Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health

Rafal Kocielnik
rkoc@uk.edu
HCDE, University of Washington

Raina Langevin
rlangevi@uw.edu
HCDE, University of Washington

James S George
JGeorge2@dhs.lacounty.gov
Harbor-UCLA Medical Center

Shota Akenaga
shota.akenaga@gmail.com
Harbor-UCLA Medical Center

Amelia Wang
aw1998@uw.edu
HCDE, University of Washington

Darwin P Jones
darwinj@uw.edu
BIME, University of Washington

Alexander Argyle
alexanderargyle7@gmail.com
HCDE, University of Washington

Callan Fockele
cfockele@uw.edu
Emergency Medicine, University of Washington School of Medicine

Layla Anderson
anderla@uw.edu
Emergency Medicine, University of Washington School of Medicine

Dennis T Hsieh
DHsieh@dhs.lacounty.gov
Harbor-UCLA Medical Center

Kabir Yadav
KYadav@dhs.lacounty.gov
Harbor-UCLA Medical Center

Herbert C Duber
hduber@uw.edu
Emergency Medicine, University of Washington School of Medicine

Gary Hsieh
garyhs@uw.edu
HCDE, University of Washington

Andrea L Hartzler
andreah@uw.edu
BIME, University of Washington

ABSTRACT

Conversational User Interfaces (CUI) are becoming increasingly utilized in Health applications due to their ability to engage patients and support clinical workflows. Yet recent reviews show that our understanding of CUI performance and user preferences towards them is still lacking. This work examines factors that explain people's preference for engaging with a social needs screening CUI in a clinical context with 41 emergency department visitors. We demonstrate that people with low health literacy and high attitude towards emotional interaction (AEI) prefer responding to questions via CUI rather than a form-based survey. Specifically, participants with low health literacy appreciate the improved understandability offered by the CUI, whereas participants with high AEI appreciate the added level of engagement offered through conversational interactions. Our work advances the understanding of the benefits of CUI for different user groups in health contexts and beyond.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

health, conversational user interfaces, chatbots, conversational agents, healthcare, social needs, vulnerable populations, health literacy, understanding users

ACM Reference Format:
Rafal Kocielnik, Raina Langevin, James S George, Shota Akenaga, Amelia Wang, Darwin P Jones, Alexander Argyle, Callan Fockele, Layla Anderson, Dennis T Hsieh, Kabir Yadav, Herbert C Duber, Gary Hsieh, and Andrea L Hartzler. 2021. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health . In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3469595.3469599

1 INTRODUCTION

Conversational User Interfaces (CUIs) are becoming increasingly popular in supporting a variety of issues in the health context. CUIs offer multiple potential benefits, such as increasing patients' engagement through dialogue-based interactions, and helping to mitigate existing insufficiencies in the clinical workforce as they are always available. CUIs have been applied to education for therapeutic care [44], treatment support of elderly patients [17], psychotherapy training [12], sexual health & substance abuse [8] and many other settings [25]. Despite this growing interest, recent reviews have shown that our understanding of CUI performance and user preferences towards them is still lacking [29, 35, 47].

CUI '21, July 27–29, 2021, Bilbao (online), Spain

ACM ISBN 978-1-4503-8998-3/21/07…$15.00
https://doi.org/10.1145/3469595.3469599

One particularly promising use case of CUIs is to support patients with limited health literacy. Clinical trials with Embodied Conversational Agents (a form of CUIs with computer animated characters) for health education and behavior change counseling has found that these interfaces are acceptable and easy to use for patients with varying health literacy levels [3]. More recent work has explored the use of CUIs for patient screening, and demonstrated the capability of multi-modal (text & voice) CUI-based administration of social needs screening to improve understandability, engagement, and generally be a preferred interaction choice for emergency department (ED) visitors with low health literacy [26]. This is particularly important as thirty-five percent of adults in the United States have a basic or below basic level of health literacy [28].

However, while dialogue interactions may be useful to support understandability, in that work the ED visitors also reported the conversational interaction led to higher workload. The inclusion of the empathetic reactions also resulted in a significantly longer time to completion as compared to a form-based survey. This resulted in the perception that these CUIs are inefficient, especially by high health literacy patients who did not need help understanding the questions. The majority of them reported preference for the form-based survey over the CUI. Similar efficiency issues have been reported in the general application of audio computer-assisted self-interviewing [34]. Some reasons for lower interaction efficiency include the elements of CUIs that are meant to mimic human-like interactions such as additional social contents [10] and delays in CUI reactions [13]. Hence one question that arises is whether the efficiency of conversational administration can be improved to address the concerns of high health literacy users, or is it an intrinsic limitation linked to conversational features?

Furthermore, past work reported mixed perceptions of the value of social and empathetic features provided by the CUI interaction [26]. Other work in the conversational space also indicated potential user-specific differences in preference for social interaction with conversational agents [30, 31], but this aspect is under-explored in general. This raises another critical question that needs to be explored: assuming that CUIs can indeed be useful for patients with low health literacy in terms of understandability, could the empathetic CUI interactions offer value or perceived value to other types of individuals?

To investigate these questions, we began by examining the CUI for social needs screening from prior work – HarborBot [26] and explored possible improvements. We addressed a number of usability problems noted in prior work, such as improving the empathetic reactions to make sure they match the dynamic context of users' answers to sensitive questions. We further improve the interaction efficiency. We provide dynamically timed and content dependent response delays as suggested in [13], redesign the dialogue to enable convenient conversational control over the use of voice readout, and support a more use-case specific and convenient answering interface.

Then through a within-subject study we replicated the study design in prior work. We compared the redesigned CUI to a form-based social needs screening in an ED setting at two separate hospitals with 41 ED visitors. Our participants included individuals with low and high health literacy. Aside from health literacy, we also investigate the explanatory power of users' attitude towards emotional interaction on the preference for interface use (form-based vs CUI). Our findings suggest that the provided redesigns rendered CUI-based administration more engaging for all participants (an improvement over the interaction effect reported in [26]). Furthermore we found no significant efficiency differences between CUI and form-survey in terms of workload or time to completion (again an improvement over prior work which reported these measures significantly higher for CUI interface [26]). Similar to prior work, we confirm the higher preference for CUI among low literacy users. In addition we demonstrate that the attitude towards emotional interaction with CUI is also significantly and independently correlated with preference.

