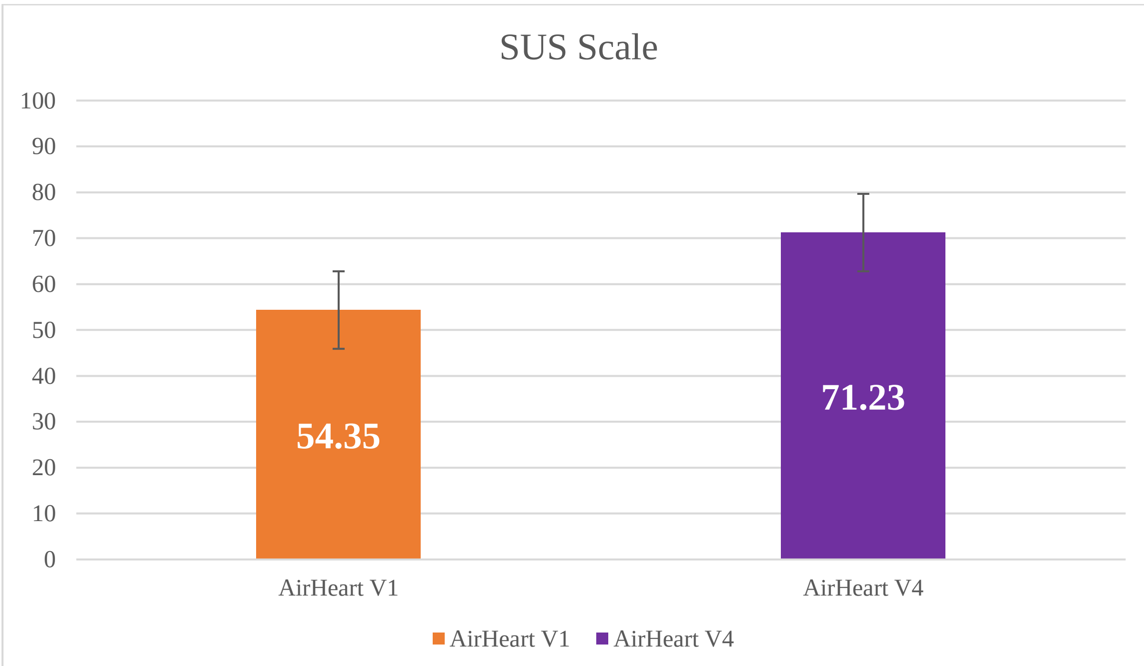
Hypothesis 4 Results:

For H4, Welch's t-test was used, as the difference in sample sizes between the two studies (N_1 = 83, N_2 = 209) violated the equality of variances assumption (p < .001). Results demonstrated a significant difference between the SUS scores from Experiment 1 (M = 54.35, SD = 18.29) and the current experiment (M = 71.23, SD = 13.36); t (118.33)

= -7.64, $p < .001$, Cohen's $d = 1.13$. Thus, H4 is supported. See Figure 4.12 for a visualization of the two SUS scores.

Figure 4.12

A Comparison of AirHeart V1 & AirHeart V4's (Current) System Usability Scores



Hypothesis 5 Results:

Similar to H1, H5 utilized 2 (Conversational vs. Non-Conversational) X 2 (Animated vs. Non-Animated) X 2 (Pre vs. Post) mixed ANOVAs to investigate the change in symptoms of anxiety, stress, and rumination between the animated and conversational conditions. For anxiety, a statistically significant difference was discovered for the main effect of time ($F(1, 205) = 9.43$, $p = .002$; $\eta p^2 = .044$) with the post-intervention anxiety scores reporting lower values ($M = 5.55$, $SD = 4.73$) than during pre-intervention ($M = 6.42$, $SD = 5.24$) across all four conditions. No significant effects were ascertained for the main effect of animated condition ($F(1, 208) = .045$, p =

.83; $\eta p^2 < .001$) or the conversational condition ($F(1, 208) = .017$, $p = .90$; $\eta p^2 < .001$). The interaction effects also failed to demonstrate a significant effect (*ps* > .05). See Table 4.7 for more information.

Table 4.7

Visualization of Main and Interaction Effects for Hypothesis 5 – Anxiety

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	PRE: 6.41 POST: 5.69	Animated	PRE: 5.42 POST: 4.87
Non-Animated *N* = 102	PRE: 6.43 POST: 5.39	Non-Animated	PRE: 5.06 POST: 4.60
Conversational *N* = 105	PRE: 6.17 POST: 5.72	Conversational	PRE: 4.72 POST: 4.74
Non-Conversational *N* = 104	PRE: 6.67 POST: 5.36	Non-Conversational	PRE: 5.27 POST: 4.73
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared (ηp^2)**
**Time Main Effect*	*9.43*	*.002*	*.044*
Animated Main Effect	.05	.83	<.001
Conversational Main Effect	.02	.90	< .001
Animated X Conversational Interaction Effect	.85	.36	.004
Time X Animated Interaction Effect	.37	.54	.002
Time X Conversational Interaction Effect	2.32	.13	.011
Time X Animated X Conversational Interaction Effect	.43	.51	.002

*Significant using alpha = 0.05

For stress, a statistically significant difference was discovered for the main effect of time ($F(1, 205) = 8.09$, $p = .005$; $\eta p^2 = .038$). This indicates that symptoms of stress were lower at two-week follow-up ($M = 15.91$, $SD = 7.67$) compared to 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; ηp^2 = .001), and all interaction effects , (*ps* >.05) were non-significant. See Table 4.8 for more information.

Table 4.8

Visualization of Main and Interaction Effects for Hypothesis 5 – Stress

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	PRE: 16.78 POST: 16.09	Animated	PRE: 6.50 POST: 7.75
Non-Animated *N* = 102	PRE: 17.27 POST: 15.73	Non-Animated	PRE: 7.14 POST: 7.61
Conversational *N* = 105	PRE: 16.97 POST: 15.67	Conversational	PRE: 6.62 POST: 7.72
Non-Conversational *N* = 104	PRE: 17.07 POST: 16.16	Non-Conversational	PRE: 7.03 POST: 7.64
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared** (ηp^2)
**Time Main Effect*	*8.09*	*.005*	*.038*
Animated Main Effect	.007	.92	<.001
Conversational Main Effect	.11	.74	.001
Animated X Conversational Interaction Effect	.38	.54	.002
Time X Animated Interaction Effect	1.20	.28	.006
Time X Conversational Interaction Effect	.24	.62	.001
Time X Animated X Conversational Interaction Effect	.26	.61	.001

*Significant using alpha = 0.05

Lastly, the post 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) for rumination ($F(1, 205) = 4.88$, *p* = .028; ηp^2 = .023) for rumination 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 effects also failed to demonstrate a significant effect (*ps* >.05). See Table 4.9 for more information. Thus, H5 is partially supported.

Table 4.9

Visualization of Main and Interaction Effects for Hypothesis 5 – Rumination

Means (*M*)		**Standard Deviation (*SD*)**	
Animated *N* = 107	PRE: 41.48 POST: 40.33	Animated	PRE: 13.25 POST: 13.19
Non-Animated *N* = 102	PRE: 42.40 POST: 40.51	Non-Animated	PRE: 14.04 POST: 12.79
Conversational *N* = 105	PRE: 42.22 POST: 41.21	Conversational	PRE: 13.51 POST: 12.34
Non-Conversational *N* = 104	PRE: 41.63 POST: 39.63	Non-Conversational	PRE: 13.78 POST: 12.97
Effect	**F-Value**	**p-value (*p*)**	**Partial Eta Squared** (ηp^2)
**Time Main Effect*	*4.88*	*.028*	*.023*
Animated Main Effect	.09	.76	<.001
Conversational Main Effect	.37	.55	.002
Animated X Conversational Interaction Effect	.39	.54	.002
Time X Animated Interaction Effect	.34	.56	.002
Time X Conversational Interaction Effect	.52	.47	.003
Time X Animated X Conversational Interaction Effect	2.45	.12	.012

*Significant using alpha = 0.05

Exploratory Results

Qualitative data was collected at the end of the experiment to ascertain potential areas of improvement for the virtual therapeutic coach. A thematic analysis was performed (Braun & Clarke, 2006); responses were collected from all 209 participants, but 41 of them failed to provide a viable answer. Utilizing axial coding, the 168 responses were sorted into five different categories: 1) Robotic Voice/Interaction, 2) More Engagement/Realism, 3) Lack of Personalization/Customization, 4) Technical Issues, and 5) General Suggestions. A frequency data table was created to help visualize this information (see Table 4.10).

Firstly, the majority of participants suggested changing the robotic voice or interactions to a more natural, human-like voice (57.7%). Multiple participants stated that the voice was too artificial or fake, which made it difficult to form a connection. Secondly, some participants stated that they wanted more engagement or realistic interactions with the virtual coach (28.6%). This suggestion came partly from individuals in the non-animated who wanted a virtual coach would use natural non-verbal body language, and individuals in the non-conversational conditions who suggested conversation as a means of improving engagement.

Thirdly, participants cited a lack of personalization and customization to their needs as an area for improvement (20.8%). This suggestion primarily included the desire for participants to have personalized responses from the virtual coach to what they had said. Participants felt that the responses the virtual coach provided were either too standardized or felt ingenuine to the personal information which was shared. Fourth, some participants experienced technical issues within the application with the virtual coach (7.7%). These issues included the virtual agent cutting off the participant when they were speaking, the application crashing, and the virtual coach's animations glitching.

Lastly, some participants provided general suggestions, such as wanting additional questions and activities, higher resolution for the coach, slowing down the speed of the modules, and providing information in less academic and easier to comprehend manner (9.5%). For visualization of this information and corresponding quotes, please see Table 4.10.

Table 4.10

Visualization of Qualitative Data for Experiment 3: Suggestions for Virtual Therapeutic Coach Improvement

Virtual Therapeutic Coach Improvement: *N* = 168					
Themes	**#s by Condition (4)***	**#s by Condition (Animation & Conversation)**	**% of Sample ****	**Examples**	**% of Depressive & Non-Depressive Samples*****
Robotic Voice/ Interaction	1: 22/44 2: 23/39 3: 25/46 4: 28/40	Animated: 45/83 Non-Animated: 53/86 Conversation: 47/90 Non-Conversation 51/79	57.7%	"Make it less robotic. The voice felt very fake to me." "I would make her less robotic, especially her voice. Her speaking sounds choppy, and I would turn it off because I did not like [it]."	D: 60.8% ND: 53.7%
More Engagement/ Realism	1: 9/44 2: 12/39 3: 18/46 4: 9/40	Animated: 21/83 Non-Animated: 27/86 Conversation: 27/90 Non-Conversation 21/79	28.6%	"I think making the movements and responses as human like as possible and adding different notes of emotion in there can be helpful." "It didn't really feel like we were having a conversation or that she was listening to my responses."	D: 28.4% ND: 25.4%
Lack of Personalization / Customization	1: 12/44 2: 10/39 3: 9/46 4: 3/40	Animated: 22/83 Non-Animated: 12/86 Conversation: 21/90 Non-Conversation 13/79	20.8%	"I think she could talk more personalized to you, seems like a robot with her responses." "I would suggest adding in a real conversation…having the therapist respond to your responses rather than following a script."	D: 18.6% ND: 23.9%
General Suggestions	1: 3/44 2: 7/39 3: 1/46 4: 5/40	Animated: 10/83 Non-Animated: 6/86 Conversation: 4/90 Non-Conversation 12/79	9.5%	"Add more activities that people could do on their own to practice." "Perhaps give more autonomy for user to pause, speed up, rewind, etc. dialogue of therapist."	D: 13.7% ND: 4.5%
Technical Issues	1: 6/44 2: 0/39 3: 6/46 4: 0/40	Animated: 6/83 Non-Animated: 6/86 Conversation: 12/90 Non-Conversation 0/79	7.7%	"Whenever I had to respond to the therapist, he often cut me off before I finished talking to him." "They had some trouble catching what I was saying to some of the prompts."	D: 5.9% ND: 9.0%

*The sample sizes for each condition exclude the participants who did not share a response.
** The sample used for column 3, "% of the Sample", is *N* = 168
***The samples used in column 5, "% of Depressive & Non-Depressive Samples", are the depressive and non-depressive samples (respectively) separately.

