



Evaluating the Usability of Pervasive Conversational User Interfaces for Virtual Mentoring

Earl W. Huff Jr.¹(✉), Naja A. Mack², Robert Cummings³, Kevin Womack³, Kinnis Gosha³, and Juan E. Gilbert²

¹ Clemson University, Clemson, SC 29630, USA
earlh@clemson.edu

² University of Florida, Gainesville, FL 32611, USA

³ Morehouse College, Atlanta, GA 30314, USA

Abstract. To improve the academic and professional achievement of underrepresented minorities in computing, a newfound interest in innovative mentoring practices has captivated STEM education researchers. Studies suggest that virtual mentoring conversational agents can be leveraged across multiple platforms to provide supplemental mentorship, offsetting the lack of access to in-person mentorship in disadvantaged communities. A within-subjects mixed-method experiment was carried out to assess the usability of a mentoring conversational agent. Mobile interfaces (Twitter and SMS) were compared to each other and against a web-based embodied conversational agent (ECA). Results suggest that mobile interfaces are more usable than the web-based ECA. The findings from this study help to identify areas for improvement in virtual learning alternatives and other potential applications for pervasive conversational interfaces.

Keywords: Virtual mentoring · Underrepresented minorities · Short message service · Twitter · Embodied conversational agents · Conversational agents

1 Background

1.1 The Need for Underrepresented Minorities in Computing

The substantial demand for projected computing careers [17] reinforces the importance of better preparing underrepresented minorities to qualify for and persevere within computing careers. There is a glaring lack of Black and Latinx representation in graduate and doctoral computing programs [14]. This deficit can be attributed to an array of socioeconomic and psychosocial circumstances [39]. One of the significant perpetuations of the deficit is the lack of faculty, staff, and administrators in computing at the graduate-level to encourage and support underrepresented minorities. Such a presence would, in turn, develop students'

sense of belonging and self-efficacy [4, 35, 38, 47]. Consequently, the underperformance of underrepresented minorities in computing positively correlates with future underemployment [15].

1.2 Conversational Agents

Many initiatives have been implemented to address the achievement gap. One method of interest is the use of virtual learning alternatives such as conversational agents (also referred to as chatbots or virtual agents), which are computer programs that engage human users in natural language conversations. Conversational agents have been used for decades to facilitate effective communication with and disseminate crucial information to users in various applications and settings [6, 10, 51]. From a usability perspective, the varying content and intended user population impact the conversational agent's visual design and its effectiveness.

An embodied conversational agent (ECA) is a Web browser interface featuring an anthropomorphic interface agent that can engage with a user in real-time dialogue by employing different channels of communication such as speech and gesture, thereby emulating face-to-face human interaction [7]. ECAs have been implemented as a virtual learning alternative in various disciplines, such as psychosocial support therapy [34, 49] and academic advice [3]. Another area where ECAs have been used involves preparing African American students for graduate school in computer science [21, 22]. However, findings from a study measuring the effectiveness of an ECA that was designed to serve as a supplemental mentor for undergraduate computer science students at a Historically Black College suggested that the ECA was not an ideal channel when there is a demand for maximizing and simplifying access [32].

In the United States, 91% of the population owns a smartphone [46]. Mobile learning has been popularized as a field, focusing on the mobility of the learner, digitalizing existing analog signals and information, and personalizing instruction [44, 52]. According to Penfold [46], Mobile learning can be more advantageous than traditional practices due to convenience, flexibility, lack of travel, lack of space, personalized engagement and interaction, and the instruction distribution [46].

1.3 The SMS Platform

Short message service (SMS), otherwise known as text messaging, is the most widely used telecommunications service and an integral part of daily communication [11, 44]. There are many advantages (typically medical or health-related) to SMS being implemented into virtual learning alternatives including: reminder and appointment capability, common use, remote access, and cost-effectiveness (as compared to traditional person-to-person advisement) [1, 2, 30, 33, 36, 43, 50]. Studies, predominantly medical or health-related, also suggest SMS to be feasible and appropriate in learning and advisement interventions targeting African

Americans [9, 31, 32, 40, 48]. SMS is particularly active with teenagers and young adults, producing lingo, abbreviations, and terminology commonly used around their focal social circles [41].

1.4 The Twitter Platform

Social media platforms can also be leveraged for virtual learning alternatives. Twitter is a microblogging social networking platform that is suggested to be effective at establishing connections between students and their desired audience in a succinct form [16, 20]. The highly accessible and interactive platform contains a timeline of information covering varying topics, self-regulated content creation, content sharing, and has been suggested to be a viable source for learning [12]. Additionally, social media allows for private communications via direct messaging addressivity between users to conduct discussions similar to face-to-face conversations [28]. Twitter has been suggested to be productive in teaching and managing research for higher education students with typical same-day responses. Content searched on Twitter as opposed to other social media platforms, tends to focus more on activities and captioned photos; many of the activities are work-related such as conferences, and many of the captioned photos are inspirational quotes and advisement [42]. Very few studies that have successfully leveraged Twitter to implement virtual learning have an African American targeted population [25, 53]; however, the Twitter's notable sociopolitical activity performed by African Americans known as Black Twitter suggests underrepresented minorities are high activity and resourcefulness on the social media platform [19, 23, 24].

2 Method

Research on virtual learning alternatives for academic and career advisement, particularly for African Americans in the computer and information sciences is minimal. An expert conversational agent was developed to provide underrepresented minority undergraduate and graduate computing students with advice on how to prepare and be successful in graduate school. The conversational agent ran on three interfaces: a web-based ECA, Twitter, and SMS. The study intended to answer the following research questions:

RQ1: How do underrepresented minorities use the graduate school conversational agent?

