

What Is the Difference? Investigating the Self-Report of Wellbeing via Conversational Agent and Web App

Raju Maharjan , Darius A. Rohani , Kevin Doherty , Per Bækgaard , and Jakob E. Bardram , *Technical University of Denmark, 2800 Kongens Lyngby, Denmark*

Adopting speech as their mode of interaction, speech-enabled conversational agent (CA) systems hold the potential to enable more natural and engaging self-report experiences than traditional media (e.g., pen and paper, web, or smartphone systems). Our recent research concerns the potential design and use of CAs to support mental health and wellbeing. In this article, we present findings from a study during which 22 individuals with affective disorders (e.g., depression, bipolar disorder) used either a speech-enabled CA or web app to self-report their emotional wellbeing. Analysis of users' experiences and engagement with the system for daily self-report suggests that despite many technical limitations, users rated self-reporting via CA as a significantly novel and attractive experience yielding similar rates of engagement to a traditional web-based method.

Mental health self-report tools such as diaries and health questionnaires are increasingly adapted to digital platforms including mobile and web apps (WAs) in line with the widespread general use of the Internet and smartphones. Due to the accessible nature and efficiency of these graphical user interface (GUI) based systems for the collection of self-reports, they are often presented as viable means of overcoming the limitations of more traditional methods including pen and paper and in-person interviews.

GUI-based systems, however, also and often limit users' capacity for self-expression, and can likewise place a reporting burden on users resulting in negative self-report experiences.¹ This burden can prove particularly substantial in the case of open-ended questions potentially leading to a decline in the use of the system over time and compromising the quality and reliability of the data due to the additional time and effort required to formulate responses.²

In contrast, recent advancements in voice recognition technology, natural language processing (NLP), and artificial intelligence (AI) have made more engaging and "natural" forms of interaction possible via speech interfaces. In recent years, speech-enabled conversational agents (CAs) in particular have grown in popularity, as reflected in the increasing trend in ownership of smart-speakers such as Google Home and Amazon Echo.³ The growing number of health and wellness applications available through these systems further reflects an increasing interest in CAs' potential to support healthcare.⁴ A recent survey of 1004 U.S. adults reported that 52% possessed an interest in using CAs, while 7.5% had already made use of such systems for a healthcare-related task,⁵ including the self-report of mental health and wellbeing.

Research suggests that speech-enabled CAs can engage users in interactions of greater duration⁶ and improve self-reporting experiences due in part to the more "natural" and conversational forms of interaction afforded by these systems.⁷

In light of these findings, we have been investigating the design and use of speech-enabled CAs for mental healthcare, primarily to support the self-report of mental health and wellbeing, as key to the diagnosis and treatment of many mental health conditions. This article reports findings from a four-week between-group "in-the-wild" study comparing users' self-report

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>
Digital Object Identifier 10.1109/MPRV.2022.3147374
Date of publication 23 March 2022; date of current version 28 June 2022.

experiences and engagement with a bespoke speech-enabled CA and a WA. This study engaged 22 individuals living with either depression or bipolar disorder, half of whom used a CA deployed via smart speaker during the study, and the other half a WA via their own personal desktop or smartphone to log their emotional wellbeing.

Findings from this study suggest that speech-enabled CAs can serve as a more engaging method for the self-report of mental health and wellbeing yielding similar rates of engagement compared to a more traditional GUI-based method.

CONVERSATIONAL AGENTS FOR THE SELF-REPORT OF MENTAL HEALTH AND WELLBEING

In recent years, the growing popularity of smart-speakers such as Google Home and Amazon Alexa has enabled researchers to investigate users' lived experiences of CAs in their home settings. One such study entailed a comparison of a custom-designed CA named "Robota," deployed as both a chat-based Slack Bot and a speech-based Alexa Skill, which asked users to journal their workplace activity by answering ten open-ended questions daily.⁷ This three-week field study examined the impact of text- and speech-based platforms on participants' reflections and self-learning, showing that speech interaction enabled users to better reflect on their work and provided opportunities for workplace-related behavior change, despite many technical limitations. Quiroz *et al.* similarly developed an Alexa Skill that enabled users to self-report their emotions, complete self-assessments for depression and anxiety, and receive suggestions for improving their mental state.⁸ Results from a pilot study of this system indicated acceptance and trust among users for the technology's adoption for the self-report and assessment of mental health.