Our work offers several important contributions: 1) We show that individual's attitude towards emotional interaction with a CUI is a significant & independent (from literacy) factor correlated with user acceptance of CUI-based social needs screening which has implications for designing tailored CUIs in health, 2) We demonstrate that the inefficiency of conversational administration is not intrinsically linked to conversational features and can be reduced to improve experiences for all user groups, 3) We confirm and improve upon the positive impact of CUI-based social needs screening in ED setting expanding the use of the system to all user groups, and providing an important reproduction in this high-stakes setting.

2 RELATED WORK

2.1 Socialization preference in CUI context

Several past works have successfully used socialization aspects in conversational agents to engage users in interaction [46], increase trust [39], and improve user satisfaction as well as credibility, believability, and success of long-term relationships with an agent [10]. In the survey administration context as well, conversational aspects have been shown to improve the quality of user responses [23], increase engagement [26, 46], and lead to preference of this form of interaction [16]. Users expect a human-like conversational introduction and ending even in constrained task-specific uses [20] and they also appreciate a certain level of social etiquette [40]. The presence of various relational behaviors (social dialogue, empathy, liking behavior, etc.) have been applied to the design of successful engaging relational agents [2]. Depending on a specific setting, social aspects have been operationalized as emotional reactions, small talk, self-disclosure or social-emotional utterances [5].

Despite these indications, other bodies of work suggests that the social aspects are not needed and may even be detrimental. Clark et al. report that users make a very strong distinction between social and functional roles of a conversation and further question the need for social aspects, focusing more on a utilitarian use of CUIs [7]. Grudin et al. found that in a counseling setting users may appreciate a more emotionally neutral interaction to avoid a judgmental tone [15]. Several negative aspects of socialization attempts have been identified. Lucas et al. report that contrary to their expectation, an attempt at rapport building (e.g., using a conversational ice breaker) to mitigate the impact of errors, had actually backfired leading to errors being more harmful to user experience [32]. In a similar vein, several works report that the use of socialization can lead to bloated expectations which, if not met, can result in disappointment [6, 14, 33]. Svenningsson et al. tried to list factors related to perceived humanness in CUIs and explore how these may lead to a positive user experience [43]. Findings from that work actually recommend avoiding small talk and maintaining a formal tone. Such indications have even led to one recommendation of CUI design with an entirely utilitarian focus, introducing the *'thread as app'* design that prioritizes simplicity and effectiveness and sees CUIs as functional replacements of mobile applications [24].

While, several of these criticisms on the use of socialization aspects in CUI stem from current technical limitations [24, 32], or specific application areas [15], few works seem to discourage the use of social aspects on more fundamental grounds [7, 43]. Yet specifically in a conversational survey context Kim et al. found that the social conversational language is an indispensable aspect of the observed benefits [23]. For conversational social needs screening, authors also report that the use of socialization & empathy aspects was beneficial, but remark that these aspects were not universally appreciated by all the users [26]. Furthermore Liao et al. in a study of an HR bot in company settings, demonstrated that an informal user-specific measure of social agent orientation was significantly correlated with individuals' preference for socialization in their CUI [30]. Yet, beyond this work, individual differences in social CUI perception have barely been explored. What makes the matter more complicated are the numerous different types of conversational agents [11] and numerous ways in which such agents can act socially [10]. Hence we seek to answer the following research question:

> **RQ1:** Does the individual's attitude towards emotional interaction with conversational agents impact CUI use preference in social needs screening? If so, is this impact independent from health literacy?

2.2 Challenges with speed, efficiency in conversational interaction

CUI-based surveys have been reported to suffer from longer completion times and lower perceived efficiency than their form-based equivalents [23, 26, 46], which can discourage their use [26]. Perceived efficiency in the broader context of CUI-based interaction is an important underlying theme of many task-oriented uses of such systems [15, 20, 33]. Luger et al.[33] in their work on speech-based CUIs pointed to time-saving as an important consideration for many users. Jain et al.[20] in their review of 8 text-based messenger CUIs highlighted that users expect an efficient interaction. Specifically in the social needs screening ED setting, prior work reported direct user feedback indicating that such inefficiencies caused high health literacy users to prefer form-based survey over CUI, even if CUI was perceived as more engaging [26]. Some possible reasons for these perceived inefficiencies include the elements of CUIs that are meant to mimic natural human-human interactions such as additional social contents [30] and delays in CUI reactions [13]. Hence we seek to answer the following research question:

> **RQ2:** Can we improve the efficiency of interaction with CUI to satisfy all the users without sacrificing the human-likeness aspects?

3 DESIGN

We have obtained and adapted a CUI for social needs screening called HarborBot from prior work (Figure 1 - Original CUI) [26]. HarborBot has the standard chat features described in [24]. It interacts via standard elliptic chat bubbles with user messages distinguished from the bot's by different colors. It uses animated ellipses with a delay to denote the bot is typing. It supports different types of answers: *Yes/No*, *options*, *multiple answers*, and *free-text input* and provides standard confirmatory reactions to user answers. On top of that HarborBot also supports several custom features targeted towards low literacy populations in sensitive social needs setting. It provides voice readout of all the questions via text-to-speech service and affords listening to the question multiple times. It also provides additional answer options: *"explain"* to allow users to ask for clarification of the question, and *"skip"* to allow users to avoid answering an uncomfortable question. Finally it occasionally adds empathetic phrases to ease the users into answering sensitive social needs questions.

Authors in [26] tested HarborBot in ED setting with low and high health literacy participants and obtained various feedback regarding its strengths and shortcoming. Specifically they report a number of issues: 1) interaction inefficiency leading most high health literacy users to prefer the form-survey over CUI, 2) disingenuous and artificial feel partially due to CUI reactions being mismatched with the context. For use in this work, we modify the interface and interaction afforded by HarborBot to try to address these issues.