DISCUSSION

The current experiment sought to investigate whether conversational and animated components of an ECA within a smartphone MHealth app would affect the reduction of symptoms of depression, anxiety, stress, and rumination, along with usability, and participant preference (trust, bonding, and experience with the agent and AirHeart app). The results from the experiment demonstrated that with the lessons, techniques, and interactions with the CBT-based AirHeart application, participants showed a significant reduction in depressive symptoms overall over the two-week span, thus partially supporting H1. This result replicates the finding from the original AirHeart experiment in which a two-week interaction with the CBT-based app led to a significant reduction in symptoms averaging across all participants (Six et al., 2022). Additionally, this same finding was repeated regarding anxiety, stress, and rumination; the levels of negative symptomology reduced over the two-week period, thus partially supporting H5. This provides support that bCBT in a smartphone application format has potential for reducing symptoms of depression, anxiety, stress, and rumination in a short amount of time (US Department of Veteran Affairs, 2013). These findings provide support to other mental health applications which use CBT and have found similar results, such as MoodMission & MoodKit (Bakker et al., 2018), iCouch (Dahne et al., 2019), Pacifica (Moberg et al., 2019), and SuperBetter (Roepke et al., 2015). Future studies should include a clinical population in addition to a non-clinical population to determine if the results remain the same. Additionally, a longitudinal experiment where assessments are taken at additional monthly intervals should be obtained to ascertain whether the effects of the two-week intervention are sustained over time.

Secondly, this experiment was set to investigate the effect of animation and conversation features of an app-based virtual therapeutic coach on change in depressive symptoms (H1) the overall experience with the app and coach (H2), bonding and trust (H3), and anxiety, stress, and rumination (H5). While it was predicted that the conversational and animated conditions would further reduce symptoms of depression, anxiety, stress, and rumination over the non-conversational and non-animated conditions, this was not supported by the current findings. Furthermore, no significant difference could be found between these conditions and overall virtual therapeutic coach and app experience and trust with the agent, thus failing to support H2 and H3. The failure to distinguish a difference between the conditions and these preferences could potentially indicate the presence of a floor effect. This floor effect could be due to the themes outlined in the exploratory qualitative results: robotic voice/interaction, lack of engagement/realism, lack of personalization/customization, & technical issues.

For the robotic voice or interactions, a large majority of participants cited this as a necessary change to be made to the AirHeart application. This negative interaction could have produced an auditory uncanny valley effect where users did not want to engage with the virtual coach, thus reducing the levels of trust, desire to interact with the virtual coach, and the overall experience with the AirHeart app. Prior research on the perception of voices, both human-like and synthetic, has shown that synthetic, artificial voices induce an eerie feeling for participants (Abdulrahman & Richards, 2022). Specifically, low-quality text to speech (TTS) resulted in lower levels of trust when compared to a recorded human voice (Schroeder, Chiou, & Craig, 2021). Prior research using a TTS conversational agent (CA) as a digital coach to help promote emotional regulation skills using CBT demonstrated a similar finding with some participants reporting that the speech felt robotic (Hopman, Richards, & Norberg, 2023). With these negative effects observed from a synthetic voice, individuals may be less encouraged to listen, learn from, and seek assistance from a virtual coach utilizing TTS. Future research should replicate this experiment utilizing a higher quality TTS or pre-recorded human voice to determine whether the perceived floor effect could be eliminated.

Another suggestion from participants' qualitative responses includes making the virtual coach more engaging and realistic in appearance and behavior. While a high priority was placed on creating a realistic agent with non-verbal behaviors, it was suggested that movements be more refined and fluid to mimic a natural interaction. Furthermore, participants also suggested adding more personalization into the modules. Many participants felt that the responses given by the virtual coach were too standardized and felt ingenuine. The conversational element was defined as "forced" or "unnatural" by participants in the conversational condition. The interactions with the virtual coach should be personalized to the individual based on their current responses by adding a more dynamic conversational component to improve perceptions of realism, trust, and bonding. Instead of using a script-based dialogue, an artificial intelligence (AI) element could be included to adapt the conversation based on the dialogue provided by the participant. This personalization could help improve levels of connection and allow the participant to feel heard and understood by the agent. Prior research conducted on comparing adaptive versus static scripting in an educational setting demonstrated that participants perceived agents with adaptive scripting as more competent (Radkowitsch et al., 2021). Additionally, an experiment utilizing a dynamic interface that selected the appropriate, personalized response to a participant's dialogue demonstrated high levels of acceptance by participants (Bresó et al., 2016).

Specific technical difficulties, such as the virtual coach cutting off the participant when they were speaking, auditory malfunctions, and lagging Wi-Fi causing participants to repeat modules could have negatively impacted their perception of the virtual coach. Specifically, this could have negatively influenced their overall perception of and trust in the coach as well as the AirHeart app overall. Future versions of the AirHeart application should work to correct these mistakes paired with beta testing prior to the next full release.

A statistically significant finding was discovered for the effect of animated condition on the bond between the participant and their virtual coach (H2). While there was no significance between the conversational conditions, a significant difference between the animated and non-animated conditions was discovered for the bond they developed with their virtual coach. The non-animated condition experienced higher levels of bonding than the animated group, which did not support the corresponding hypothesis (H3). This suggests that individuals felt more connected with their virtual coach when they were viewed as a stationary figure without non-verbal body movements and facial expressions. These findings are not in line with work showing that animation is superior to a static image in a therapeutic setting (Bickmore & Gruber, 2010). However, other research indicates that non-verbal cues, like facial expressions, may actually lead to lower performance on tasks, potentially due to their distracting nature (Frechette & Moreno, 2010). Further research has failed to produce an interaction between animation styles (animated vs. non-animated) and levels of intimacy with the agent (Potdevin, Clavel, & Sabouret, 2021). These findings may suggest that while users want to be engaged during their virtual therapeutic sessions, artificial movements may produce negative reactions which impede forming a bond with the virtual coach. Perhaps, the lack of movement from the virtual coach is preferred to the unnatural non-verbal body language.

Lastly, the usability scores from the current AirHeart application were much higher compared to the first version, taking it from a below average score in the first version to an average score in the current version (Six et al., 2022). While all components of the original AirHeart application remained embedded within the current version, new updates were made based on user testing and heuristic analysis. Principles of human factors and UX should be maintained and utilized during the re-design and development of the next generation AirHeart application.

Chapter 5: FINAL DISCUSSION

The concept of utilizing smartphone applications to distribute non-clinical therapeutic assistance to a diverse and widespread population is not a novel proposition. However, the element which is lacking from a majority of the 20-30,000 health applications on the virtual e-stores (Clay, 2021) is the inclusion of therapeutic practices supported by prior research, and the inclusion of mental health and UX professionals in the design and testing processes. Furthermore, little to no research has been conducted to investigate the use of ECAs within interactive CBT-based MHealth applications aimed at reducing symptoms of depression. This is surprising as CBT-based mental health treatment often includes talk therapy in a one-one-one or group setting (Mayo Clinic Staff, 2019). While ECAs have been used in other CBT-based computer-based programs (Burton et al., 2016) and online systems (Suganuma, Sakamoto, & Shimoyama, 2018), neither utilized a guided approach with a virtual therapeutic coach leading the user through a lesson and activities. This is unexpected considering the strong support for an interactive design for therapeutic assistance (Bird et al., 2018; Gaffney et al., 2019; Moher et al., 2009). The purpose of the current three-part research endeavor was to investigate whether the inclusion of virtual characters would help enhance the usability of the AirHeart app, positively impact user experience, such as connection, likeability, comfort, trust, satisfaction, and bonding with the virtual character and app enjoyment, and further reduce symptoms of depression.

Experiment 1 investigated whether individuals demonstrated high levels of preference for customization or evolution of a virtual companion. While most smartphone applications involve some element of customization, evolution or progression as a motivator has been largely under-investigated. Results demonstrated a stronger preference for customization from individuals in the depressive group when compared to the non-depressive group; however, no difference was discovered regarding connection, likeability of the character, or enjoyability of the activity between the customization and evolution conditions. Furthermore, a congruence was discovered with both depressive and non-depressive groups citing the same order of preference for various virtual characters: 1) self-representative character, 2) an animal character, 3) a fantasy animal agent, 4) a human, non-representative virtual character, and 5) a plant character.

In Experiment 2, two high anthropomorphic virtual agents were tasked with the role of acting as a virtual therapeutic coach. This experiment investigated whether participants expressed a preference between a passive, video-like therapeutic experience or an interactive therapeutic experience. During the passive experience, participants were given a static, non-moving virtual therapeutic coach that vocally taught the material of two CBT-modules without any collaboration. For the interactive experience, an engaging, animated virtual therapeutic coach asked the participant questions to ensure understanding and connection towards the CBT modules. These results demonstrated that levels of worry and rumination significantly decreased after interacting with the virtual therapeutic coach. While no difference between depressive groups was indicated, a significant preference for the interactive agent was discovered for the overall sample. Furthermore, higher levels of trust and overall satisfaction were found in the interactive condition when compared to the passive condition. Participants reported a stronger preference for the active condition primarily due to the level of engagement and ability to maintain their attention, the connection they felt with the ECA, and the self-reflection that was encouraged. For selecting which ECA, Leo or Val, participants reported that they often selected the ECA that matched their gender.

Lastly, Experiment 3 further investigated the effect of conversation and animation features of a virtual therapeutic coach in relation to user preference and reduction in mental health symptoms over the course of two-weeks. Results demonstrated that participants saw a significant reduction of symptoms of depression, anxiety, stress, and rumination over the two-week period. Additionally, the non-animated condition experienced higher levels of bonding with the virtual therapeutic coach when compared to the animated condition. There was no difference between conversational and animated conditions regarding overall experience with the virtual coach, the AirHeart app, and trust with the agent. Lastly, the usability score for AirHeart V4 (the version used in the Experiment 3) is significantly higher than the score for AirHeart V1 (Six et al., 2022).

Of all the features examined in this research, the most crucial for both depressive and non-depressive individuals is the option to include customization. The ability to customize not only one's virtual agent, but the application itself is crucial for reasons of autonomy, preference, usability, and overall satisfaction. Design choices, such as interaction style (active vs. passive), agent voice selection, and personalization towards the end user can mitigate or enhance the supporting features, such as journaling, mood tracking, and psychotherapeutic modules (CBT, mindfulness, EFT, etc.). From customization, other subjective preferences may improve, such as likability of the application or virtual agent. For optional mental health applications, ample precedence should be placed on making the application enjoyable for all types of users (i.e. age, race, gender, etc.) for reasons of adherence. The clinical effectiveness of the application may become void if it is paired with high rates of attrition; if people do not use the application, they will not receive the benefits. As a multitude of mental health applications experience high rates of attrition in studies (Bosso, 2020; Fish and Saul, 2019; Kladnitski et al., 2020; Reid et al., 2011), a high emphasis should be placed on improving enjoyment.