The foundational research question was to determine the behaviors and opinions of using the conversational agent for a short term. The research team will be able to use the user experience and usability data for future tool development and improvement. Each interface should have similar responses for the base virtual mentoring system as well as interface-specific behaviors and opinions.

RQ2: How usable are the mobile conversational agent interfaces compared to the web-based ECA interface?

It is hypothesized that mobile conversational agent interfaces will have better-perceived usability than the web-based ECA. Users should feel more at ease to engage with the conversational agent through mobile interfaces than through a Web browser. As Twitter direct messaging and SMS operate similarly, it was hypothesized that users would have similar subjective usability for Twitter and SMS.

2.1 Participants

The study was conducted in two settings: (1) a computing conference where the majority of attendees were African American students, faculty, and professionals in computing; (2) a historically black college in the southeastern United States. The target population was African American students who are pursuing either an undergraduate or graduate degree in computing. Thirty-five African American students participated in the study in some capacity. All participants used at least one interface and completed a qualitative user experience assessment about the conversational agent. Twenty participants completed the assessments for user experience and usability for all three interfaces. Random sampling was used to select from interested volunteers at the computing conference. Convenience sampling was used at the historically Black college from a computer science course.

2.2 Conversational Agent Development

Virtual Mentor. The virtual mentoring system (VMS) is a computer software application developed to engage in natural language conversation with the user. The VMS is comprised of several components: (1) the content knowledgebase, (2) a natural language understanding engine, and (3) the user interface. The content knowledgebase provides the information that the VMS provides to the user based on the questions. Obtaining the content involved developing questions that users would want to ask a mentor and formulating answers that mentors would give based on the questions. The process required identifying “experts” with knowledge of applying to graduate school, funding opportunities for graduate school, and employment options for post-graduation. The experts identified work as administrators, faculty, and professionals in industry. This process is similar to the one performed in Gosha’s study for developing the first virtual mentoring system [21]. The experts’ responses were transcribed and developed into a final answer for each question added to the VMS.

At the core of the VMS, a natural language understanding (NLU) engine is necessary for the system to engage in natural language conversation with the user. For the VMS developed, Google’s Dialogflow was chosen as the NLU engine. Dialogflow, formerly Api.ai, is a natural language platform built using Google’s machine learning suite and runs on the Google Cloud Platform. Dialogflow can be used to build text-based and voice-based applications for many platforms and devices including, but not limited to, websites, smartphones, Google Assistant,

Facebook Messenger, and other devices and platforms. Dialogflow’s applicability to a vast number of platforms and devices made it an ideal choice for our engine. Integrating the expert content into Dialogflow’s engine involve mapping intended questions from the user utterances/intents to the answers provided by the experts. When a user asks the VMS a question, Dialogflow will use machine learning and natural language processing to determine which question best fit what the user is asking if it is not the same question that is stored in the system. After a match is found, the answer is returned to the VMS and displayed through the user interface (see Fig. 1 for visual representation).

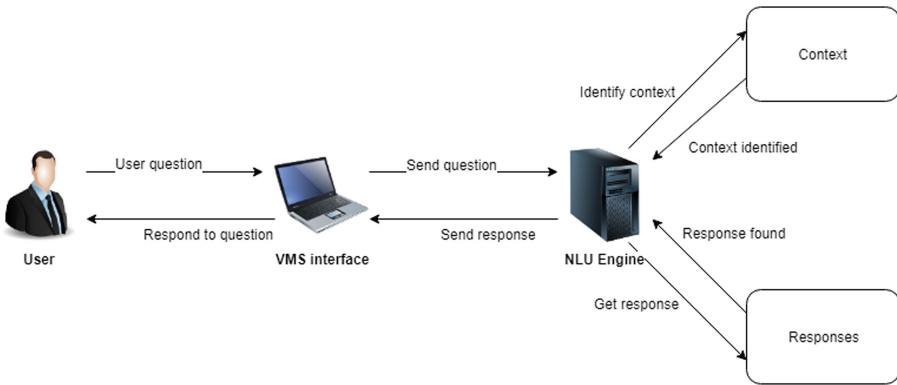


Fig. 1. VMS dialogue flow

Interfaces. The first virtual mentoring system used a graphical avatar as its interface; this is called an embodied conversational agent (ECA) [21]. The VMS designed for the present study also developed a web-based ECA to serve as the baseline. The ECA was developed using SitePal, a website the enables users to create Flash-based graphical avatars with the ability to engage in conversations with users. Once an avatar is created, it can be inserted into any website using the provided embed code. Using JavaScript code, the Dialogflow engine can be connected to the ECA. Besides the web-based ECA interface, two additional interfaces were developed: one for Twitter and one using short message service (SMS).

The Twitter-based interface uses Twitter’s direct messaging feature to engage users. Integrating Dialogflow into Twitter involves (1) creating a twitter account to serve as the virtual agent “interface” and (2) connecting it to Dialogflow through the Twitter Developer Platform. We created a Twitter account to serve as the user interface. Then we created a Twitter app on the Twitter Developer Platform. Connecting Dialogflow to the interface involves connecting the engine to the Twitter app and then connecting the Twitter app to our Twitter account. For users to engage the VMS via Twitter, the user must send a direct message to the VMS Twitter account and ask their questions.

The SMS-based interface was developed using Twilio, a cloud-based communications platform that offers APIs for building SMS, voice and messaging applications. Twilio enables users to obtain a phone number and attach a service (messaging, SMS, or voice) to that number. Setting up the SMS interface involves obtaining a phone number from Twilio, adding the Programmable SMS Messaging Service to the number, and adding the phone number, account token, and service ID to Dialogflow. To interact with the SMS interface, users merely send their question as a text message to the phone number connected to the VMS.