While CAs as a medium for self-report therefore have the potential to improve reporting experiences, these technologies also have their limitations.

Kocielnik *et al.*,⁷ for example, found that Amazon's Alexa Skills allow a maximum of 12 s for each user response, limiting opportunities for open-ended self-reports.

CAs' inability to interpret or recognize free utterances have likewise been reported to cause confusion during conversations leading to further errors.⁹ For depression-related utterances, for example, a sensitivity of 80% was reported along with a positive predictive value of 83%, and for clinician-identified harm-related sentences a word error rate of 34%.¹⁰

Most CAs are not capable of engaging in dynamic conversation, and users, therefore, often have to know in advance what they can or cannot say to the CA, frequently making it cognitively demanding to interact.¹¹

Combined, the conceptual gap between CAs' capacities and users' expectations of emotional exchanges, relationship building, and human-human like conversations have been shown to lead to drastic declines in the use of CAs.¹²

Findings from these studies, on the one hand, suggest that speech-enabled CAs can improve users' self-reporting experiences and potentially address the significant challenge of sustaining user engagement associated with GUI-based mental health self-report technologies. On the other hand, CAs' current technical limitations might seriously and adversely impact users' self-report experiences and behaviors, undermining their potential advantages.

CONVERSATIONAL AGENT SOFIA

To investigate users' experiences and behaviors in using a speech-enabled CA to self-report their mental health and wellbeing, we developed a CA named *Sofia*. It enabled users to both maintain a daily journal of their feelings by responding to a series of open-ended questions and respond to the World Health Organization-Five Well-Being Index (WHO-5)¹³ questionnaire in Likert scale form fortnightly.

We combine open-ended questions with the standardized WHO-5 scale in order to enable users to experience different types of self-report, and in turn explore and potentially address through design the current and diverse limitations of CA technologies. We chose to adopt the WHO-5 questionnaire as it is a well-established health questionnaire posing five simple questions and containing less invasive questions than many clinical scales.

Figure 1 provides an overview of the conversational design of *Sofia*. These conversational dialogues were designed with the involvement of the authors and a single external human-computer interaction (HCI) researcher; using a "Wizard of Oz" method to test the design with three university students. The final design of the conversation enabled users to freely express their emotional wellbeing by responding to three open-ended questions starting with a greeting. Multiple variations of the questions were designed to reduce the repetition of the conversation.

To render the WHO-5 questionnaire more conversational in nature, slight variations in the wording of the questions and preamble were incorporated.