This set of studies investigated whether virtual characters would be effective as therapeutic companions and coaches in reducing symptoms of depression and eliciting a positive user experience. While a small effect was observed in Experiment 3 ($\eta p^2 = 0.05$), the moderate effect size from the first AirHeart study ($\eta p^2 = .075$) suggests that the AirHeart version without the ECA exhibited a greater impact on depressive symptoms over the two-week period (Six et al., 2022). This reduction in effect size could have been due to the distaste for the virtual therapeutic coach experienced by many of the participants. While this remains speculation, this experiment demonstrates a scenario where negative user experiences with an application can minimize or negate the positive benefits, such as connection, likeability, and trust shown through scientifically supported research (Gaffney et al., 2019; Lee et al., 2021). Further development and iterative testing of ECAs regarding voice, animations, and personalization of material to the individual is critical for successful integrations into MHealth applications. Alternatively, app developers could utilize numerous features, including conversational, animated virtual agents, journaling, and journey customization, but provide users with the *option* to use or not utilize these features. Essentially, as long as evidence-based CBT content remains at the core of the MHealth apps, allowing users to personalize additional features within the app that they can choose to engage with may be key to optimizing user experience. This customization would not only enhance autonomy but provide a chance to reduce potential negative effects of disliking certain features. As mental health is never a "one size fits all" scenario, mental health applications should follow the same principle.

Mental Health Application Recommendations

Findings from this research showed that both depressive and non-depressive individuals strongly prefer customization of a virtual agent and the application over evolution or progression. While both could be included in a mental health application, customization should be prioritized. Additionally, mental health applications should include both interactive and passive options for therapeutic modules to provide the user with autonomy. As most individuals greatly preferred the active option in Experiment 2, app developers may consider setting an interactive therapy style as the standard with the option to change to a passive interaction design.

Mental health applications that include the use of a virtual therapeutic coach should also consider utilizing a high-quality TTS program which closely mimics the vocal patterns, pitches, and tones of a human. Facial expressions and body and mouth movements should be well defined and in-sync with one another to create a unified, realistic therapeutic coach. Interactions with the virtual therapeutic coach should go through multiple rounds of alpha and beta testing with a diverse population to ensure the agent appears realistic, while also taking care not to fall into the Uncanny Valley. Lastly, a generalized virtual therapeutic interaction will not be as effective as one personalized to the individual. A priority should be placed on finding ways to introduce adaptive or dynamic conversational interactions into the therapeutic modules. This will allow the participant to feel more connected to both the virtual therapeutic coach and the therapeutic learning module.

Contributions

The current set of studies provides several contributions to the literature on bCBT, customization, interactive experience with a virtual agent, and the effects of low-quality TTS. First, prior research has suggested that bCBT is effective in smartphone mental health application contexts for reducing symptoms of depression and anxiety (Bakker et al., 2018; Cully & Teten, 2008; Dahne et al., 2019; Moberg et al., 2019; Richards & Richards, 2012; Roepke et al., 2015; Six et al., 2022). Experiment 2 demonstrated a reduction in worry and rumination over the course of four therapeutic modules; however, this finding should be interpreted with caution as it was only one 45-minute intervention. Experiment 3 demonstrated a significant reduction in symptoms of depression, anxiety, stress, and rumination over a two-week duration with eight therapeutic modules. These findings support the idea that bCBT can be used as an effective short-term solution for helping improve the quality of life of a variety of individuals.

Secondly, prior research has demonstrated that customization is viewed as crucial to users; when customization was restricted or removed, it severely negatively affected motivation and adherence (Oyebode et al., 2020; Starwarz et al., 2014; Zhang et al., 2021). Furthermore, research has shown that customization in mental health applications increased levels of engagement and preference amongst users (Alqahtani & Orji, 2020). Our findings support this idea as it was observed that users preferred the customization condition over progression or evolution in Experiment 1 and provided repeated suggestions for improved customization in Experiment 3. Thirdly, participants value having an active role in their therapeutic experience, and those who take an active role show a significant reduction in negative symptoms and improved quality of life (Kertes et al., 2011l Uzoka, 1983). Furthermore, virtual agents have shown promise acting in a mental health counselor role (Burton et al., 2016; Fitzpatrick et al., 2017; Fulmer et al., 2018; Gaffney et al., 2019). The findings of Experiment 2 support this idea. To our knowledge, Experiment 2 was one of the first studies to demonstrate a majority preference for an active therapeutic interaction as opposed to a passive, video-like interaction. These findings provide support that, with more research, virtual agents could be effective in the role of a virtual mental health coach, as an adjuvant to traditional therapy.

Fourth, prior research has demonstrated that artificial voices can induce eerie feelings, and low-quality TTS often results in lower levels of trust in the virtual agent than a pre-recorded human voice (Abdulrahman & Richards, 2022; Schroeder et al., 2021). The findings of Experiment 3 support this research; a majority of participants reported distaste for the choice of TTS voice utilized for the virtual therapeutic coach. Low levels of trust with the agent were seen across all four experimental conditions, and the coach's voice may have played a large role. Furthermore, Experiment 3 demonstrates how virtual agents may achieve the opposite effect as intended if participants perceive the agent negatively. Lastly, to our knowledge, this is the first experiment to empirically investigate the effects of animation and conversation features in a virtual agent in a therapeutic context over a longitudinal period.

Future Research

Preferences for One-Time vs. Extended Use for Virtual Agents

In the current set of experiments, Experiment 2 and Experiment 3 both utilized virtual agents in a therapeutic coach type role, however, the amount of time the participant spent with the virtual agent differed. In Experiment 2, the participant interacted with the agent for around 45 minutes, while in Experiment 3, the participant spent around two weeks with the agent. Future research should investigate whether a difference in subjective preference for the agent's traits exists for different amounts of time a user interacts with an agent. Voice, levels of animation, dimensionality, and agency should be investigated regarding the difference between a one-time agent, such as a virtual nurse who simply administers forms before the individual has their appointment with the doctor, and an extended use agent, such as a therapeutic agent who interacts with the individual daily.

Introduction of a Module Bank

This research provides empirical evidence that bCBT delivered through an MHealth app can effectively reduce depressive symptoms over a two-week period. However, additional content using other therapeutic approaches, such as acceptance and compassion therapy (ACT) and self-compassion, may also be beneficial in addition to the CBT content. Future research could implement a "module bank" which contains 20-30 different options for a variety of different therapeutic options, ACT and self-compassion, and different topics with which users may struggle, such as anxiety, stress, rumination, and grief. By introducing this bank, users could self-select topics that would relate to their struggles in addition to the CBT content. More specifically, a quick assessment could be conducted through the app, and based on the results, the app could recommend content to users. This would imitate an in-person experience with a counselor who personalizes the lessons and coping strategies to the patient. Additionally, this would improve the autonomy within the AirHeart application.

Flexible Emotional Cognitive Modeling

As a means of improving connection with the participant, developers can implement a flexible emotional cognitive modeling style to enhance the perceived empathy of the ECA (Burton et al., 2016; Martínez-Miranda, Bresó, & García-Gómez, 2014). This type of cognitive architecture combines systems engineering, computer science, and psychology to create theoretical frameworks about emotion regulation (Bosse, Pontier, & Treur, 2010; Gross, 1998). By introducing a flexible emotional cognitive model, the virtual therapeutic coach could observe or calculate the participant's current emotional state, personality, diagnosis (if applicable), and negative symptoms and return them to a state of emotional equilibrium (Pico et al., 2024). The coach could have access to psychological measurements completed by the participant, such as the PHQ-8, GAD-7, PSS-10, and BFI, and physiological measurements, like heart rate, skin conductance, and eye blink rate, to help determine current emotional state. An acute in-lab experiment could be conducted to assess comfort and experience with the agent prior to a longitudinal study outside of the lab with the ECA connected to a MHealth application and to a physiological tracker, such as a Fitbit or Apple Watch.

High Quality Text to Speech (TTS) compared to Pre-Recorded Voices

A separate experiment should be conducted how a high-quality TTS engine would compare to a pre-recorded human voice in a therapeutic setting. In Experiment 3, low levels of trust and overall experience with the coach and AirHeart application may have been caused by a robotic sounding TTS. A short-term experiment, similar to Experiment 2, could be conducting using low-quality TTS, high-quality TTS, and pre-recorded voices to determine which held the highest level of preference amongst participants. This could also help confirm whether trust and overall experience with the coach and application is severely impacted by voice quality.

Animation Styles in a Therapeutic Setting

Research on animations within a virtual healthcare setting seems to be conflicting on whether they are perceived as helpful towards making the agent seem more realistic or distracting (Bickmore & Gruber, 2010; Frechette & Moreno, 2010). Additional research should be conducted in a virtual therapeutic setting with different degrees of subtly for animations: subtle: just facial expressions and light mouth movements, moderate: facial expressions, normal mouth movements, nodding, and some subtle body movements, and severe: exaggerated facial expressions, dramatic mouth movements, nodding, additional gesturing, and body movements. The impact of these differences on trust, bonding, and user experience could then be examined. Researchers could also determine whether a difference in animation styles exists between a therapeutic, educational, and entertainment setting.

Virtual Agent vs. Simple Text Condition

Future work may consider testing whether having a virtual agent in the role of a therapeutic coach adds further value to the mental health application compared to CBT content alone in terms of reducing symptoms of depression. Personality differences, fluency with technology, experience with virtual agents, trust in technology, and depressive symptoms could be included to determine whether an individual's character or past experiences heavily impacts their relationship with the agent. This could be compared to the option of simply reading through CBT material and performing activities on their own as opposed to an agent guiding them through the process.

Conclusion

Overall, mental health applications hold potential for providing therapeutic benefits as an adjuvant to in-person therapy While other CBT-based mental health programs have used ECAs to reduce symptoms of depression (Burton et al., 2016; Suganuma et al., 2018), AirHeart V4 is the first to combine ECAs in the role of a virtual therapeutic coach who guides participants through lessons and activities in a smartphone application format. This research provides support that CBT-based mental health apps can help reduce symptoms of depression, but whether virtual agents provided an added level of therapeutic benefits to mental health applications remains relatively unclear. Customization and personalization of an MHealth application is crucial for success in improving overall quality of life and reducing negative symptoms. Conducting research related to MHealth apps and ways to improve their features, such as mood tracking, journaling, customization, and virtual characters, provides necessary insight for developers. This research provides novel, evidence-based techniques to create an effective application used in conjunction with current mental health treatments. Future research designs should prioritize efficacy, improvement, and trans-diagnostic capabilities of virtual characters while being mindful of cost to prevent unwarranted barriers to mental health assistance.