3.1 Interaction speed improvements

HarbotBot used response delays dictated by the duration of the audio readout and also with a predefined minimal length. This design choice caused users who did not benefit from auto readout (mostly high literacy participants) to still have to wait as if it was present. Furthermore, the delays were not dependent on the length of the the text itself. Hence reducing this response delay offers a clear opportunity for interaction speed improvement. Prior work indicates the value of reduced latency [18], but also warns that removing response delays altogether can be detrimental to user experience as response delays are deeply intertwined with social expectations [10, 13]. We improve the speed of interaction for users not needing audio, by calculating the delay dynamically based on the length of preceding question (Figure 1.c). We assume an average 250 words per minute reading speed based on [38].

3.2 Usability improvements

High literacy respondents in prior work also did not use the function to disable voice readout due to the option being hidden in the interface. We have included an explicit choice whether to continue with audio or not as one of the questions early on in the dialogue: *"Last question before we begin. Would you like me to continue reading the questions out loud?"* (Figure 1.a). In that case the option is more visible and users can take advantage of the faster interaction without having to wait for audio if they don't need it (9 of 21 high health literacy participants took advantage of this option in our study). They can also reinstate audio conversationally by changing their response in the same fashion as for any other question (Figure 1.b). The GUI interface option is also available.

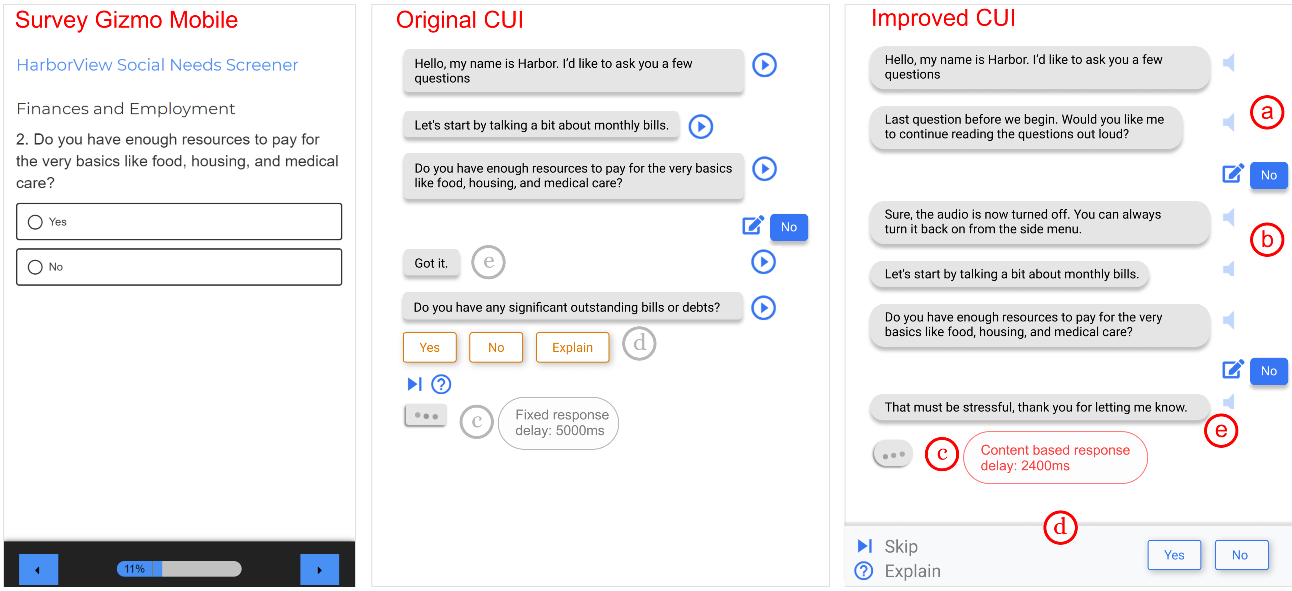

Figure 1: Survey Gizmo form-based survey rendered on a tablet (left). Original CUI interface of HarborBot from prior work (middle) and the redesigned version used in this work (right).

Another usability and efficiency change we introduced was to place answer options consistently at the bottom of the interface (Figure 1.d). Originally the answer options were presented next to the question in CUI log, which required moving the hand each time to the middle of the tablet. While this is a small improvement, in emergency department setting many visitors are tired, often with impaired attention, and sometimes even in pain. This change allows them to put less physical effort into the interaction.

3.3 Engagement improvements

Prior work used fixed reactions to user answers that were related to the question and not to the combined context of the question and answer. For example a question *"Do you have enough resources to pay for the very basics like foor, housing, and medical care?"*, would be followed by reaction *"Got it."* regardless of whether the user answer was *"Yes"* or *"No"*. At the same time a question *"Within the past 12 months the food you bought just didn't last and you didn't have money to get more. Was this:"* would always be followed by *"That must be stressful, I'm sorry to hear that."*. In most cases the original work did not use empathetic reactions to avoid such mismatches, but this might have also led to the users describing the reactions as *"it felt like defaults rather than someone 'feeling for you'"* in that work [26].

We have improved on this limitation by matching the reaction to user answers in the question context and making them empathetic whenever needed. In the improved design the question *"Do you have enough resources to pay for the very basics like food, housing, and medical care?"* would be met with reaction *"That must be stressful, thank you for letting me know."* in case the user responded *"No"*, and *"That's great to hear."* in case the user responded *"Yes"*. (Figure 1.e).

4 USER STUDY

We conducted a within-subject study with 41 ED visitors to compare the experience of answering social needs survey via CUI and form-based Survey[1]. We recruited a balanced sample of users with high and low health literacy from two recruitment sites: Greater Seattle & Los Angeles areas.