2.3 Study Procedure

A mixed method approach was used in this repeated measures study to explain the usability and user experience of a graduate school expert conversational agent to be used to prepare underrepresented minorities for graduate school in computing. Selected participants were provided information on how to contact the graduate school expert conversational agent: the SMS phone number, Twitter username handle, and URL to the ECA interface in a Web browser. Participants were also provided with the URL link to complete an online survey. The online survey instrument included a qualitative user experience assessment and a quantitative usability assessment. To increase participant confidentiality, each participant was provided with an identification number to use on all three assessments. The user experience questionnaire asked participants about their short-term experience using the conversational agent. Six open-ended user experience questions were assessing prominent user experience themes. These themes were determined by supporting the user experience themes established by the Interactive Design Foundation [29] (usefulness, usability, credibility, desirability, accessibility, value) and by reviewing user experience literature to further validate those themes [13,26,27,37]. Alternatively, the accessibility theme was assessed in a dichotomous yes-no question, along with if participants would recommend the tool (see Table 1). The final open-ended question was if there were more appropriate audiences for the tool to be used. The assessment also asked participants if their reaction changed their reaction from before using the conversational agent. The System Usability Scale (SUS) was another part of the online survey assessment [8]. The SUS is a 10-item questionnaire that utilizes 5-point Likert scales from strongly disagree to strongly agree. Participants were not given a time restriction nor a question limit for interacting with the system. Participants were instructed to use all three interfaces one at a time, completing the online survey immediately after they were finished using an interface. Study participants were given compensation through online gift card after completing the study.

2.4 Data Analysis

Qualitative open-ended responses were analyzed through a hybrid deductive-inductive content analysis approach [18]. The research team developed a coding guide following the user experience themes [29] including expectation/reaction

change and recommended audience (see Table 2). Four researchers coded the responses based on user experience literature; results were compared and modified excluding the theme of findability for the nature of this investigation. Data from the assessment was summarized, and themes were determined, contextualized, and synthesized. Quantitative user experience data was observed with simple descriptives (see Table 3).

A repeated-measure multilevel linear model was used to compare mobile (Twitter and SMS) interface usability to the usability of the web-based ECA. A multilevel linear model provides a number of affordances over a repeated-measures ANOVA including (1) avoiding any assumption of independence, homogeneity, and sphericity, (2) being able to use more than a two-level hierarchy and (3) able to work with missing data.

Table 1. Descriptive user experience responses

| Code | Twitter | | SMS | | ECA | |
|------------------|----------|--------|---------|--------|---------|---------|
| | Yes | No | Yes | No | Yes | No |
| Recommend tool? | 70% (14) | 30%(6) | 70%(14) | 30%(6) | 40%(8) | 60%(12) |
| Accessible tool? | 80%(16) | 20%(4) | 95%(19) | 5%(1) | 85%(17) | 15%(3) |

Note. Two user experience questions were dichotomous (yes or no). This table shows the percent of participants (number of participants) who selected yes and no per interface.

3 Results

3.1 User Experience and Usability

Demographics. There were a total of 20 participants that used all three interfaces. From the pool of participants, 50% (10) identified as Ph.D. students, 20% as Ph.D. candidates, 15% as other (postdoctoral or faculty), 10% as undergraduate seniors, and 5% as undergraduate juniors. When asked if they have ever used or interacted with a conversational agent prior to the study, 95% (19) of the participants answered ‘Yes’. Prior to the study, they were asked about the likelihood of using a system for learning about applying to graduate computing programs and careers in computing. Forty-five (45) percent indicated either likely or very likely while 35% indicated somewhat likely. After interacting with the conversational agent, 14 (70%) participants for SMS, 14 (70%) participants for Twitter, and 8 (40%) participants for the web-based ECA agreed that they would recommend the conversational agent to someone. For SMS, 19 (95%) participants agreed that they felt the conversational agent is easily accessible to everyone, compared to 17 (85%) for ECA and 16 (80%) for Twitter.

Table 2. User experience descriptions

| Code | Definition |
|----------------------|---|
| Useful | The conversational agent has a purpose and meaning allowing one to accomplish a task using the tool |
| Usable | The conversational agent is effective and efficient at achieving its objective |
| Credible | The content provided by conversational agent is trustworthy, accurate, well-intentioned, and will remain so for a reasonable amount of time |
| Desirable | The conversational agent has an attractive aesthetic and its content reflects a positive brand identity |
| Valuable | The conversational agent and its content is worth a great deal and could provide solutions to challenging and expensive problems |
| Expectation | After using the conversational agent, I had a satisfying reaction different than what i had expected prior to using the agent |
| Recommended audience | The conversational agent and its content would benefit a population |

Usability. The SUS has 10 usability items. The evaluation yielded a composite measure of the overall usability of each conversational agent. The maximum score for conversational agent usability is 100. Twitter conversational agent usability scores ranged from 47.5 to 100 with the median being 68.75. SMS conversational agent usability scores ranged from 17.5 to 100 with the median being 73.75. ECA usability scores ranged from 27.5 to 97.5 with the median being 60.

User Experience. Short responses were collected from six open-ended user experience questions from six user experience codes. The results of the hybrid thematic analysis are listed in Table 3 with themes for each code. In each interface, participants indicated that the conversational agent is useful for helping to make decisions about pursuing graduate studies and applying to graduate school. Participants reported that they could use the conversational agent *“to get another perspective and/or idea about what [they] should do to approach a task or objective”*. The conversational agent serves particularly useful *“as a starting point to begin searching for more answers. If you didn’t know anything, this at least gives you a place to start. It mentions various resources to find specific information”*. There were a few variations in usefulness per interface. Participants specifically found *“linking to URLs via the SMS”* useful as well as *“[liking] SMS because I can go back and review at a later date”*. One participant expressed that they believed the Twitter conversational agent *“would help students use social media as a way to not only stay connected but to obtain advice”*.