APPENDIX A

Figure A1: Female User Persona Demonstrating the None-Mild Depression Category

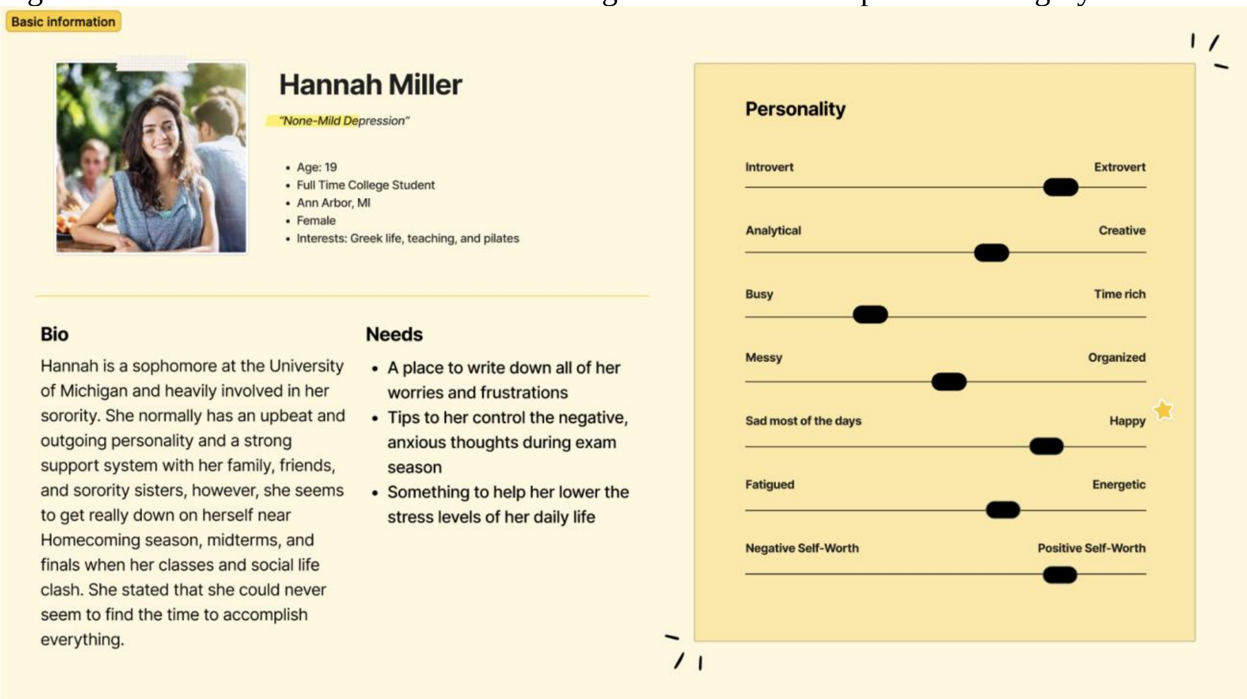


Figure A2: Male User Persona Demonstrating the Mild Depression Category

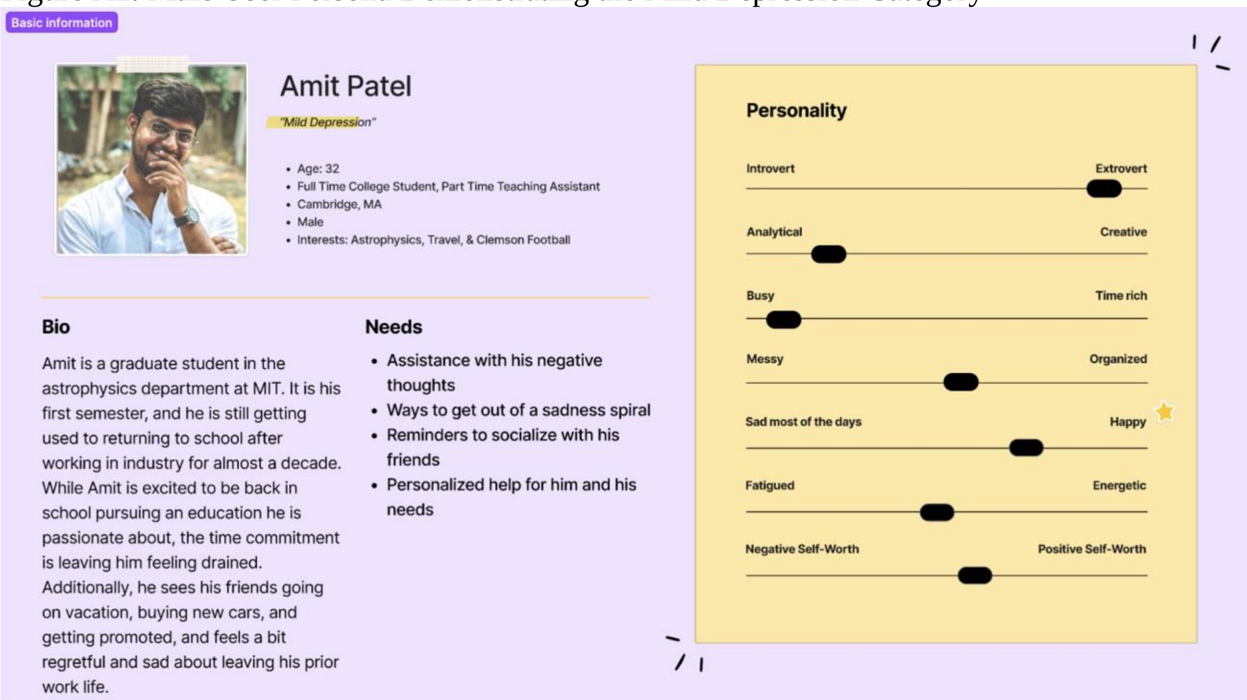


Figure A3: Female User Persona Demonstrating the Mild Depression Category

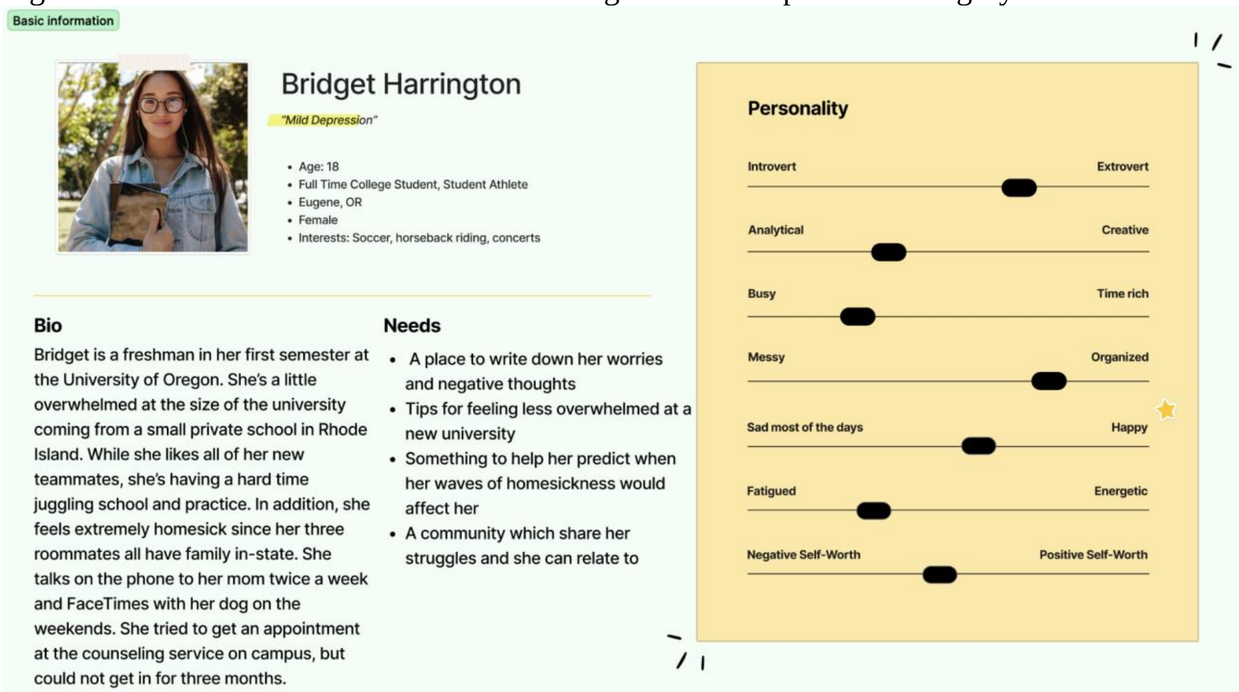


Figure A4: Male User Persona Demonstrating the Mild-Moderate Depression Category

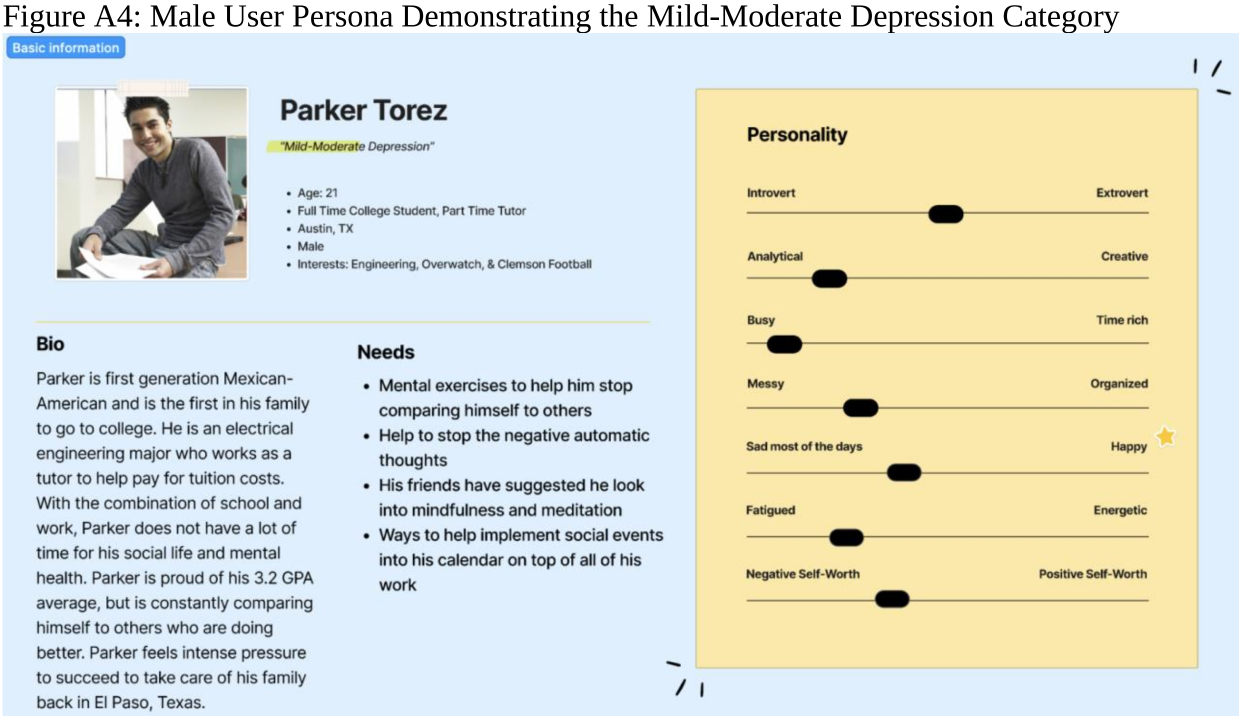


Figure A5: Non-Binary User Persona Demonstrating the Mild-Moderate Depression Category