4.1 Recruitment

The ED visitors were approached with a study participation offer by the research team (not triage nurses, to avoid pressure to participate). The recruitment in the Seattle area took place at nights between 8pm and 1am local time and in Los Angeles between 4pm and 4am due to shift schedules and room availability constraints. Participants were 18 or older (except for one 12 year old participant interviewed with a parent present) and had a conversational level of English proficiency. Some participants were excluded from recruitment due to safety concerns - participants that could be violent (e.g., in handcuffs brought in by the police), or were suffering from severe pain or mental issues that could impair communication. The recruitment has been performed in coordination and with help from ED staff. The study has been approved by the IRB at both sites.

4.2 Study Procedure

The participants were taken to the visitor room located at the ED (to ensure they could hear the announcements and would not miss their turn). They were asked to read and sign the consent and then were evaluated on health literacy using REALM and NVS instruments (detailed in measures below §4.4). They then interacted with both

[1] SurveyGizmo - a popular survey platform was used: www.surveygizmo.com

interfaces (in randomized order) on a tablet's web browser. After each interaction they reported their experience and perceptions. After completing both, they were interviewed.

4.3 Social Needs Survey

Participants in both interfaces responded to a 36 question long social needs survey developed by the Los Angeles County Health Agency (LACHA)[19]. This survey has been used in prior work [26]. It includes questions related to demographics, financial situation, employment, education, housing, food, and utilities as well as questions related to physical safety, access to care, and legal needs. Some questions can be considered sensitive, such as: *"Have you ever been pressured or forced to have sex?"*.

4.4 Measures

Participants evaluated both interfaces in terms of workload (NASA TLX [41]), engagement (adapted from O'Brian's engagement [37], e.g., *"I was really drawn into answering questions."*, *"I was absorbed in answering questions."*), understandability of contents (*"I understood the questions that were asked of me."*), willingness to share information (*"I was comfortable answering the questions."*), which were used in prior studies of CUIs. We also measured attitude towards emotional interaction with CUI (AEI) (questions adapted from Negative Attitude Towards Emotional Interaction sub-scale from robotics domain [36], e.g., *"I would feel uneasy if an AI agent or chatbots really had emotions"*, *"If AI agent or chatbot had emotions, I would be able to make friends with them"*).

We evaluated participant's health literacy using Rapid Estimate of Adult Health Literacy (REALM)[9] and Newest Vital Signs (NVS)[45]; REALM measures one's ability to read health materials and instructions, at a comprehension level of high school or lower, while NVS assess likelihood of limited health literacy based on numeracy, prose and document literacy measures.

During the interview we asked about preference for each survey interface, comfort with sharing, feedback on interface features, and perceptions of CUI socialization and empathy aspects.

4.5 Participants

We recruited 41 participants (12 female, 23 male, 1 transgender, 5 declined to answer) with ages ranging from 12 to 70 (M=40.79, SD=15.74). The one participant aged 12 was interviewed in the presence of a parent. Participants reported completing an average of 11.47 years of education (SD=3.34). 31 reported English as their primary language, 4 Spanish, 1 Russian, and 5 declined to answer. Participants represented diverse ethnic backgrounds: 13 Hispanic or Latino, 9 White, 7 Black or African American, 3 Multi race, and 9 reported other ethnicity or declined to answer. In terms of health and social needs, 17 participants reported having a stable housing situation, 4 had only temporary housing, 16 were homeless, and 4 declined to answer. 19 participants reported *"Good"*, *"Very good"* or *"Excellent"* overall health, while 17 reported overall health as *"Poor"* or *"Fair"* and 5 declined to answer.

Recruitment was balanced in terms of health literacy: 21 participants were assessed as high health literacy and 20 as low. Participants were considered low literacy if they scored at a seventh to eighth grade level or below on the REALM scale, or got a score that suggests high likelihood (50% or more) of limited literacy on the NVS scale. In terms of emotional interaction preference measured with AEI, 10 participants scored below the mid-point and 18 below the median of 3.25.

4.6 Analysis

Collected quantitative data included CUI and form-survey responses and interaction logs, self-reported evaluations of each interface, as well as participant specific measures (e.g., health literacy, social interaction preference). The reliability of the measures was high (Cronbach's α for O'Brien's engagement=0.85, AEI=0.79, TLX=0.74). Statistical analysis was performed using general linear models to control for interface ordering and study site. Where appropriate user id was included as a random effect to control for repeated per user interactions.

The interviews took between 5 and 28 minutes (M=11:33, SD=6:13), conducted by five authors and then analyzed by eight authors. The interviews were audio recorded (4 participants refused recording and only notes were taken; one participant was called out before the interview could be conducted), then auto-transcribed using an online service[2] and further manually corrected.

The interviews were then distributed among the authors for qualitative analysis. The analysis process was divided into initial coding and the development of themes. In the initial coding, we limited the chance of biased interpretation by adopting and discussing a shared codebook obtained from prior work. Furthermore, each author coded a sample of interviews they had not conducted. The initial coding was reviewed among the authors and the disagreements were resolved via discussion. For the development of themes, the coded interview excerpts (with source quotes) were further used. The identified emerging themes were also discussed among the authors until a consensus was reached. This process also involved revisiting raw audio and transcriptions of interviews (e.g., to retrieve broader context as needed) to further ensure validity.

5 QUANTITATIVE RESULTS

5.1 Preference

Table 1 reports logistic regression analysis modeling interface preference based on various factors. We present 3 models including each factor separately and then in combination to communicate the stability and independent impact of the factors. Model 1 shows that interface preference is correlated with literacy level, with low health literacy (LL) participants being 11.88 times more likely to prefer CUI compared to high health literacy (HL) ones (β=2.47, χ^2=8.53, p<0.01)[3]. Model 2 shows that AEI is also significantly correlated with interface preference with each 1 point increase in AEI (on a 5-point likert scale) leading to a participant being 3.06 times more likely to prefer CUI (β=1.12, χ^2=5.47, p<0.05). Finally Model 3 shows that both health literacy and AEI are independently and simultaneously correlated with preference. In this model LL participants are 15.53 times more likely to prefer CUI (β=2.74, χ^2=8.504, p<0.01) and each 1 point increase in AEI also leads to 3.61 times more likely preference for CUI (β=1.29, χ^2=5.00, p<0.05). Across

[2]Otter.AI (www.otter.ai) was used for automated audio transcription
[3]Reported χ^2 from Wald test in logistic regression model controlling for order and site [1]

Table 1: Logistic regression analysis modeling interface preference using various factors.