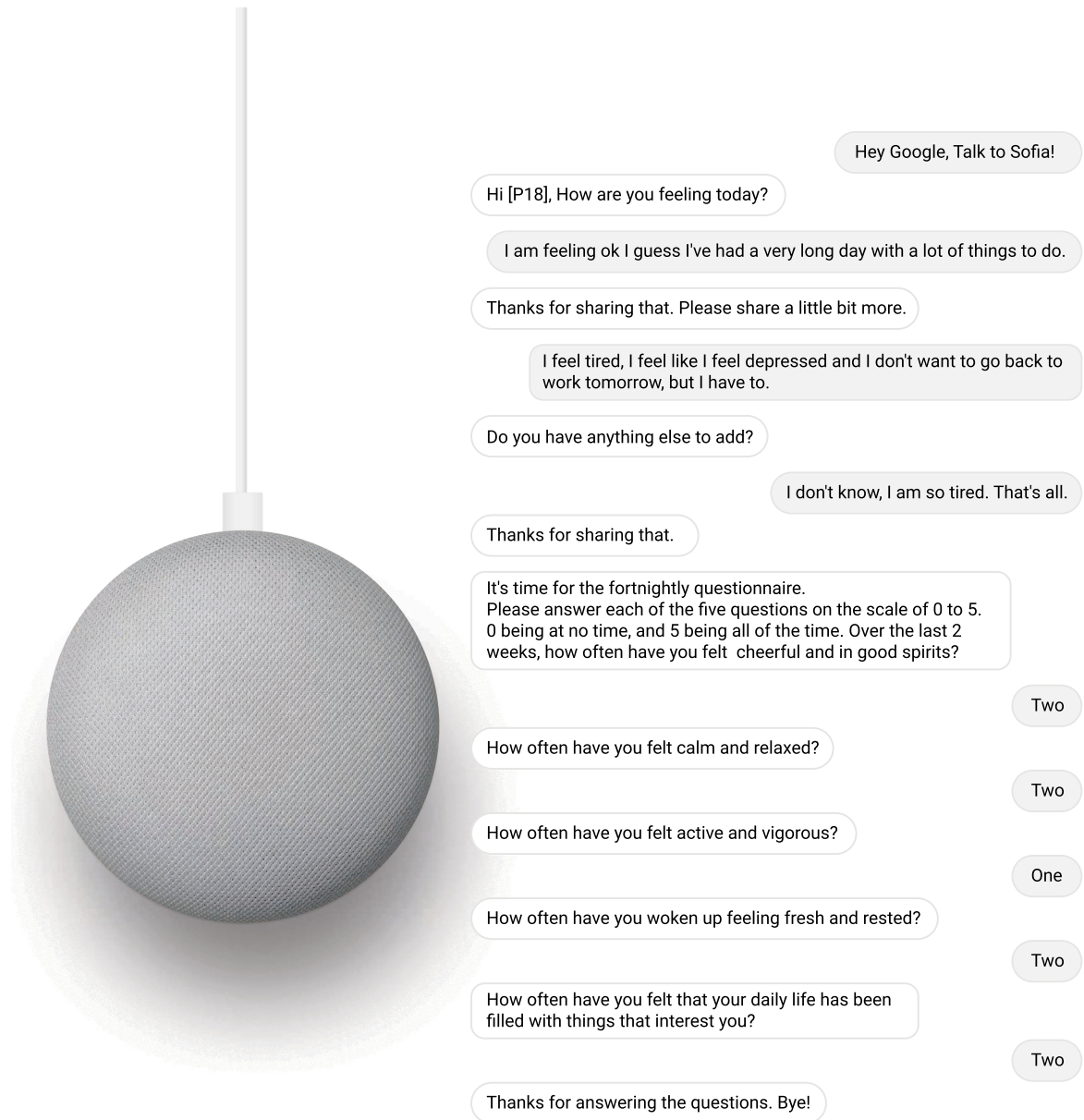


FIGURE 1. Sample conversation between P18 and the CA.

Sofia handles “no response” (respondent takes too long to respond) and “no match” (Sofia does not understand the respondent) errors by providing fallback prompts. It also repeats the question when a user states “repeat” or “what was the question?,” and reminds the user of the response options by repeating the preamble (e.g., “You can answer the question on the scale of 0 to 5; 0 being at no time, and 5 being all of the time”) during the fortnightly WHO-5 questionnaire when asked for “help.”

The conversational design was implemented using Dialogflow¹⁴ and deployed via Google Nest Mini smart speakers.

WEB APP

To enable us to compare users’ mental health self-report experiences via CA with more traditional methods, we also developed a WA implementing the CA’s content entry features. As shown in Figure 2, this simply comprised a text-field labeled with the daily

How are you feeling today?

SUBMIT

Please indicate for each of the five statements which is closest to how you have been feeling over the last two weeks. Notice that higher numbers mean better well-being. Example: If you have felt cheerful and in good spirits more than half of the time during the last two weeks, put a tick in the box with the number 3 in the upper right corner.

Over the last two weeks	All of the time	Most of the time	More than half of the time	Less than half of the time	Some of the time	At no time
I have felt cheerful and in good spirits	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0
I have felt calm and relaxed	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0
I have felt active and vigorous	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0
I woke up feeling fresh and rested	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0
My daily life has been filled with things that interest me	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0

SUBMIT

FIGURE 2. WA for the self-report of mental health and wellbeing. Interface for open-ended diary entry on the left. Fortnightly WHO-5 questionnaire on the right.

open-ended question “How are you feeling today?” and a submit button. To keep the design of the WA as comparable to the CA as possible, the open-ended question in the WA was presented in three different variations (“How have you been feeling today?,” “How are you feeling today?,” and “How do you feel today?”), each time the user logged into the WA. In addition to this daily entry, users were also asked to respond to the WHO-5 questionnaire fortnightly. To ensure reliable collection of these self-reports, appropriate data validation methods including user accounts and error messages were incorporated into the app.

IN-THE-WILD STUDY

We then conducted a four-week study “in the wild” comprising a between-group study design.

Participant Recruitment

Inclusion criteria for participating in this study required participants to:

- 1) be over the age of 18;
- 2) identify as diagnosed with an affective disorder (AD); and
- 3) have continuous access to WiFi.