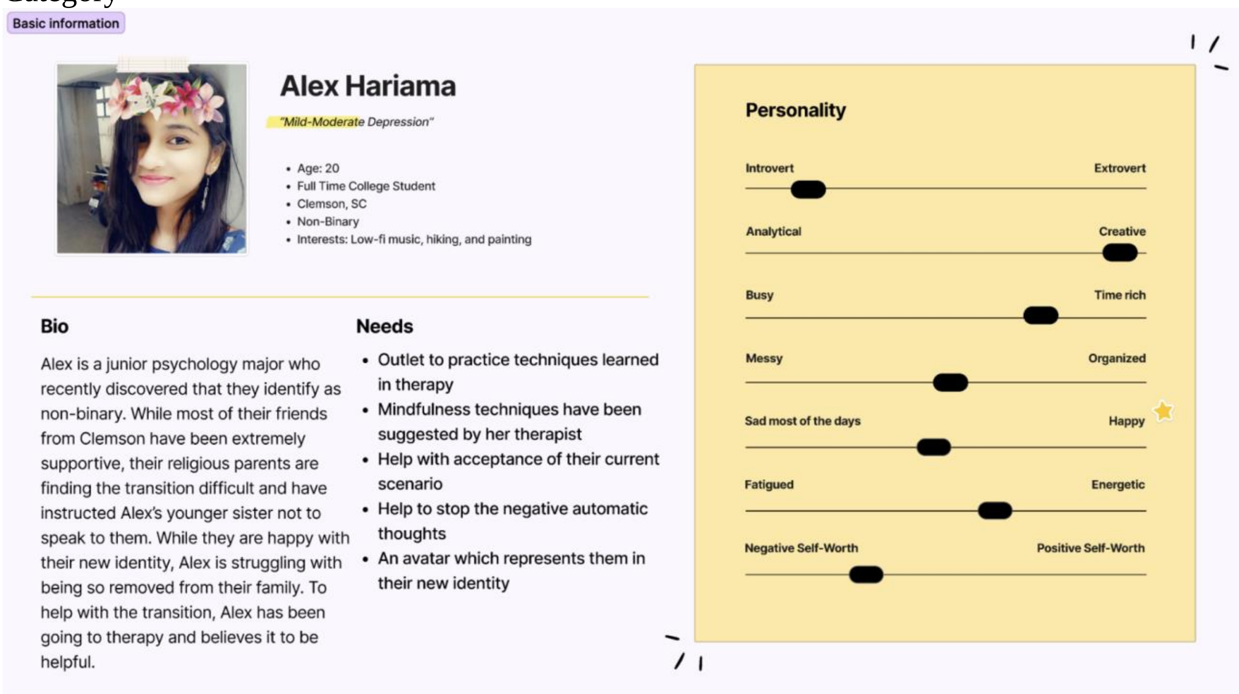


Figure A6: Female User Persona Demonstrating the Mild-Moderate Depression Category

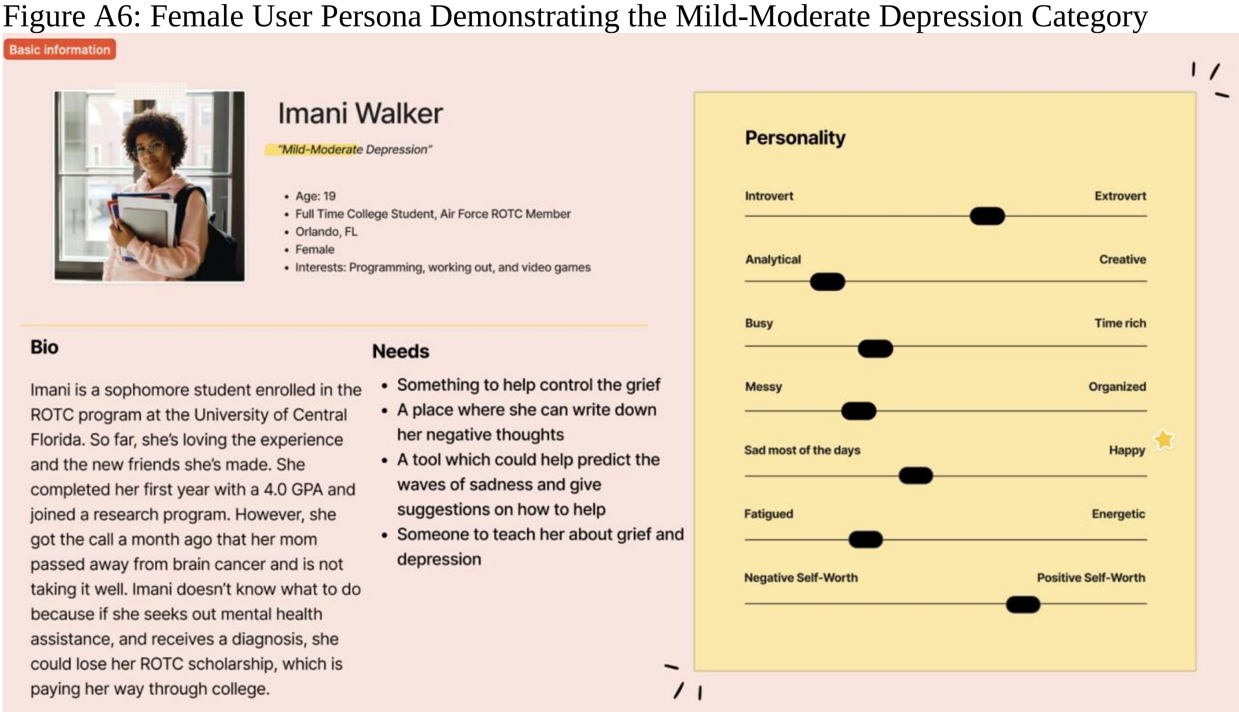


Figure A7: Amit’s Journey Map

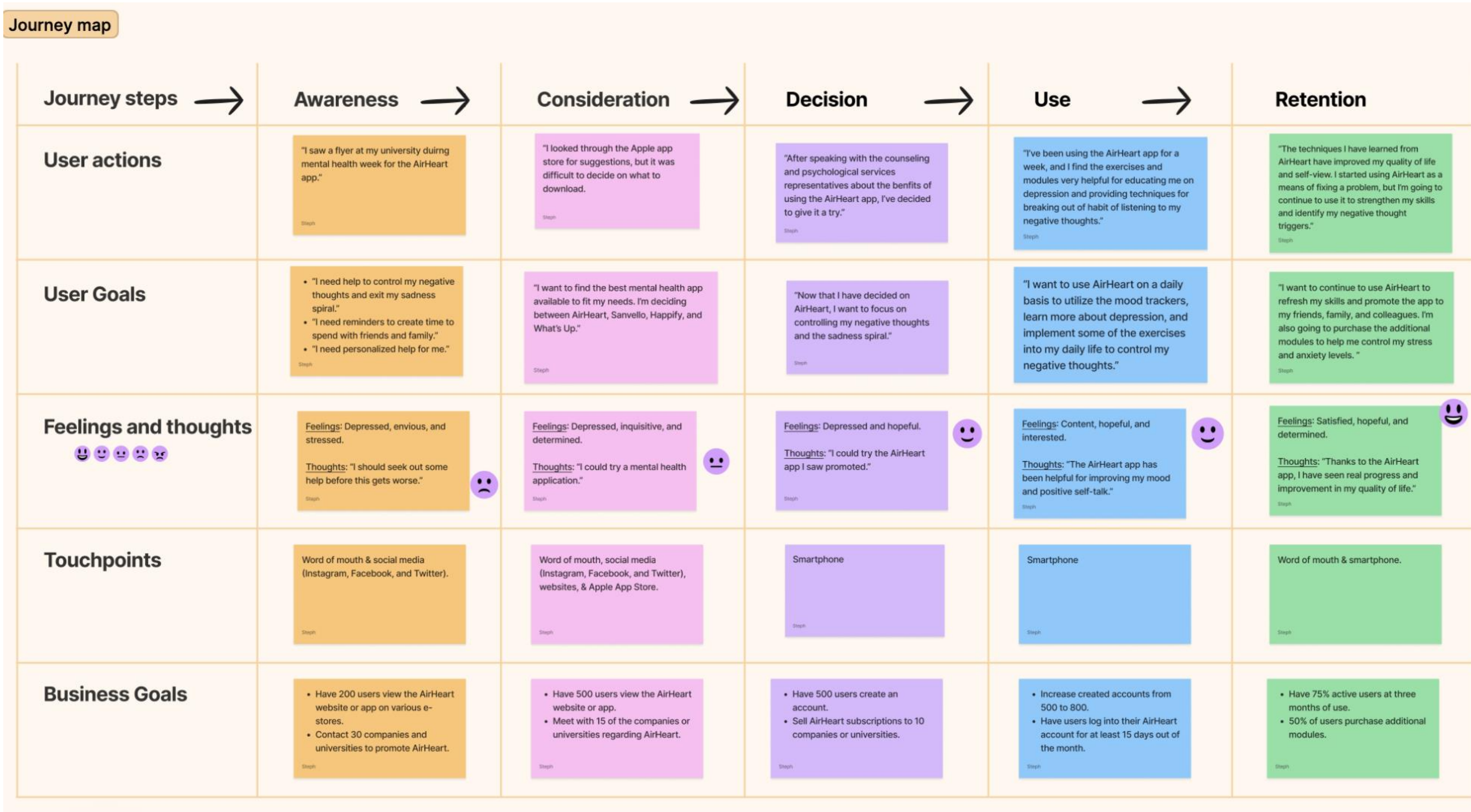


Table A8: Competitor Analysis of AirHeart Against Five Other MHealth Apps

Feature	Youper	What's Up?	Sanvello	Replika	Woebot	AirHeart
Cognitive Behavioral Therapy (CBT)	CBT sessions	Cognitive restructuring	CBT tools	-	Step-by-step guidance using CBT.	Seven replay-able CBT modules.
Problem-solving	Understanding a problem, crafting a solution, and executing it.	-	-	-	-	Problem solving techniques and activities
Goal setting	Provides guidance to set goals.	A positive and negative habit tracker.	-	Work towards a variety of goals.	App checks in with the user about their goals.	Provides guidance and activities to help with goal setting.
Journal	CBT Diary	Diary with a mood scale	-	-	-	Thought journal
Mood tracker	Mood tracker and identifies patterns	-	Track emotions and progress over time	Mood tracker for thoughts & feelings	-	Weekly mood tracker
Symptom tracker	Symptom tracker for mental health disorders	-	-	-	-	Symptom tracker in weekly mood tracker
Virtual Character	-	-	-	-	-	Provide a customizable virtual character
Relaxation and meditation	-	Breathing techniques	-	-	Master skills to reduce stress and live happier	Provides modules for meditation and relaxation

Feature	Youper	What's Up?	Sanvello	Replika	Woebot	AirHeart
Rating	4.8/5	4.4/5	4.8/5	4.5/5	4.7/5	82.5/100

Table A9: Competitor Analysis of The New AirHeart Prototype with Five Popular MHealth Apps

Features	Youper	What's Up?	Sanvello	Replika	Woebot	AirHeart
Cognitive Behavioral Therapy(CBT)	✔	✔	X	X	✔	✔
Problem-solving	✔	✔	X	X	X	✔
Goal setting	✔	X	X	✔	✔	✔
Journal	✔	X	X	X	X	✔
Mood tracker	✔	X	✔	✔	✔	✔
Symptom tracker	✔	X	X	X	X	✔
Virtual Character	X	X	X	X	X	✔
Relaxation and meditation	X	✔	✔	X	✔	✔

REFERENCES

A-tjak, J. G., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A., & Emmelkamp, P. M. (2015). A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics*, *84*(1), 30-36.

Abdulrahman, A., & Richards, D. (2022). Is natural necessary? Human voice versus synthetic voice for intelligent virtual agents. *Multimodal Technologies and Interaction, 6*(7), 51.

Ackerman, C. E. (2021, May 20). *25 CBT techniques and worksheets for cognitive behavioral therapy*. PositivePsychology.com. Retrieved October 29, 2021, from https://positivepsychology.com/cbt-cognitive-behavioral-therapy-techniques-worksheets/.

Alqahtani, F. & Orji, R. (2020). Insights from user reviews to improve mental health apps. Health informatics journal, 26(3), 2042-2066.

Alqahtani, F., Winn, A., & Orji, R. (2021). Co-designing a mobile app to improve mental health and well-being: focus group study. *JMIR formative research*, *5*(2), e18172.

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Asbury, E. T., Casey, J., & Desai, K. (2018). Family eJournal: benefits of online guided group journaling for women. *Journal of Public Mental Health*.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2018). A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behaviour Research and Therapy*, *109*, 75-83.

Balakrishnan, K., & Honavar, V. (2001). Evolutionary and neural synthesis of intelligent agents.