	Preference					
	Model 1		Model 2		Model 3	
	Exp(β)	*S.E.*	Exp(β)	*S.E.*	Exp(β)	*S.E.*
Order [CUI]	0.27	0.80	0.41	0.73	0.28	0.89
Site [Seattle]	1.25	0.77	0.88	0.73	0.96	0.87
Literacy [Low]	**11.88****	0.85			**15.53****	0.97
AEI			**3.06***	0.48	**3.61***	0.57
Negelkerke R^2		0.37		0.28		0.53

Significance: * p<0.05, ** p<0.01

the models neither order in which the interfaces were presented nor the study site had a significant impact on preference, suggesting no systematic difference between the sites or ordering effects. Furthermore AEI is not correlated with literacy (r=0.08, p=0.63).

5.2 Engagement

Analysis of *engagement* revealed a main effect of *interface*[4] (β=0.19, $F(1, 39)$=4.652, p<0.05), with CUI reported as significantly more engaging than Survey. We also found main effect of *AEI*[4] (β=0.38, $F(1, 36)$=12.460, p<0.01), with participants rated high on AEI (HE) being more engaged than those that rated low on AEI (LE). There was no main effect of health literacy and no interaction effects.

5.3 Comfort with sharing & Understandability

Analysis of *comfort with sharing* revealed a main effect of *AEI*[4] (β=0.45, $F(1, 36)$=4.176, p<0.05), with HE participants being more willing to share than LE ones. We also found a weakly significant main effect of *health literacy* on *understandability*[4] (β=−0.41, $F(1, 36)$=3.580, p<0.1), with LL participants reporting lower *understandability* than HL ones. There were no other significant effects of literacy, AEI or interface on comfort with sharing or understandability.

5.4 Interaction Efficiency: Time to Completion, Response Equivalence & Workload

5.4.1 Time to Completion: Participants spent a similar amount of time responding to the questions via CUI (M=10:57 min, Med=8:12 min, SD=6:19 min) and Survey (M=10:13 min, Med=8:02 min, SD=8:31 min), there was no statistically significant difference. Also average completion times for LL participants (M=11:56 min, Med=9:35 min, SD=7:28 min) were not significantly different than for HL participants (M=10:04 min, Med=7:52 min, SD=5:07 min). There were no significant main or interaction effects of AEI, ordering of interface presentation, or the study site. This is an improvement over prior work [26], where CUI interaction took significantly longer.

5.4.2 Workload: There was no significant difference in workload between CUI (M=2.20, SD=1.08) and Survey (M=2.15, SD=1.16). The differences for health literacy and AEI were also not significant and there were no significant interaction effects either. This is again an improvement over prior work where CUI interaction incurred significantly higher workload [26].

5.4.3 Response Equivalence: Participants could skip answering questions in either interface, but the skip rate in CUI (M=6.80%, SD=11.15%) was not significantly different than in Survey (M=6.28%, SD=13.23%). In terms of data equivalence 85.69% (SD=12.39%) of the responses per user were the same across the two interface versions. This is comparable to the skip and equivalence rates established in prior work for this setting [26]. Manual analysis of the discrepancy causes also matched the findings from prior work.

[4]The reported main effect is from mixed-effects model controlling for order, site and interface, with user id as random effect.

6 QUALITATIVE RESULTS

Low health literacy (LL) participants prefer using CUI over Survey (17 of 20), while high health literacy (HL) participants have a slight preference for Survey (13 of 21). Also for the median-split AEI score, the high AEI (HE) participants prefer CUI over Survey (18 of 23), while low AEI (LE) participants prefer Survey (11 of 18). Table 2 summarizes the qualitative reasons for user CUI preferences.

6.1 Understandability

6.1.1 Low Health Literacy Users. Participants preferred using the CUI because it was understandable (10 out of 20) and explained that it was easy to understand the questions by listening to the audio: *"reading and hearing the voice makes processing the information easier"* (LA17). In another example, LA2 did not have their reading glasses to help them see the questions and preferred to use the CUI. While participants found the CUI more understandable due to its use of audio, one person with low health literacy said they preferred the survey because they disliked the CUI's voice, saying that it sounded robotic and creepy: *"I feel like I personally don't like the Siri voice. Creepy"* (LA4).

6.1.2 High Health Literacy Users. Participants with high health literacy may prefer familiar interactions with surveys compared to low health literate individuals who face difficulties with understanding survey questions. Participants mentioned that they would prefer to read (SE18), rather than listen to audio, and that they would become tired of listening to the voice over time (SE5). SE11 felt more comfortable using the survey since it was a familiar format for answering questions: *"like most surveys that I've done, have been in that format, so it just felt more familiar."* High health literacy participants may be capable of reading through a survey by themselves and it may not be necessary to activate the audio. Using participants' preferred mode of interaction could help to make the CUI easier and faster to use.

6.2 Perception of social utterances, reactions & human likeness

6.2.1 High Attitude Towards Emotional Interaction Users. Participants in this group preferred human-like characteristics of the CUI and felt that they had a more engaging conversation (7 out of 23). LA12 said *"it felt like you were making more of a contact...I felt like I was talking to something or somebody that was on my side"*. Participants found the reactions to be similar to empathetic responses they might receive from a real person, and it gave good feedback. One participant felt like the CUI is really considering what they

Table 2: Reasons for CUI preference across health literacy and attitude towards emotional interaction (AEI). Each participant could have mentioned more than one reason in the interview. The SE and LA next to the participant number refers to a participant from Seattle or Los Angeles areas respectively.