All participants found each interface to be easy to understand and use. There were few errors. One participant reported that *“there was a little bit of an issue for me to start using the chatbot as I was not clear on the type of questions I should be asking, but once I got the hang of it, it got much easier to use”*. One participant *“found [Twitter] more usable compared to the other two (SMS and text). I believe the platform flowed better because I was used to seeing such responses via twitter”*. Another participant noted that *“the advice was spotty, the bot would sometimes answer the question completely wrong”*. One participant said that *“the voice of the [ECA] was creepy”*. In all three interfaces, the vast majority of the participants indicated that they found the conversational agent to be an accessible tool.

For credibility, participants believed the systems’ *“answers seemed legit”*. *“The mentor seemed authentic because there were detailed explanations of the words I typed into the messaging system.”* The feedback also included *“no spelling errors, [and] professional dialogue”*. Twitter users believed the advice to be even more credible because the account profile picture: *“photo of an actual person made the advice seem more credible. Not having an artificial voice or visage made it seem like I was chatting with a real person”*. SMS did not feature a profile photo, leaving participants to express that *“the missing face/embodyment in the other two interfaces makes me trust this one a little less”*; *“I don’t know who was texting me, so that left me questioning the information.”* ECA credibility responses varied. One participant explained *“the mentor was an older woman who appeared very professional, giving a credible presentation. She also seemed relatable to me, being a woman of color”*. Another believed *“the mentor did seem credible [however] the page itself may need a little more embellishment to make me feel stronger about this”*.

Although believing the ECA was valuable was the most frequent response, there was no clear saturation of value for ECAs. Themes ranged from very valuable to not valuable. The conversational agent ranged in value from *“Eh”*, *“Not much”* to *“very valuable”* with the majority believing the conversational agent to be valuable. Some of the poorer rationales included that the system was *“personally not for me perhaps for someone who is inquiring about what to do...I wanted immediate answers and the virtual mentors didn’t provide me with such. After receiving the response, I was kinda unsatisfied because it was a chunk of text, it wasn’t too personal which I needed”*. There were a decent amount of participants who believed the content was too generic for a mentoring system: *“It was good generic advice. It’d be great if it were personalized but I realize this was our first encounter. It was almost like a Google search”*. Another participant commented on the conversational flow: *“The advice seemed genuine but the conversation didn’t flow”*. Most would agree that *“it was good for the initial steps”*.

All three interfaces met the expectations of the participants during their interaction. When asked about how their reaction to the tool changed after interacting with each interface. One participant admitted that *“my reaction slightly softened, but I still prefer human interaction”*. Another participant said *“my initial reaction was of confusion but afterwards once the interface became more*

clear, it was easier to get". Participants' experiences ranged from no change to positive for the Twitter interface. A Twitter user expressed that *"I was apprehensive, especially, to the twitter platform. but a LOT of students use twitter and this very quick interaction might just hit the mark"*. *"I like the DM interaction. It feels a little more private and personal."* Very few participants' reactions changed for the SMS interface. One participant believed that *"I was hesitant about this one, but it was okay. I think i may prefer the other two over this one as the pictures helped to feel like i was 'talking' to a 'person'"* and *"It takes time to type on the phone, but useful for mobile and travel"* while others claim to have *"liked the SMS best"*, *"I would consider it. It was the simplest version to use"*, and *"Stayed the same, this was probably the one I felt most comfortable with"*. ECA reactions ranged most greatly, being either no change, a generally positive change, or being rather surprised. One participant mentioned how *"it was faster for me to read what she was saying than wait to hear from her, which in my mind makes [the ECA] unnecessary."* Whereas another participant said *"At first, I was apprehensive to the thought of a virtual mentor, but after interacting with it, I think it was pretty cool"*.

For the all interfaces, the conversational agent was recommended for novices in the field such as high school students, school counselors, undergraduates, family members of computer scientists, and anyone looking to get into computing. The Twitter platform is recommended for active Twitter users and SMS interfaces, Most of the participants indicated that they would recommend the conversational agent to others in the Twitter and SMS interface. Most participants indicated that they would not recommend the ECA. Some recommendations for the ECA included fixing the slow-pace and freezing face glitch, considering a voice change option, making the speak button have an indicator to show it is working, ensuring the speak option does not cut off mid-sentence and to allow the ECA on a mobile platform.

3.2 Multilevel Linear Model for Interface Usability

We hypothesized that using either mobile interfaces (SMS or Twitter) would result in higher perceived usability than using the web-based ECA. It was also hypothesized that the SMS interface would result in similar usability scores as the Twitter interface. Figure 2 shows a bar chart of the means and 95% confidence interval of the SUS scores for each interface. From the chart, the SMS and Twitter interfaces have a significantly higher mean score than the ECA interface. This may indicate that participants favor using both of the mobile interfaces as compared to the ECA.