Baymard Institute. (2022, August 8). *The step-by-step guide to UX competitive analysis*. Baymard Institute. https://baymard.com/learn/competitive-analysis-ux.

Beckham, E. E., Leber, W. R., Watkins, J. T., Boyer, J. L., & Cook, J. B. (1986). Development of an instrument to measure Beck's cognitive triad: the Cognitive Triad Inventory. *Journal of consulting and clinical psychology*, *54*(4), 566.

Bell, A. C., ' D'Zurilla, T. J. (2009). Problem-solving therapy for depression: a meta analysis. *Clinical psychology review*, *29*(4), 348-353.

Bethune, Sophie. (2021). Demand for mental health treatment continues to increase, say psychologists. American Psychological Association. https://www.apa.org/news/press/releases/2021/10/mental-health-treatment-demand

Bickmore, T., & Gruber, A. (2010). Relational agents in clinical psychiatry. *Harvard review of psychiatry*, *18*(2), 119-130.

Bird, T., Mansell, W., Wright, J., Gaffney, H., & Tai, S. (2018). Manage your life online: a web based randomized controlled trial evaluating the effectiveness of a problem-solving intervention in a student sample. *Behavioural and cognitive psychotherapy, 46*(5), 570-582.

Birk, M. V., Mandryk, R. L., & Atkins, C. (2016, October). The motivational push of games: The interplay of intrinsic motivation and external rewards in games for training. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 291-303).

Bodenschatz, C. M., Czepluch, F., Kersting, A., & Suslow, T. (2021). Efficient visual search for facial emotions in patients with major depression. *BMC psychiatry*, *21*, 1-12.

Bosse, T., Pontier, M., & Treur, J. (2010). A computational model based on Gross' emotion regulation theory. *Cognitive systems research, 11*(3), 211-230.

Bosso, K. B. (2020). *The Effects of Mindfulness Training on BDNF Levels, Depression, Anxiety, and Stress Levels of College Students* (Doctoral dissertation, Florida Atlantic University).

Brach, T. (2016). Feeling Overwhelmed? Remember RAIN. *X*, *January*, *13*.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Bresó, A., Martínez-Miranda, J., Botella, C., Baños, R. M., & García-Gómez, J. M. (2016). Usability and acceptability assessment of an empathic virtual agent to prevent major depression. *Expert Systems, 33*(4), 297-312.

Bubolz, S., Mayer, G., Gronewold, N., Hilbel, T., & Schultz, J. H. (2020). Adherence to established treatment guidelines among unguided digital interventions for depression: quality evaluation of 28 web-based programs and mobile apps. *Journal of Medical Internet Research*, *22*(7), e16136.

Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., & Mohr, D. C. (2011). Harnessing context sensing to develop a mobile intervention for depression. *Journal of medical Internet research*, *13*(3), e1838.

Burton, C., Szentagotai Tatar, A., McKinstry, B., Matheson, C., Matu, S., Moldovan, R., ... & Help4Mood Consortium. (2016). Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. *Journal of telemedicine and telecare*, *22*(6), 348-355.

Cambridge University Press. (n.d.). Customization. In Cambridge Dictionary. Retrieved October 30, 2022, from https://dictionary.cambridge.org/us/dictionary/english/customization

Carter, M. C., Burley, V. J., Nykjaer, C., & Cade, J. E. (2013). Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. Journal of medical Internet research, 15(4), e2283.

Chang, T. R., Kaasinen, E., & Kaipainen, K. (2012, December). What influences users' decisions to take apps into use? A framework for evaluating persuasive and engaging design in mobile apps for well-being. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia* (pp. 1-10).

Chaves, C., Lopez-Gomez, I., Hervas, G., & Vazquez, C. (2017). A comparative study on the efficacy of a positive psychology intervention and a cognitive behavioral therapy for clinical depression. Cognitive Therapy and Research, 41(3), 417-433.

Cheek, C., Bridgman, H., Fleming, T., Cummings, E., Ellis, L., Lucassen, M. F., ... & Skinner, T. (2014). Views of young people in rural Australia on SPARX, a fantasy world developed for New Zealand youth with depression. *JMIR Serious Games*, *2*(1), e3183.

Christ, C., Schouten, M. J., Blankers, M., van Schaik, D. J., Beekman, A. T., Wisman, M. A., ... & Dekker, J. J. (2020). Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(9), e17831.

Chu, H. Y., & Shen, Y. (2022, June). User Feedback Design in AI-Driven Mood Tracker Mobile Apps. In *Human-Computer Interaction. User Experience and Behavior: Thematic Area, HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part III* (pp. 346-358). Cham: Springer International Publishing.

Clay, R. A. (2021). Mental health apps are gaining traction. Monitor on Psychology. Retrieved November 5, 2021, from https://www.apa.org/monitor/2021/01/trends-mental-health-apps.

Cohen S, Kamarck T, Mermelstein R (1983) A global measure of perceived stress. *Journal of Health and Social Behavior* 24(4): 385–396.

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. *Measuring stress: A guide for health and social scientists*, *10*(2), 1-2.

Cole, D. A., Nick, E. A., Zelkowitz, R. L., Roeder, K. M., & Spinelli, T. (2017). Online social support for young people: does it recapitulate in-person social support; can it help?. Computer in human behavior, 68, 456-464.

Cooper, A. (1999). The inmates are running the asylum (pp. 17-17). Vieweg+ Teubner Verlag.

Cooper, A., Reimann, R., & Cronin, D. (2015). About face. *The Essentials of Interaction Design*, *3*.

Conroy, Jessica, Lin, Luona, & Ghansess, Amrita. (2020). Why people aren't getting the care they need. American Psychological Association. https://www.apa.org/monitor/2020/07/datapoint-care

Cosco, T. D., Prina, M., Stubbs, B., & Wu, Y. T. (2017). Reliability and validity of the Center for Epidemiologic Studies Depression Scale in a population-based cohort of middle-aged US adults. *Journal of nursing measurement*, *25*(3).

Cully, J. A., & Teten, A. L. (2008). A therapist's guide to brief cognitive behavioral therapy. *Houston: Department of Veterans Affairs South Central MIRECC*.

D'Alfonso, S. (2020). AI in mental health. *Current Opinion in Psychology*, *36*, 112-117.

Dahlin, M., Johansson, A., Romare, H., Carlbring, P., & Andersson, G. (2022). Worry specific versus self-tailored internet-based treatments for generalized anxiety disorder, with scheduled support or support on demand: A pilot factorial design trial. Internet Interventions, 28, 100531.

Dahne, J., Collado, A., Lejuez, C. W., Risco, C. M., Diaz, V. A., Coles, L., ... & Carpenter, M. J. (2019). Pilot randomized controlled trial of a Spanish-language Behavioral Activation mobile app (¡ Aptívate!) for the treatment of depressive symptoms among united states Latinx adults with limited English proficiency. *Journal of affective disorders*, *250*, 210-217.

David, A., & Glore, P. (2010). The impact of design and aesthetics on usability, credibility, and learning in an online environment. *Online Journal of Distance Learning Administration*, *13*(4), 43-50.

Davids, M. R., Chikte, U. M., & Halperin, M. L. (2013). An efficient approach to improve the usability of e-learning resources: the role of heuristic evaluation. *Advances in physiology education*, *37*(3), 242-248.

DiSalvo, C., Gemperle, F., & Forlizzi, J. (2005). Imitating the human form: Four kinds of anthropomorphic form. *Unpublished manuscript. Accessed April*.

Doherty, G., Coyle, D., Matthews, M. (2010). Design and evaluation guidelines for mental health technologies. Interacting with computers, 22(4), 243-252.

Donker, T., Griffiths, K. M., Cuijpers, P., & Christensen, H. (2009). Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. *BMC medicine*, *7*(1), 1-9.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M. R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of medical Internet research*, *15*(11), e247.

Donovan, C. L., Poole, C., Boyes, N., Redgate, J., & March, S. (2015). Australian mental health worker attitudes towards cCBT: What is the role of knowledge? Are there differences? Can we change them?. *Internet Interventions*, *2*(4), 372-381.

Eftychiou, L., & El Morr, C. (2017). Mobile mental health virtual communities: challenges and opportunities. *The Digitization of Healthcare: New Challenges and Opportunities*, 257-275.

Erickson, T. (1997). Designing agents as if people mattered. *Software agents*, 79-96.

Everitt, N., Broadbent, J., Richardson, B., Smyth, J. M., Heron, K., Teague, S., & Fuller Tyszkiewicz, M. (2021). Exploring the features of an app-based just-in-time intervention for depression. *Journal of affective disorders*, *291*, 279-287.

Fernandes, B., Biswas, U. N., Mansukhani, R. T., Casarín, A. V., & Essau, C. A. (2020). The impact of COVID-19 lockdown on internet use and escapism in adolescents. *Revista de psicología clínica con niños y adolescentes*, *7*(3), 59-65.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*, *16*(3), 287-298.

Fish, M. T., & Saul, A. D. (2019). The gamification of meditation: A randomized controlled study of a prescribed mobile mindfulness meditation application in reducing college students' depression. *Simulation & Gaming*, *50*(4), 419-435.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). 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 mental health*, *4*(2), e7785.

Forand, N. R., Barnett, J. G., Strunk, D. R., Hindiyeh, M. U., Feinberg, J. E., & Keefe, J. R. (2018). Efficacy of guided iCBT for depression and mediation of change by cognitive skill acquisition. *Behavior therapy*, *49*(2), 295-307.

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical agents' presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology: Theories, Methods, and Applications*, *22*(2), 61–72.

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR mental health*, *5*(4), e9782.

Gallego, J., Aguilar-Parra, J. M., Cangas, A. J., Langer, Á. I., & Mañas, I. (2014). Effect of a mindfulness program on stress, anxiety and depression in university students. *The Spanish journal of psychology*, *17*, E109.

Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR mental health*, *6*(10), e14166.

Gautam, M., Tripathi, A., Deshmukh, D., & Gaur, M. (2020). Cognitive behavioral therapy for depression. *Indian journal of psychiatry*, *62*(Suppl 2), S223.

Greenberg, P. E., Fournier, A. A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S. H., & Kessler, R. C. (2021). The economic burden of adults with major depressive disorder in the United States (2010 and 2018). *Pharmacoeconomics*, *39*(6), 653-665.

Gould, R. L., Coulson, M. C., & Howard, R. J. (2012). Cognitive behavioral therapy for depression in older people: A meta-analysis and meta-regression ofrandomized controlled trials. Journal of the American Geriatrics Society, 60(10), 1817-1830.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of general psychology*, *2*(3), 271-299.

Grosse Holtforth, M., Krieger, T., Zimmermann, J., Altenstein-Yamanaka, D., Dörig, N., Meisch, L., & Hayes, A. M. (2019). A randomized-controlled trial of cognitive behavioral therapy for depression with integrated techniques from emotion focused and exposure therapies. Psychotherapy Research, 29(1), 30 44.

Gu, J., Strauss, C., Bond, R., & Cavanagh, K. (2016). " How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies": Corrigendum.

Hamilton, L., Elliott, D., Quick, A., Smith, S., & Choplin, V. (2023). Exploring the Use of AI in Qualitative Analysis: A Comparative Study of Guaranteed Income Data. International Journal of Qualitative Methods, 22, 16094069231201504.

Hallford, D. J., Sharma, M. K., & Austin, D. W. (2020). Increasing anticipatory pleasure in major depression through enhancing episodic future thinking: A randomized single-case series trial. *Journal of Psychopathology and Behavioral Assessment*, *42*, 751-764.

Hayes, Marianne. (2021). How much does therapy cost, and how do you pay for it? Northwestern Mutual. https://www.northwesternmutual.com/life-and-money/how-much-does-therapy-cost-and-how-do-you-pay-for-it/.