Median split AEI	**Low Health Literacy** (20 participants)	**High Health Literacy** (21 participants)
Low AEI (18 participants)	*5/8 participants preferred CUI* - **Understandability*** (LA2, LA11, LA17) - More entertaining (SE17, LA8) - Chat faster (SE17, LA17) - Forgot reading glasses (LA2)	*2/10 participants preferred CUI* - Convenient to have it read (LA3, LA5) - More entertaining (LA3) - Would "skim" otherwise (LA3)
High AEI (23 participants)	*12/12 participants preferred CUI* - **Understandability*** (SE1, SE7, SE14, SE16, LA6, LA12, LA19) - **Human-like feel**† (SE2, SE4, LA6, LA12, LA1) - Modern, novel (SE4, SE7, SE10) - More entertaining (LA6) - Chat faster (SE10)	*6/11 participants preferred CUI* - Convenient to have it read (SE15, LA13, SE6) - **Human-like feel**† (SE20, SE21) - Chat faster (LA13, LA18) - Modern, novel (LA18) - Irritated eye due to surgery (SE6)

∗ - Understandability is a more common reason for CUI preference among low health literacy participants

† - Human-like feel is more common as a preference reason for High AEI participants

answer: *"I feel good because I was giving them answers and they were actually giving me good advice back"* (LA19). One participant appreciated also that the CUI was understanding of their background: *"when I was answering the questions it was catching up with… my background and like situations, you know getting a better view me"* (SE20).

While two participants in the high emotional attitude group acknowledged that the CUI has limited capabilities, they also agreed that the social utterances were genuine and similar to reactions you might hear in real life. One participant noted that *"artificial intelligence at the moment lacks emotion. It doesn"t have, We don't have enough information in this day and age for them to have emotions… it gave polite responses, but it still felt robotic…It wasn't talking in a way that would express in a way that human would express. Let's put it that way."* (LA18). The social utterances were realistic statements that participants had heard before, but the nature of the conversation was robotic due to the tablet interface and participants' past knowledge of conversational agents.

6.2.2 Low Attitude Towards Emotional Interaction Users. Participants with low emotional attitude may find that chatting with an automated agent is the opposite of engaging. One participant with low emotional attitude was reminded of frustrating experiences with automated phone services when they wanted to reach a real person: *"I don't like automated voices.… In the last 20 years most companies turned to automated voices…[gives customer service examples] waiting on the phone forever.… Any business call you do on the phone, you get these automated voices… it's very difficult to get a real person on the phone"* (SE5). In this case, the human-like reactions from the CUI reminded the participant of past negative interactions where automated systems could not fill the role of a human.

6.3 Comfort with sharing

6.3.1 High Attitude Towards Emotional Interaction Users. The majority of participants with high emotional attitude (22 out of 23) felt comfortable sharing information with the CUI. Participants who were comfortable cited that they felt like they were having a real conversation, the CUI had a friendly tone and there was no pressure to answer questions. SE16 said they were comfortable sharing information because *"it was like talking to you and I"*. Participants also felt that the questions asked were typical and general questions: *"There's a lot of apps that do that [ask for information]"* (LA20). The participant who was not comfortable said they were comfortable sharing information with the CUI, but did not want their answers to be read out loud: *"I wanted whatever I was getting to the system to kind of be private. I don't care how you guys evaluate it after but I didn't want [it] sitting out loud asking me questions"* (SE10).

6.3.2 Low Attitude Towards Emotional Interaction Users. Conversely, 11 out of 18 participants with low emotional attitude felt comfortable. Five were uncomfortable sharing private information, particularly in a public setting: *"I'd be comfortable, but with the room full of people it's kind of, Yeah. It's kind of weird"* (SE11). SE19 did not feel comfortable discussing personal information with the CUI because they were not sure who they were conversing with: *"I don't want to chat about this kind of question about sex because I don't know sure if it's like a robot or not"*. One participant was annoyed with the CUI's automated voice. When using the CUI, it may be important to make clear to participants with low preference for emotional interactions who they are interacting with and where their information will be sent. Designing CUIs to resemble real interactions may cause individuals with low attitude towards emotional interaction to feel more uneasy about the conversation.

7 LIMITATIONS

There are a few limitations of our work that we need to acknowledge. First, the specific socialization design relying on empathetic reactions is particularly suited for our setting and the findings may not generalize to all CUI designs. However, given common challenges of sensitive topics and low literacy patients in health settings, we believe our findings to be informative for more general health-related applications. Second, due to the availability of rooms and hospital processes we could only recruit ED visitors at late hours, which could have effected their focus and attention. Third, researchers were present in the room with the participants, which could have introduced desirability bias particularly in the ED setting where patients seek care. Finally, some of the participants had limited exposure to conversational interfaces before the study, which could have introduced novelty effects.

8 DISCUSSION

In this work, we demonstrate that CUI based social needs screening can offer a significantly more engaging interaction for all ED visitors compared to form-based administration, while maintaining high response equivalence of 86% (in line with prior work [26, 34]). We show that a CUI is significantly more preferred among low literacy users due to the understandability benefits it provides (replicating and confirming results from [26]). The CUI is further preferred among users reporting high attitude towards emotional interaction with CUI (AEI) due to its social & empathetic qualities. We also directly improve on the prior work in this context [26], by showing that the CUI interaction efficiency can be improved to also satisfy high health literacy users without sacrificing the above valuable qualities. We further discuss these findings and their implications for the broader application of CUIs.

8.1 CUI Preference in Relation to Emotional Interaction Attitude

Our main finding in this work is that CUI preference is significantly correlated with both health literacy and emotional interaction attitude towards CUI (measured by AEI [36]). Prior work established the predictive power of health literacy in an ED context [26]. Our current findings offer an important confirmation and replication of these results in this high-stakes context. Furthermore we found that AEI is not correlated with health literacy and constitutes an independent significant factor. Our qualitative findings also show that the reasons for preferring CUI among users with low health literacy emphasize understandability benefits more, while users with high AEI tend to focus more on human-likeness aspects.