Figure 3 shows boxplots for the SUS scores of each interface. The median for SMS is the highest of the three; however, both SMS and Twitter have fairly extended boxes, suggesting the middle 50% of scores are variable. Additionally, the median score for the ECA is closer to the lower quartile (Q1), leading us to believe that the lower half of participant scores have less variation than the upper half. The long lower whiskers of both ECA and SMS interfaces indicate

Table 3. Qualitative user experience responses

| UX Codes | Twitter Themes | SMS Themes | ECA Themes |
|----------------------|--|---|--|
| Useful | To deal with specific questions To make grad school decisions To ask beginner questions | Tool saved advice when needed Useful in applying to grad school To obtain helpful resources Get quick answers To ask about funding opportunities | To make grad school decisions To ask beginner questions To create a schedule/plan To asked questions a user feels ashamed of |
| Usable | Tool was easily understandable Tool features were limited Tool did not understand question | Tool was understandable Tool was simple to use Needed instructions Tool did not converse well Tool did not understand question | Tool was easily understandable Tool was simple to use Tool had errors |
| Credible | Profile picture felt real Tool not credible Tool gave predictable, yet credible answers Ethnicity of profile felt authentic | Tool used professional language (punctuation, grammar) Tool gave predictable, yet credible answers Tool not credible Information felt believable Tool felt authentic Tool did not feel authentic Tool provided supporting resources | Tool gave predictable, yet credible answers Tool felt authentic Tool not credible |
| Desirable | Tool met expectations Tool desirable Tool was not personable enough | Tool met expectations Tool was not personable enough Information was desirable Tool not desirable Tool was repetitive Tool was ok | Tool met expectations Tool was repetitive Tool was not personable enough Tool not desirable |
| Valuable | Tool was valuable Tool somewhat valuable Tool not valuable | Tool was valuable Tool somewhat valuable Tool not valuable | Tool was valuable Tool not valuable Tool somewhat valuable |
| Expectation | Positive change Little to no change Negative change | Little to no change Faster than anticipated More comfortable than anticipated Negative change | Positive change Little to no change Positively surprised by virtual human Negatively surprised by virtual human |
| Recommended Audience | Prospective graduate students High school students Twitter users Undergraduate student Computing novices School counselors Family members of computer scientists | High school students Prospective graduate students Computing novices Undergraduate students School counselors Family members of computer scientists | High school students Undergraduate students Computing novices Prospective graduate students School counselors Family members of computer scientists |

views in the lower 25% quartile range are variable while the short upper and lower whiskers for Twitter show that scores among participants at both ends of the quartile range are less variable.

We tested our hypotheses using a repeated-measures multilevel linear model. The usability scores were dependent of the interfaces. To execute the multilevel linear model, we created two linear models: one as the baseline with perceived usability as the outcome variable and a random intercept denoting the absence of the interface variable; another one with the interface variable added to the model. This is to show if adding the interface variable has a significant overall effect. To assess if the interface-added model was a significant improvement of fit over the baseline, we observe the likelihood ratio and its corresponding p-value. The different interfaces were the independent variable and dependent variable is the perceived usability. From our tests, we can conclude that the type of

interface had a significant effect on the perceived usability of the virtual mentor, $\chi^2(2) = 8.46$, $p = .015$. While our test proved that the type of interface used affects usability, it did not tell us which interfaces had an effect.

Hence, we additionally ran planned orthogonal contrasts to ascertain the direction of the effect on usability. Two contrasts were created in order to test our hypotheses: web vs. mobile and SMS vs. Twitter. Planned contrasts revealed that perceived usability was significantly higher for mobile interfaces than for web-based ECA ($b = 6.19$, $t(38) = 2.74$, $p_{one-tailed} = .005$) and there was no significant difference between SMS and Twitter ($b = 3.13$, $t(38) = 1.2$, $p_{one-tailed} = .238$). Therefore, we accept our hypothesis that using mobile interfaces will result in higher perceived usability than using a web-based ECA and that there is no significant difference between SMS and Twitter-based interfaces.

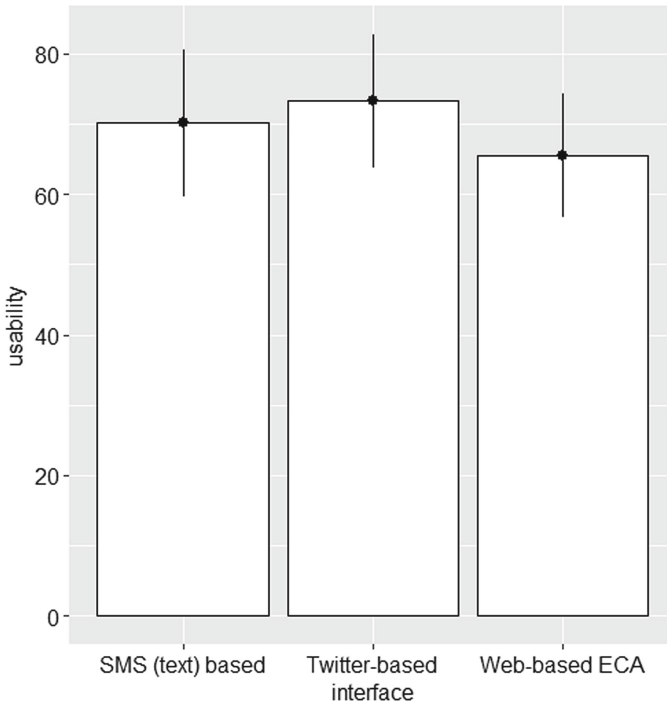


Fig. 2. Bar chart of means of usability scores by each interface

4 Discussion

4.1 Usability

It is important to note that the short-form System Usability Scale only measures the individual experience of users as an entirety. Thus individual item results of the usability scale will not be implied. According to Bangor, Kortum and Miller [5],