Hopman, K., Richards, D., & Norberg, M. M. (2023). A digital coach to promote emotion regulation skills. *Multimodal Technologies and Interaction*, *7*(6), 57.

Howard, T. (2014). Journey mapping: A brief overview. *Communication Design Quarterly Review*, *2*(3), 10-13.

Hudson, G., Negbenose, E., Neary, M., Jansli, S. M., Schueller, S. M., Wykes, T., & Jilka, S. (2022). Comparing professional and consumer ratings of mental health apps: mixed methods study. *JMIR formative research*, *6*(9), e39813.

Hyde, J., Carter, E. J., Kiesler, S., & Hodgins, J. K. (2016). Evaluating animated characters: facial motion magnitude influences personality perceptions. *ACM Transactions on Applied Perception (TAP)*, *13*(2), 1-17.

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, *6*(11), e12106.

Kaltenthaler, E., Sutcliffe, P., Parry, G., Beverley, C., Rees, A., & Ferriter, M. (2008). The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. *Psychological medicine, 38*(11), 1521-1530.

Kang, H. J., Kim, S. Y., Bae, K. Y., Kim, S. W., Shin, I. S., Yoon, J. S., & Kim, J. M. (2015). Comorbidity of depression with physical disorders: research and clinical implications. *Chonnam medical journal*, *51*(1), 8-18.

Kauer, S. D., Reid, S. C., Crooke, A. H. D., Khor, A., Hearps, S. J. C., Jorm, A. F., ... & Patton, G. (2012). Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. *Journal of medical Internet research*, *14*(3), e1858.

Kenny, R., Dooley, B., & Fitzgerald, A. (2016). Developing mental health mobile apps: Exploring adolescents' perspectives. *Health informatics journal*, *22*(2), 265-275.

Kertes, A., Westra, H. A., Angus, L., & Marcus, M. (2011). The impact of motivational interviewing on client experiences of cognitive behavioral therapy for generalized anxiety disorder. *Cognitive and Behavioral Practice*, *18*(1), 55-69.

Khubchandani, J., Brey, R., Kotecki, J., Kleinfelder, J., & Anderson, J. (2016). The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Archives of psychiatric nursing*, *30*(4), 457-462.

Kim, Y., Ji, S., Lee, H., Kim, J. W., Yoo, S., & Lee, J. (2016, May). " My Doctor is Keeping an Eye on Me!" Exploring the Clinical Applicability of a Mobile Food Logger. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5620-5631).

Kirkham, J. G., Choi, N., & Seitz, D. P. (2016). Meta-analysis of problem solving therapy for the treatment of major depressive disorder in older adults. *International journal of geriatric psychiatry*, *31*(5), 526-535.

Kladnitski, N., Smith, J., Uppal, S., James, M. A., Allen, A. R., Andrews, G., & Newby, J. M. (2020). Transdiagnostic internet-delivered CBT and mindfulness-based treatment for depression and anxiety: A randomised controlled trial. *Internet interventions*, *20*, 100310.

Koh, J., Tng, G. Y., & Hartanto, A. (2022). Potential and pitfalls of mobile mental health apps in traditional treatment: An umbrella review. *Journal of Personalized Medicine*, *12*(9), 1376.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ–4. *Psychosomatics*, *50*(6), 613-621.

Krpan, K. M., Kross, E., Berman, M. G., Deldin, P. J., Askren, M. K., & Jonides, J. (2013). An everyday activity as a treatment for depression: the benefits of expressive writing for people diagnosed with major depressive disorder. Journal of affective disorders, 150(3), 1148-1151.

Larsen, M. E., Nicholas, J., & Christensen, H. (2016). A systematic assessment of smartphone tools for suicide prevention. *PloS one*, *11*(4), e0152285.

Lee, E. H. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian nursing research*, *6*(4), 121-127.

Lee, S. K., Kavya, P., & Lasser, S. C. (2021). Social interactions and relationships with an intelligent virtual agent. *International Journal of Human-Computer Studies*, *150*, 102608.

Lee, H., Lee, J. A., Brar, J. S., Rush, E. B., & Jolley, C. J. (2014). Physical activity and depressive symptoms in older adults. *Geriatric nursing*, *35*(1), 37-41.

Lenz, A. S., & Hollenbaugh, K. M. (2015). Meta-analysis of trauma-focused cognitive behavioral therapy for treating PTSD and co-occurring depression among children and adolescents. *Counseling Outcome Research and Evaluation*, *6*(1), 18-32.

Li, B., Munoz, J. P., Rong, X., Chen, Q., Xiao, J., Tian, Y., ... & Yousuf, M. (2018). Vision-based mobile indoor assistive navigation aid for blind people. *IEEE transactions on mobile computing*, *18*(3), 702-714.

Lintvedt, O. K., Griffiths, K. M., Sørensen, K., Østvik, A. R., Wang, C. E., Eisemann, M., & Waterloo, K. (2013). Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial. *Clinical psychology & psychotherapy*, *20*(1), 10-27.

Litvin, S., Saunders, R., Maier, M. A., & Lüttke, S. (2020). Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: A randomized controlled trial. *PloS one*, *15*(9), e0237220.

Long, F. (2009, May). Real or imaginary: The effectiveness of using personas in product design. In *Proceedings of the Irish Ergonomics Society annual conference* (Vol. 14, pp. 1-10).

Lui, J. H., Marcus, D. K., & Barry, C. T. (2017). Evidence-based apps? A review of mental health mobile applications in a psychotherapy content. Professional Psychology: Research and Practice, 48(3), 199.

Luo, Y. (2021, June). Designing Multimodal Self-Tracking Technologies to Promote Data Capture and Self-Reflection. In *Companion Publication of the 2021 ACM Designing Interactive Systems Conference* (pp. 11-15).

MacIsaac, A., Mushquash, A. R., & Wekerle, C. (2022). Writing Yourself Well: Dispositional Self-Reflection Moderates the Effect of a Smartphone App-Based Journaling Intervention on Psychological Wellbeing across Time. *Behaviour Change*, 1-17.

Mantani, A., Kato, T., Furukawa, T. A., Horikoshi, M., Imai, H., Hiroe, T., ... & Kawanishi, N. (2017). Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. *Journal of medical Internet research*, *19*(11), e373.

Marshall, J. M., Dunstan, D. A., & Bartik, W. (2020a). Apps with maps—anxiety and depression Mobile apps with evidence-based frameworks: systematic search of major app stores. *JMIR mental health*, *7*(6), e16525.

Marshall, J. M., Dunstan, D. A., & Bartik, W. (2020b). Clinical or gimmickal: The use and effectiveness of mobile mental health apps for treating anxiety and depression. *Australian & New Zealand Journal of Psychiatry*, *54*(1), 20-28.

Martell, C. R., Dimidjian, S., & Herman-Dunn, R. (2021). *Behavioral activation for depression: A clinician's guide*. Guilford Publications.

Martínez-Miranda, J., Bresó, A., & García-Gómez, J. M. (2014). Modelling two emotion regulation strategies as key features of therapeutic empathy. *Emotion Modeling: Towards Pragmatic Computational Models of Affective Processes*, 115-133.

Mayo Clinic Staff. (2019, March 16). *Cognitive behavioral therapy*. Mayo Clinic. https://www.mayoclinic.org/tests-procedures/cognitive-behavioral-therapy/about/pac-20384610#:~:text=Cognitive%20behavioral%20therapy%20(CBT)%20is a%20limited%20number%20of%20sessions.

Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes?. *Design studies*, *32*(5), 417-430.

Micheaux, A., & Bosio, B. (2019). Customer journey mapping as a new way to teach data-driven marketing as a service. *Journal of Marketing Education*, *41*(2), 127-140.

Miller, W. (2014). Interactive journaling as a clinical tool. Journal of Mental Health Counseling, 36(1), 31-42.

Mims, R. (2015). Military veteran use of visual journaling during recovery. *Journal of Poetry Therapy*, *28*(2), 99-111.

Miner, K., Tischler, S., DeSena, D., & Rimer, J. (2016). CBT Group Program for Depression Patient Manual. Cognitive-Behavioral Therapy (CBT) Group Program for Depression Adult Patient Manual. Retrieved October 29, 2021, from https://medicine.umich.edu/sites/default/files/content/downloads/CBT-Group-Program-for-Depression-Patient Manual.pdf

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 Jun 8;21(6):e12556. PMID: 31199319

Moran, K.. (2017). *The Aesthetic-Usability Effect*. Accessed: Jan. 25, 2020. [Online]. Available: https://www.nngroup.com/articles/aesthetic-usability-effect/

Muñoz, R. F., Miranda, J., & Aguilar-Gaxiola, S. (2000). *Individual therapy manual for cognitive-behavioral treatment of depression*. Santa Monica, Calif: Rand.

Mustaffa S, Musa A M, Abu S, Yusof R M. Effects of cognitive behavior approaches in anxiety for cancer patient undergone chemotherapy. *International Journal of Fundamental Psychology & Social Sciences.* 2012;2:70–73

National Institute of Mental Health. (2022). Major Depression. National Institute of Mental Health. https://www.nimh.nih.gov/health/statistics/major-depression.

Neary, M., & Schueller, S. M. (2018). State of the field of mental health apps. *Cognitive and Behavioral Practice*, *25*(4), 531-537.

Ng, M. M., Firth, J., Minen, M., & Torous, J. (2019). User engagement in mental health apps: a review of measurement, reporting, and validity. *Psychiatric Services*, *70*(7), 538-544.

Nicholas, J., Larsen, M. E., Proudfoot, J., & Christensen, H. (2015). Mobile apps for bipolar disorder: a systematic review of features and content quality. *Journal of medical Internet research*, *17*(8), e198.

Nielson, J. (2012). *Usability 101: Introduction to Usability*. Nielson Norman Group. Retrieved April 26, 2023.

Noh, J. W., Kwon, Y. D., Park, J., Oh, I. H., & Kim, J. (2016). Relationship between physical disability and depression by gender: a panel regression model. *PloS one*, *11*(11), e0166238.

Ntoumanis, N., Ng, J. Y., Prestwich, A., Quested, E., Hancox, J. E., Thøgersen-Ntoumani, C., ...& Williams, G. C. (2021). A meta-analysis of self-determination theory-informed intervention studies in the health domain: Effects on motivation, health behavior, physical, and psychological health. Health psychology review, 15(2), 214-244.

O'Halloran, A. M., Kenny, R. A., & King-Kallimanis, B. L. (2014). The latent factors of depression from the short forms of the CES-D are consistent, reliable and valid in community-living older adults. *European Geriatric Medicine*, *5*(2), 97-102.

Oliveira, L., Bradley, C., Birrell, S., Davies, A., Tinworth, N., & Cain, R. (2019). Understanding passengers' experiences of train journeys to inform the design of technological innovations. *Design Discourse on Business and Industry: Re: Research, Volume 6*, *5*, 111.

Orgilés Amorós, M., Espada Sánchez, J. P., Delvecchio, E., Francisco, R., Mazzeschi, C., Pedro, M., & Morales Sabuco, A. (2021). Anxiety and depressive symptoms in children and adolescents during COVID-19 pandemic: A transcultural approach. *Psicothema*.

Otte, C. (2022). Cognitive behavioral therapy in anxiety disorders: current state of the evidence. *Dialogues in clinical neuroscience*.

Oud, M., De Winter, L., Vermeulen-Smit, E., Bodden, D., Nauta, M., Stone, L., ... & Stikkelbroek, Y. (2019). Effectiveness of CBT for children and adolescents with depression: A systematic review and meta-regression.

Parmar, D., Olafsson, S., Utami, D., Murali, P., & Bickmore, T. (2022). Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Autonomous agents and multi-agent systems*, *36*(1), 17.