Aside from preference, CUI interaction was more engaging for all participants regardless of their literacy or AEI score. This highlights the engagement benefits of CUIs, even if users don't prefer it for understandability aspects. Another interesting finding relates to the High AEI participants being significantly more *willing to share* and significantly more *engaged* in general (i.e., main effect). One possible explanation could be that the attitude towards emotional interaction with CUI is indicative of a more general need for emotional connection among these users. This need combined with specific sensitive and personal nature of the social needs questions themselves could lead to such higher scores (i.e., these users resonated with the sensitive questions more, regardless of whether asked by CUI or form-survey).

8.2 CUI Interaction Efficiency

The interaction speed and usability improvements we introduced to the CUI in this work seem to have improved the issues reported in [26]. We found that our CUI does not introduce higher workload and does not significantly lengthen time to completion. While qualitatively users still reported the CUI interaction to feel a bit longer, which is arguably unavoidable due to additional conversational contents, this perception was not strong enough to significantly affect efficiency measures and dictate user preferences. With this result we show that careful usability design (e.g., accessibility of answer options), timing of reactions (i.e., content-based & reading speed based delays [13]) as well as visibility of choices (i.e., voice readout setting integrated into the conversation to make it easier for users not to stick with a default setting [21]) can effectively balance efficiency with human-likeness and understandability features. It is interesting that only 9 of 21 high health literacy users decided to disable voice readout when asked about it early in the CUI interaction. Some of these users later still complained about this feature causing efficiency issues. It is possible that due to the novelty effect, users decided to stick to voice due to curiosity even if they did not need it.

8.3 Future Directions

There are several aspects that could be investigated in future work. First, the AEI measure relates to the perception of the emotional interaction with CUI which we adapted from the robotics domain [36]. This measure could be shaped by the user's prior experiences, especially with technology [27]. As such it could be related to some more fundamental characteristic such as tech savviness [4], value orientation [42] or a personality trait [48]. Future work could investigate this aspect further. Second, in our work we uncovered user characteristics significantly correlated with CUI preference, but we did not act on them in the design. Future work could explore a tailored design that could adjust CUI empathy level to such preferences. This could be accomplished in multiple ways: 1) as different interface alternatives the user could pick from, e.g., empathetic CUI or utilitarian CUI before the interaction starts. 2) through integration of AEI or similar questions into the interaction itself to quickly assess preferences (similar fashion as we integrated question about audio in our redesign). 3) finally, one could attempt to assess the preference from interaction behavior itself, similar to the personality assessment in [22, 46].

9 CONCLUSION

In this work, we have designed and evaluated a CUI for application in Health. In a within-subject study with 41 ED visitors at two sites, we have compared a CUI to web administered social needs screening. We demonstrate the significant preference of low health literacy users for CUI use, confirming and replicating findings from prior work. Improving on the prior work we demonstrate that with several usability and efficiency redesigns as well as enhancements in empathy design, the CUI interaction can be more engaging for all users (including those with high health literacy) and as efficient as a form-based survey in terms of workload and time to completion. We further demonstrate that attitude towards emotional interaction (AEI) is an additional and independent (from health literacy) factor correlated with CUI preference. This highlights the empathetic interaction benefits CUIs can offer to high health literacy users who may not need the understandability features. Our work advances the understanding of the benefits of CUIs for different user groups in health contexts and beyond.