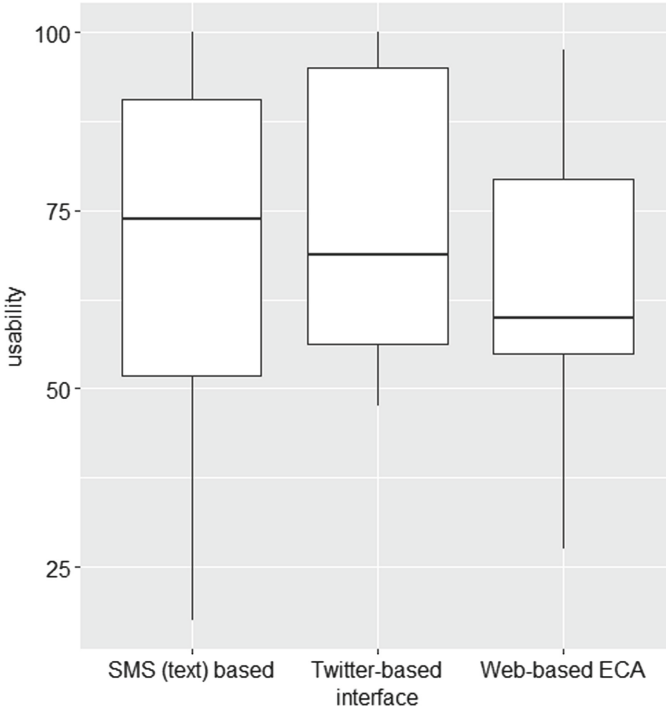


Fig. 3. Boxplots of usability scores for each interface

the SMS median usability score of 73.75, Twitter median usability score of 68.75, and ECA median usability score of 60 were all good scores, with SMS being low-range excellent. Good scores identify the conversational agent is a usable virtual mentoring system with minor defects.

The conversational agents also presented the information in a simple, easily understandable manner, with responses that can remain saved in the dialogue flow between the user and the conversational agent. This allows the user to return to the response when dealing with the context of their questions rather than having to ask an advisor every time the context of their questions arises.

The results suggest mobile (SMS and Twitter) interfaces are more usable than the web-based ECA [1,30,32]. After analyzing user experience data, some additional features that may have contributed to the lower results include the ECA's face freezing or having a delay, the voice recognition cutting off at the tail end of a user's voice input, and formatting preferences of the website. The ECA also needed an indicator to show that the ECA is listening when the users hit the speak button, whereas SMS provides clear delivery and speech notification [30]. The web-based product also has less accessibility than mobile sources due to its lack of mobile compatibility. SMS had higher usability scores than Twitter. This may be attributed to Twitter requiring users to have a Twitter

account and internet access. These data support the participants' preferences in recommending the conversational agent on mobile interfaces.

4.2 User Experience

Usefulness and Usability. The graduate school preparatory conversational agent could be used to make graduate school decisions. The conversational agent for Twitter direct messages was useful at work/school-related activities and advisement [42]. The feature of having advisement saved in SMS dialogue for remote access was also suggested be useful [1,30]. There were common themes for recommendations to improve the usefulness and usability of the tool. There is a need to clarify the instructions of usage. There were participants that were confused at what questions to ask. The tool should be better integrated with mentoring interventions and scenarios to better assist a user's understanding of the types of questions to ask the conversational agent. On Twitter, the conversational agent currently only runs through direct message, however a user reported *"I initially tried tweeting the @"*. Therefore, interface-specific instructions need to be provided to users. Conversational theory can be applied to better improve the flow of conversations. For instance *"being able to follow up on responses the bot just provided without rewriting the entire question."* is essential for a healthy conversation. Furthermore, greeting, ebonics, and abbreviations are necessary for the conversational agent to understand, particularly with mobile interfaces. Participants reported *"allow for abbreviations. People tend to abbreviate when texting"* and *"because it's through text messaging I would want to greet the mentor before engaging in a conversation. Would the system understand ebonics?"*. The ability to link resources on all interfaces can be very useful in communicating information that is length or very detailed.

Expectation. Prospective participation in virtual graduate school mentorship for minorities in computing is broad, as most everyone used conversational agents previously and 80% were at least somewhat likely to use an underrepresented minority-computing graduate school preparation conversational agent. The conversational agents were what many of the participants expected as active users of Twitter and SMS [11,24]. Users of all interfaces generally had a low-level positive change of their feelings towards the tool from before using the tool and after using the tool: *"I wouldn't mind interacting with a mentor via twitter since I already use the platform"*. There was a significant number of participants who were surprised by the ECA. Though the tool met their expectations, they still were not used to interacting with an ECA. Some participants had a positive reaction and found the moving ECA to be memorable or notable, while one participant found it to be *"freaky"*. These responses have many implications on how the ECA can be tweaked, particularly in ensuring the ECA is as personable. Personality should compliment users' comfort of seeing a character rather than just text. The image of the ECA may need to be reworked to eliminate some uncanniness.

Credibility. Credibility also varied between the interfaces. Twitter users felt the profile of an African American computer science professor made it feel like they were speaking to someone rather than a conversational agent, supporting the value of personability [44,52] and the importance of having an advisor for underrepresented minorities be of their same ethnicity [14,39]. While using SMS, participants believed the professional language the conversational agent used made the conversational agent seem more credible, supporting the informal use of texting [41].

Value. The conversational agents were generally valuable to users. As many of the participants were Ph.D. students or candidates, the valuability may have been limited, as many participants agreed the conversational agents at its current potential is more valuable to high school students, undergraduate students, or prospective graduate students. Participants pointed out that they were already familiar with the responses to many of their questions making the system credible, notably for SMS and ECA, and applicable for all interfaces. This is another supportive finding that the content needs to be more relevant to graduate students and is currently more support. It is necessary to make the content more in-depth, less generic, and more personalized in order to reflect a mentoring relationship.

Other Implications. Other themes varied. A few participants wished the conversational agents would be more personable, an implication that would support the notion that graduate school preparation needs both expert knowledge and responsive mentoring [21,44,52]. This is especially apparent due to the underrepresented minority target demographic [9,21]. Other participants mentioned how terminology and abbreviations should be better utilized within the system's knowledgebase. Other comments include usability issues and suggestions for improvement such as the conversational agent's response if it doesn't understand the user and allowing for the user to greet the conversational agent. Though many factors show SMS and Twitter to be more usable and have preferred user experiences, it is unclear why there is such a significant variance in the prospective recommendation in mobile tools and the web-based ECA.