Pew Research Center. (2021, April 7). *Demographics of mobile device ownership and adoption in the United States*. Pew Research Center. Retrieved November 2, 2021, from http://www.pewresearch.org/internet/factsheet/mobile/.

Pico, A., Taverner, J., Vivancos, E., Botti, V., & García-Fornes, A. (2024). Towards an Affective Intelligent Agent Model for Extrinsic Emotion Regulation. *Systems*, *12*(3), 77.

Pinto, M. D., Greenblatt, A. M., Hickman, R. L., Rice, H. M., Thomas, T. L., & Clochesy, J. M. (2016). Assessing the critical parameters of eSMART-MH: a promising avatar-based digital therapeutic intervention to reduce depressive symptoms. *Perspectives in psychiatric care*, *52*(3), 157-168.

Pinto, M. D., Hickman Jr, R. L., Clochesy, J., & Buchner, M. (2013). Avatar-based depression self-management technology: promising approach to improve depressive symptoms among young adults. *Applied nursing research*, *26*(1), 45-48.

Potdevin, D., Clavel, C., & Sabouret, N. (2021). Virtual intimacy in human-embodied conversational agent interactions: the influence of multimodality on its perception. *Journal on Multimodal User Interfaces*, *15*, 25-43.

Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research*, *19*(5), e151.

Pruitt, J., & Adlin, T. (2010). *The persona lifecycle: keeping people in mind throughout product design*. Elsevier.

Puyat, J. H., Kazanjian, A., Goldner, E. M., & Wong, H. (2016). How often do individuals with major depression receive minimally adequate treatment? A population-based, data linkage study. *The Canadian Journal of Psychiatry*, *61*(7), 394-404. DOI: 10.1177/0706743716640288

Querstret, D., Cropley, M., Kruger, P., & Heron, R. (2016). Assessing the effect of a Cognitive Behaviour Therapy (CBT)-based workshop on work-related rumination, fatigue, and sleep. *European Journal of Work and Organizational Psychology*, *25*(1), 50-67.

Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively–Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, *75,* 101487.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10 item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, *41*(1), 203-212.

Rebedew, D. (2018). Four Mobile Apps to Encourage Healthy Habits. *Family Practice Management*, *25*(1), OA1-OA3.

Reid, S. C., Kauer, S. D., Hearps, S. J., Crooke, A. H., Khor, A. S., Sanci, L. A., & Patton, G. C. (2011). A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomized controlled trial. *BMC family practice*, *12*, 1-14.

Richards, D., Murphy, T., Viganó, N., Timulak, L., Doherty, G., Sharry, J., & Hayes, C. (2016). Acceptability, satisfaction and perceived efficacy of "Space from Depression" an internet-delivered treatment for depression. *Internet Interventions, 5*, 12-22.

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clinical psychology review*, *32*(4), 329-342.

Roepke, A. M., Jaffee, S. R., Riffle, O. M., McGonigal, J., Broome, R., & Maxwell, B. (2015). Randomized controlled trial of SuperBetter, a smartphone-based/internet based self-help tool to reduce depressive symptoms. *Games for health journal*, *4*(3), 235-246.

Rojas, S. M., Gold, S. D., Bryan, C. J., Pruitt, L. D., Felker, B. L., & Reger, M. A. (2022). Brief cognitive-behavioral therapy for suicide prevention (BCBT-SP) via video telehealth: a case example during the COVID-19 outbreak. *Cognitive and behavioral practice*, *29*(2), 446-453.

Rost, T., Stein, J., Löbner, M., Kersting, A., Luck-Sikorski, C., & Riedel-Heller, S. G. (2017). User acceptance of computerized cognitive behavioral therapy for depression: systematic review. Journal of medical Internet research, 19(9), e309.

Royuela-Colomer, E., & Calvete, E. (2016). Mindfulness facets and depression in adolescents: Rumination as a mediator. *Mindfulness*, *7*, 1092-1102.

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. A. (2018). Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 1-23.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.

Saeri, A. K., Cruwys, T., Barlow, F. K., Stronge, S., & Sibley, C. G. (2018). Social connectedness improves public mental health: Investigating bidirectional relationships in the New Zealand attitudes and values survey. *Australian & New Zealand Journal of Psychiatry*, *52*(4), 365-374.

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB neuroscience*, *11*(2), 88-95.

Schindler, A., Hiller, W., & Witthöft, M. (2013). What predicts outcome, response, and drop-out in CBT of depressive adults? A naturalistic study. *Behavioural and Cognitive Psychotherapy*, *41*(3), 365-370.

Schroeder, N. L., Chiou, E. K., & Craig, S. D. (2021). Trust influences perceptions of virtual humans, but not necessarily learning. *Computers & Education*, *160*, 104039.

Schueller, S. M., Tomasino, K. N., & Mohr, D. C. (2017). Integrating human support into behavioral intervention technologies: The efficiency model of support. *Clinical Psychology: Science and Practice*, *24*(1), 27.

Seppala, E., Rossomando, T., & Doty, J. R. (2013). Social connection and compassion: Important predictors of health and well-being. *Social Research: An International Quarterly*, *80*(2), 411-430.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications* (Vol. 19, No. 4, pp. 451-484). London: Wiley.

Simon, N., McGillivray, L., Roberts, N. P., Barawi, K., Lewis, C. E., & Bisson, J. I. (2019). Acceptability of internet-based cognitive behavioural therapy (i-CBT) for post-traumatic stress disorder (PTSD): a systematic review. *European journal of psychotraumatology*, *10*(1), 1646092.

Six, S. G., Byrne, K. A., Aly, H., & Harris, M. W. (2022). The Effect of Mental Health App Customization on Depressive Symptoms in College Students: Randomized Controlled Trial. *JMIR Mental Health*, *9*(8), e39516.

Smith, P., Scott, R., Eshkevari, E., Jatta, F., Leigh, E., Harris, V., ... & Yule, W. (2015). Computerised CBT for depressed adolescents: Randomised controlled trial. Behaviour research and therapy, 73, 104-110.

So, M., Yamaguchi, S., Hashimoto, S., Sado, M., Furukawa, T. A., & McCrone, P. (2013). Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. *BMC psychiatry*, *13*(1), 1-14.

Steinkamp, J. M., Goldblatt, N., Borodovsky, J. T., LaVertu, A., Kronish, I. M., Marsch, L. A., & Schuman-Olivier, Z. (2019). Technological interventions for medication adherence in adult mental health and substance use disorders: a systematic review. JMIR mental health, 6(3), e12493.

Sturmey, P. (2009). Behavioral activation is an evidence-based treatment for depression. *Behavior modification*, *33*(6), 818-829.

Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018). An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR mental health*, *5*(3), e10454.

Tandon, S. D., Leis, J. A., Mendelson, T., Perry, D. F., & Kemp, K. (2014). Six-month outcomes from a randomized controlled trial to prevent perinatal depression in low-income home visiting clients. Maternal and child health journal, 18(4), 873-881.

TestFlight. (n.d.). App Store. https://apps.apple.com/us/app/testflight/id899247664

Thase, M. E., Wright, J. H., Eells, T. D., Barrett, M. S., Wisniewski, S. R., Balasubramani, G. K., ... & Brown, G. K. (2018). Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. *American Journal of Psychiatry*, *175*(3), 242-250.

Torous, J., Nicholas, J., Larsen, M. E., Firth, J., & Christensen, H. (2018). Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *BMJ Ment Health*, *21*(3), 116-119.

Tursi, M. F. D. S., Baes, C. V. W., Camacho, F. R. D. B., Tofoli, S. M. D. C., & Juruena, M. F. (2013). Effectiveness of psychoeducation for depression: a systematic review. *Australian & New Zealand Journal of Psychiatry*, *47*(11), 1019-1031.

Twomey, C., O'Reilly, G., Byrne, M., Bury, M., White, A., Kissane, S., ... & Clancy, N. (2014). A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. *British Journal of Clinical Psychology*, *53*(4), 433-450.

Umegaki, Y., Nakagawa, A., Watkins, E., & Mullan, E. (2022). A rumination-focused cognitive-behavioral therapy self-help program to reduce depressive rumination in high-ruminating Japanese female university students: A case series study. Cognitive and Behavioral Practice, 29(2), 468-484.

US Department of Veterans Affairs, V. H. A. (2013, December 12). *Va.gov: Veterans Affairs*. A Provider's Guide to Brief CBT. Retrieved October 20, 2021, from https://www.mirecc.va.gov/visn16/guide-to-brief-cbt-manual.asp.

Uzoka, A. F. (1983). Active versus passive therapist role in didactic psychotherapy with Nigerian clients. *Social psychiatry*, *18*, 1-6.

Vaghefi, I., & Tulu, B. (2019). The continued use of mobile health apps: insights from a longitudinal study. *JMIR mHealth and uHealth*, *7*(8), e12983.

Verhagen, M., Van Der Meij, A., Van Deurzen, P. A. M., Janzing, J. G. E., Arias Vasquez, A., Buitelaar, J. K., & Franke, B. (2010). Meta-analysis of the BDNF Val66Met polymorphism in major depressive disorder: effects of gender and ethnicity. *Molecular psychiatry*, *15*(3), 260-271.

Von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior*, *26*(6), 1641-1650.

Wang, M., Ma, Y., Yuan, W., Su, K., & Li, M. D. (2016). Meta-analysis of the COMT Val158Met polymorphism in major depressive disorder: effect of ethnicity. *Journal of Neuroimmune Pharmacology*, *11*, 434-445.

Wang, K., Varma, D. S., & Prosperi, M. (2018). A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. *Journal of psychiatric research*, *107*, 73-78.

Warttig, S. L., Forshaw, M. J., South, J., & White, A. K. (2013). New, normative, English-sample data for the short form perceived stress scale (PSS-4). *Journal of health psychology*, *18*(12), 1617-1628.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063.

Watts, S. E., Turnell, A., Kladnitski, N., Newby, J. M., & Andrews, G. (2015). Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. *Journal of affective disorders*, *175*, 152-167.

Wisniewski, H., Liu, G., Henson, P., Vaidyam, A., Hajratalli, N. K., Onnela, J. P., & Torous, J. (2019). Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. *Evidence-based mental health*, *22*(1), 4-9.

World Health Organization. (2022, March 2). *COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide*. World Health Organization. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

Wu, Y., Babu, S. V., Armstrong, R., Bertrand, J. W., Luo, J., Roy, T., ... & Fasolino, T. (2014). Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE transactions on visualization and computer graphics*, *20*(4), 626-635.

Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., ... & Thombs, B. D. (2020). Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. *Psychological medicine*, *50*(8), 1368-1380.

Young, C. L., Mohebbi, M., Staudacher, H., Berk, M., Jacka, F. N., & O'Neil, A. (2021). Assessing the feasibility of an m-Health intervention for changing diet quality and mood in individuals with depression: the My Food & Mood program. *International Review of Psychiatry*, *33*(3), 266-279.

Zhang, P. (2009). Theorizing the relationship between affect and aesthetics in the ICT design and use context.

Zhang, R., E. Ringland, K., Paan, M., C. Mohr, D., & Reddy, M. (2021, May). Designing for Emotional Well-being: Integrating Persuasion and Customization into Mental Health Technologies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Zhou, L., Bao, J., Setiawan, I. M. A., Saptono, A., & Parmanto, B. (2019). The mHealth App Usability Questionnaire (MAUQ): development and validation study. *JMIR mHealth and uHealth*, *7*(4), e11500.

Zhou, L., Bao, J., Watzlaf, V., & Parmanto, B. (2019). Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. *JMIR mHealth and uHealth*, *7*(4), e11223.