ACKNOWLEDGMENTS

We would like to thank the triage nurses and all the emergency department staff at the Seattle and Los Angeles locations for all their help and for facilitating the recruitment process. We would also like to thank all the emergency department patients who participated in the study for their help, time, and feedback. This project was supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1 TR002319. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] Viv Bewick, Liz Cheek, and Jonathan Ball. 2005. Statistics review 14: Logistic regression. *Critical care* 9, 1 (2005), 1–7.
[2] Timothy Bickmore and Justine Cassell. 2005. Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems*. Springer, 23–54.
[3] Timothy W Bickmore, Laura M Pfeifer, Donna Byron, Shaula Forsythe, Lori E Henault, Brian W Jack, Rebecca Silliman, and Michael K Paasche-Orlow. 2010. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. *Journal of health communication* 15, S2 (2010), 197–210.
[4] Kelly Chandler-Olcott and Donna Mahar. 2003. "Tech-savviness" meets multiliteracies: Exploring adolescent girls' technology-mediated literacy practices. *Reading Research Quarterly* 38, 3 (2003), 356–385.
[5] Ana Paula Chaves and Marco Aurelio Gerosa. 2020. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human–Computer Interaction* (2020), 1–30.
[6] Leon Ciechanowski, Aleksandra Przegalinska, Mikolaj Magnuski, and Peter Gloor. 2019. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems* 92 (2019), 539–548.
[7] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
[8] Rik Crutzen, Gjalt-Jorn Y Peters, Sarah Dias Portugal, Erwin M Fisser, and Jorne J Grolleman. 2011. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. *Journal of Adolescent Health* 48, 5 (2011), 514–519.
[9] Terry C Davis, Sandra W Long, Robert H Jackson, EJ Mayeaux, Ronald B George, Peggy W Murphy, and Michael A Crouch. 1993. Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Family medicine* 25, 6 (1993), 391–395.
[10] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies* 132 (2019), 138–161.
[11] Sherwin Fernandes, Rutvij Gawas, Preston Alvares, Macklon Femandes, Deepmala Kale, and Shailendra Aswale. 2020. Survey on various conversational systems. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. IEEE, 1–8.
[12] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 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 (2017), e19.
[13] Ulrich Gnewuch, Stefan Morana, Marc TP Adam, and Alexander Maedche. 2018. Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction.. In *ECIS*. 113.
[14] Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2017. Towards Designing Cooperative and Social Conversational Agents for Customer Service.. In *ICIS*.
[15] Jonathan Grudin and Richard Jacques. 2019. Chatbots, humbots, and the quest for artificial general intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.
[16] S Harms, C. & Schmidt. 2017. Conversational Survey Frontends: How ChatbotsCan Improve Online Surveys. https://www.researchgate.net/publication/315199377_Conversational_Survey_Frontends_How_Can_Chatbots_Improve_Online_Survey. [accessed Feb 04 2021].
[17] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Relating conversational expressiveness to social presence and acceptance of an assistive social robot. *Virtual reality* 14, 1 (2010), 77–84.
[18] Thomas Holtgraves and Tai-Lin Han. 2007. A procedure for studying online conversational processing using a chat bot. *Behavior research methods* 39, 1 (2007), 156–163.
[19] Liu P Johnson S Hong C, Hsieh D. 2018. "White paper: Development of a social and behavioral determinants of health short screener.". (2018).
[20] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N Patel. 2018. Evaluating and informing the design of chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 895–906.
[21] Liyin Jin. 2011. Improving response rates in web surveys with default setting: The effects of default on web survey participation and permission. *International Journal of Market Research* 53, 1 (2011), 75–94.
[22] Yuta Katsumi, Suhkyung Kim, Keen Sung, Florin Dolcos, and Sanda Dolcos. 2017. When nonverbal greetings "make it or break it": the role of ethnicity and gender in the effect of handshake on social appraisals. *Journal of Nonverbal Behavior* 41, 4 (2017), 345–365.
[23] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.
[24] Lorenz Cuno Klopfenstein, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo. 2017. The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 conference on designing interactive systems*. 555–565.
[25] A Baki Kocaballi, Juan C Quiroz, Liliana Laranjo, Dana Rezazadegan, Rafal Kocielnik, Leigh Clark, Q Vera Liao, Sun Young Park, Robert J Moore, and Adam Miner. 2020. Conversational Agents for Health and Wellbeing. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.
[26] Rafal Kocielnik, Elena Agapie, Alexander Argyle, Dennis T Hsieh, Kabir Yadav, Breena Taira, and Gary Hsieh. 2019. HarborBot: A Chatbot for Social Needs Screening. In *AMIA Annual Symposium Proceedings*, Vol. 2019. American Medical Informatics Association, 552.
[27] Rafal Kocielnik, Saleema Amershi, and Paul N Bennett. 2019. Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
[28] Mark Kutner, Elizabeth Greenburg, Ying Jin, and Christine Paulsen. 2006. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. NCES 2006-483. *National Center for Education Statistics* (2006).
[29] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258.
[30] Q Vera Liao, Matthew Davis, Werner Geyer, Michael Muller, and N Sadat Shami. 2016. What can you do? Studying social-agent orientation and agent proactive interactions with an agent for employees. In *Proceedings of the 2016 acm conference on designing interactive systems*. 264–275.
[31] Q Vera Liao, Muhammed Mas-ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, N Sadat Shami, and Werner Geyer. 2018. All work and no play?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[32] G. Lucas, J. Boberg, D. Traum, Ron Artstein, J. Gratch, Alesia Gainer, E. Johnson, A. Leuski, and M. Nakano. 2017. The Role of Social Dialogue and Errors in Robots. *Proceedings of the 5th International Conference on Human Agent Interaction* (2017).
[33] Ewa Luger and Abigail Sellen. 2016. " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286–5297.
[34] David S Metzger, Beryl Koblin, Charles Turner, Helen Navaline, Francesca Valenti, Sarah Holte, Michael Gross, Amy Sheon, Heather Miller, Philip Cooley, et al. 2000. Randomized controlled trial of audio computer-assisted self-interviewing: utility and acceptability in longitudinal studies. *American journal of epidemiology* 152, 2 (2000), 99–106.
[35] Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. 2019. Survey of conversational agents in health. *Expert Systems with Applications* 129 (2019), 56–67.
[36] Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and Kensuke Kato. 2008. Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE transactions on robotics* 24, 2 (2008), 442–451.
[37] Heather L O'Brien and Elaine G Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69.
[38] The University of Chicago Student Health and Counseling Services. 2018. Speed Reading. https://web.archive.org/web/20180307083117/https://wellness.uchicago.edu/page/speed-reading/. [Online; Archived from the original on 7 March 2018.

Retrieved 30 December 2017.].

[39] Manuel Portela and Carlos Granell-Canut. 2017. A new friend in our Smartphone? Observing Interactions with Chatbots in the search of emotional engagement. In *Proceedings of the XVIII International Conference on Human Computer Interaction*. 1–7.

[40] Susan Robinson, David R Traum, Midhun Ittycheriah, and Joe Henderer. 2008. What would you Ask a conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting.. In *LREC*.

[41] Susana Rubio, Eva Díaz, Jesús Martín, and José M Puente. 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied psychology* 53, 1 (2004), 61–86.

[42] Shalom H Schwartz. 2012. An overview of the Schwartz theory of basic values. *Online readings in Psychology and Culture* 2, 1 (2012), 2307–0919.

[43] Nina Svenningsson and Montathar Faraon. 2019. Artificial intelligence in conversational agents: A study of factors related to perceived humanness in chatbots. In *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*. 151–161.

[44] Catharine Wang, Timothy Bickmore, Deborah J Bowen, Tricia Norkunas, MaryAnn Campion, Howard Cabral, Michael Winter, and Michael Paasche-Orlow. 2015. Acceptability and feasibility of a virtual counselor (VICKY) to collect family health histories. *Genetics in Medicine* 17, 10 (2015), 822–830.

[45] Barry D Weiss, Mary Z Mays, William Martz, Kelley Merriam Castro, Darren A DeWalt, Michael P Pignone, Joy Mockbee, and Frank A Hale. 2005. Quick assessment of literacy in primary care: the newest vital sign. *The Annals of Family Medicine* 3, 6 (2005), 514–522.

[46] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 3 (2020), 1–37.

[47] Zhaopeng Xing, Fei Yu, Jian Du, Jennifer S Walker, Claire B Paulson, Nandita S Mani, and Lixin Song. 2019. Conversational interfaces for health: bibliometric analysis of grants, publications, and patents. *Journal of medical Internet research* 21, 11 (2019), e14672.

[48] Lisa M Pytlik Zillig, Scott H Hemenover, and Richard A Dienstbier. 2002. What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin* 28, 6 (2002), 847–858.