4.3 Limitations

The research study featured a small sample size. For a simple usability assessment, the sample size is sufficient [45]. However, the comparative statistical analysis requires more population normality for validity. As the SUS and the qualitative user experience data are sound, caution should be used in observing the statistical analysis data. Data were collected from 35 participants and was used in the usability and user experience results. All 35 participants identified as Black/African American. Only 20 participants were used in the comparative analysis, including the demographic questions. This participant disparity was due to the lack of interacting and completing the assessment for all three interfaces.

User experience questions were developed from the Interactive Design Foundation's [29] seven factors that influence user experience: valuability, accessibility, desirability, usefulness, usability, findability, and credibility. As participants were not required to find the conversational agents on their own, the findability factor was omitted from the assessment. Survey items rating the accessibility of the tool and their recommended audience were asked explicitly and not a part of a validated scale. No analysis for these two factors was performed.

5 Conclusion

Mobile interfaces are a viable direction in improving the quality of mentoring conversational agents. The conversational agent has many areas where improvement could be realized, yet showed promising results for future implementation. In the near future, more in-depth conversational theory will be applied to the conversational agent to improve conversational flow. Mentoring attributes, particularly those in psychosocial areas that are critical to African Americans who are pursuing computing, will be included to the virtual mentoring system to provide a more thorough and reliable supplement for traditional mentorship. There will be additional scrutiny in the content development process to ensure potential users have the best insightful responses to their questions and concerns with pursuing graduate studies in computing. Other fields can use these findings to aid their virtual learning alternatives in their respective disciplines.

References

1. Abu Ziden, A., Rosli, M., Gunasegaran, T., Azizan, S.: Perceptions and experience in mobile learning via SMS: a case study of distance education students in a Malaysian public university. *Int. J. Interact. Mobile Technol. (iJIM)* **11**, 116 (2017)
2. Aguilera, A., Muñoz, R.: Text messaging as an adjunct to CBT in low-income populations: a usability and feasibility pilot study. *Prof. Psychol.: Res. Pract.* **42**, 472–478 (2011)
3. Andre, E., Pelachud, C.: Interacting with embodied conversational agents. In: Chen, F., Huggins, R.M. (eds.) *Speech Technology*, pp. 123–149. Springer, Heidelberg (2010)
4. Aronson, J., Fried, C., Good, C.: Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *J. Exp. Soc. Psychol.* **38**, 113–125 (2002)
5. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. *J. Usability Stud.* **4**, 114–123 (2009)
6. Benotti, L., Martinez, M., Schapachnik, F.: Engaging high school students using chatbots. In: *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education*. pp. 63–68. ACM, New York (2014)
7. Bickmore, T., Cassell, J.: Relational agents: a model and implementation of building user trust. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 396–403. ACM, New York (2001)
8. Brooke, J.: SUS: a retrospective. *J. Usability Stud.* **8**, 29–40 (2013)

9. Carter, L., Corneille, M., Hall-Byers, N., Clark, T., Younge, S.: Exploring user acceptance of a text-message base health intervention among young African Americans. *AIS Trans. Hum.-Comput. Interact.* **7**, 110–124 (2015)
10. Coniam, D.: Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL* **20**(01), 98–116 (2008)
11. Corrocher, N.: The Development of Short Message Services. <https://www.cairn.info/revue-economique-2013-1-page-149.htm>
12. Dabbagh, N., Kitsantas, A.: Personal learning environments, social media, and self-regulated learning: a natural formula for connecting formal and informal learning. *Internet and High. Educ.* **15**, 3–8 (2012)
13. Deaton, M.: The elements of user experience. *Interactions* **10**, 49 (2003)
14. Dillon, E.C., Gilbert, J.E., Jackson, J.F.L., Charleston, L.: Expanding the pipeline, the state of African-Americans in computer science: the need to increase representation. *Comput. Res. News* **27**(8), 2–6 (2015)
15. DiSalvo, B., Guzdial, M., Meadows, C., Perry, K., McKlin, T., Bruckman, A.: Workifying games: successfully engaging African American gamers with computer science. In: *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, pp. 317–322. ACM, New York (2013)
16. Evans, C.: Twitter for teaching: can social media be used to enhance the process of learning? *Brit. J. Educ. Technol.* **45**, 902–915 (2013)
17. *Fastest Growing Occupations: Occupational Outlook Handbook: U.S. Bureau of Labor Statistics.* <https://www.bls.gov/ooh/fastest-growing.htm>
18. Fereday, J., Muir-Cochrane, E.: Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *Int. J. Qual. Methods* **5**, 80–92 (2006)
19. Freelon, D., Lopez, L., Clark, M., Jackson, S.: How Black Twitter and other social media communities interact with mainstream news (2018)
20. Gikas, J., Grant, M.: Mobile computing devices in higher education: student perspectives on learning with cellphones, smartphones & social media. *Internet High. Educ.* **19**, 18–26 (2013)
21. Gosha, K.: The application of embodied conversational agents for mentoring African American STEM doctoral students (2013)
22. Gosha, K., Gilbert, J., Middlebrook, K.: Virtual graduate school mentoring using embodied conversational agents. *J. Comput. Sci. Inf. Technol.* **3**(2), 15–38 (2015)
23. Graham, R., Smith, S.: The content of our characters. *Sociol. Race Ethn.* **2**, 433–449 (2016)
24. Graham, R.: *Digital Sociology in Action: A Case Study of Research on Black Twitter.* SAGE Research Methods (2018)
25. Hampton, L., Gosha, K.: Development of a Twitter graduate school virtual mentor for HBCU computer science students. In: *Proceedings of the ACMSE 2018 Conference*, p. 42. ACM, New York (2018)
26. Hassenzahl, Marc: The thing and I: understanding the relationship between user and product. In: Blythe, M., Monk, A. (eds.) *Funology 2.* HIS, pp. 301–313. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-68213-6_19
27. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. *Behav. Inf. Technol.* **25**, 91–97 (2006)
28. Honeycutt, C., Herring, S.: Beyond microblogging: conversation and collaboration via twitter. In: *Proceedings of the 42nd Hawaii International Conference on System Sciences*, pp. 1–10. IEEE (2009)

29. Interaction Design Foundation: The 7 Factors that Influence User Experience. <https://www.interaction-design.org/literature/article/the-7-factors-that-influence-user-experience>
30. Iribarren, S., et al.: Scoping review and evaluation of SMS/text messaging platforms for mHealth projects or clinical interventions. *Int. J. Med. Inf.* **101**, 28–40 (2017)
31. James, D., Harville, C., Sears, C., Efunbumi, O., Bondoc, I.: Participation of african americans in e-Health and m-Health studies: a systematic review. *Telemed. e-Health* **23**, 351–364 (2017)
32. Julian, L., Gosha, K., Huff Jr., E.: The development of a conversational agent mentor interface using short message service (SMS). In: *Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research*, pp. 123–126. ACM, New York (2018)
33. Kaphle, S., Chaturvedi, S., Chaudhuri, I., Krishnan, R., Lesh, N.: Adoption and usage of mHealth technology on quality and experience of care provided by frontline workers: observations from rural india. *JMIR mHealth and uHealth* **3**, 61 (2015)
34. Kenny, P., Parsons, T.D., Gratch, J., Rizzo, A.A.: Evaluation of justina: a virtual patient with PTSD. In: Prendinger, H., Lester, J., Ishizuka, M. (eds.) *IVA 2008. LNCS (LNAI)*, vol. 5208, pp. 394–408. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85483-8_40
35. Khalifa, M.A., Gooden, M.A., Davis, J.E.: Culturally responsive school leadership: a synthesis of the literature. *Rev. Educ. Res.* **86**(4), 1272–1311 (2016)
36. king, C., et al.: Electronic data capture in a rural african setting: evaluating experiences with different systems in Malawi. *Global Health Act.* **7**, 25878 (2014)
37. Kuniavsky, M.: *Smart Things*. Morgan Kaufmann, Burlington (2010)
38. Kyoung Ro, H., Lattuca, L., Alcott, B.: Who goes to graduate school? Engineers’ math proficiency, college experience, and self-assessment of skills. *J. Eng. Educ.* **106**, 98–122 (2017)
39. Lewis, S., Simon, C., Uzzell, R., Horwitz, A., Casserly, M.: A Call for Change: The Social and Educational Factors Contributing to the Outcomes of Black Males in Urban Schools. <https://eric.ed.gov/?id=ED512669>
40. Le, D., et al.: Feasibility and acceptability of SMS text messaging in a prostate cancer educational intervention for african american men. *Health Inf. J.* **22**, 932–947 (2016)
41. Ling, R., Bertel, T., Sundsøy, P.: The socio-demographics of texting: an analysis of traffic data. *New Media Soc.* **14**, 281–298 (2011)
42. Manikonda, L., Meduri, V., Kambhampati, S.: Tweeting the mind and instagramming the heart: exploring differentiated content sharing on social media. In: *ICWSM*, pp. 639–642. AAAI (2016)
43. McNabb, M., et al.: Assessment of the quality of antenatal care services provided by health workers using a mobile phone decision support application in northern nigeria: a prepost-intervention study. *PLOS One* **10**, e0123940 (2015)
44. Moura, A., Carvalho, A.A.: Mobile learning: using SMS in educational contexts. In: Reynolds, N., Turcsányi-Szabó, M. (eds.) *KCKS 2010. IAICT*, vol. 324, pp. 281–291. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15378-5_27
45. Nielsen, J.: Why You Only Need to Test with 5 Users. <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
46. Penfold, S.: Why mobile learning is important for elearning. <https://www.elucidat.com/blog/why-mobile-learning-is-important/>

47. Sadler, J., Brown, M.: Focus on teacher education: culturally relevant instruction by student teachers: a case study approach. *Childhood Educ.* **91**(2), 150–154 (2015)
48. Sheats, J., Petrin, C., Darensbourg, R., Wheeler, C.: A theoretically-grounded investigation of perceptions about healthy eating and mhealth support among african american men and women in new orleans. Louisiana. *Family Commun. Health* **41**, S15–S24 (2018)
49. Tartaro, A., Cassell, J.: Using virtual peer technology as an intervention for children with autism. In: Lazar, J. (ed.) *Universal Usability: Designing Computer Interfaces for Diverse User Populations*, pp. 231–262. Wiley, New Jersey (2007)
50. Van Dam, J., et al.: An open-access mobile compatible electronic patient register for rheumatic heart disease (‘eRegister’) based on the world heart federation’s framework for patient registers. *Cardiovasc. J. Afr.* **26**, 227–233 (2015)
51. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* **9**(1), 36–45 (1966)
52. West, D.: *Mobile Learning: Transforming Education, Engaging Students, and Improving Outcomes*. The Brookings Institute, Washington (2013)
53. Xu, S., Markson, C., Costello, K., Xing, C., Demissie, K., Llanos, A.: Leveraging social media to promote public health knowledge: example of cancer awareness via Twitter. *JMIR Public Health and Surveill.* **2**, 17 (2016)