A total of 22 participants were recruited through the national recruitment site <http://www.forsoegsperson.dk>, social media, university internal email, and posters.

This study was exempt from ethical approval by the Danish National Committee on Health Research Ethics in accordance with Section 14 (2) of the Danish Act on Research Ethics Review of Health Research Projects, as states that “Notification of questionnaire

surveys and medical database research projects to the system of research ethics committee system is only required if the project involves human biological material.”¹⁵

Procedure

Prior to the study, all participants were asked to sign a consent form and provide demographic information including their name, email, age, gender, education level, employment, year when diagnosed with the AD, symptoms last experienced, other known health conditions, and technical ability.

Participants were randomly assigned to one of the two systems. Participants assigned to use the CA (Group CA) were provided with a Google Nest Mini 2 smart speaker and walked through the setup process. This involved pairing the smart speaker with the Google Home app on their personal mobile phone, creating an account with the CA, and going through a sample conversation. Each participant using the WA (Group WA) was provided with a unique ID with which to log into the WA and asked to access the WA on their own devices.

Each group was provided with corresponding handouts describing instructions on how to use the system, information on the data collected through the system, and expectations of participants during the study. Participants were encouraged to adopt their own preferred methods to remind themselves to report their mental health each day. CA participants were additionally made aware of the Google Assistant’s “Routine” feature and the possibility of setting up a reminder by asking the assistant to remind them to speak to Sofia.

The CA handout included information regarding invocation phrases, instructions for asking *Sofia* to repeat the question and end the conversation, as well as details of what *Sofia* could and could not do.

The handout for the WA contained the uniform resource locator (URL) of the app, suggested browsers for users accessing the WA, and instructions for logging in and submitting self-reports via the app.

Data collected through the systems included participants' responses to the open-ended questions and WHO-5 questionnaires and timestamps of their interactions. CA-based self-reports were automatically transcribed and stored in the database.

On completion of the study, participants in each group were asked to fill out the user experience questionnaire (UEQ) scale¹⁶ as a subjective assessment of their self-report experiences.

CA participants were granted the possibility of keeping the device used during the study. And WA participants were offered a Google Nest Mini device or a gift card corresponding to Danish Kroner 300 (~US\$50) for their participation.

Measures

We assessed participants' self-report experiences based on their responses to the UEQ scale, and behaviors in terms of their engagement with the system for the daily practice of self-report and over time.

As additional context for users' self-report behaviors, we furthermore analyzed participants' WHO-5 scores collected and performed text and sentiment analysis of the open-ended responses.

We found that Google Action allowed a maximum of 12 s for each open-ended user response which significantly impacted self-report duration and the number and frequency of the words used to express their emotions. In contrast, the WA supported unlimited text input over any period of time. For these reasons, we did not consider self-report duration nor the word count of the open-ended self-reports as primary outcome measures in this case although prior studies of a similar nature have used such data as indicators of user engagement.⁶

User Experience Questionnaire

The UEQ is a fast and reliable questionnaire for the measurement of the user experience of interactive products, consisting of 26 items representing six key user experience dimensions; attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Each item of the questionnaire describes users' subjective assessment of their experience using the

system and is rated on a 7-point Likert scale randomly ordered from positive to negative extremes.

Engagement

Across the HCI literature, engagement has been defined in relation to a wide variety of contexts and technologies, and often tied to measurement in practice.¹⁷ For the purposes of this work, we define engagement as the total number of days for which a participant self-reported their wellbeing during the study period. Multiple entries made by the same participant in a single day were therefore counted as one daily self-report. Each participant in the study was expected to provide 28 daily self-reports resulting in a total of 308 reports per user group ($n = 11$). We also measured the trend of engagement for each user group in order to compare users' continued willingness to use the system.

Data Analysis

Data collected from each of the systems were analyzed in R (Version 3.6.2). For each statistical analysis conducted, normality was assessed by inspecting the distribution of the residuals and visually comparing the outer quantiles with a standard normal distribution. For all the dependent variables, Welch two sample *t*-tests were conducted. We conducted a sentiment analysis of the open-ended responses using Sentimentr,¹⁸ which scores each response on the scale of -2 (negative) to 2 (positive). We carried out a linear regression to investigate the relationship between the number of daily self-reports and days within CA and WA users.

RESULTS

We report our results from a comparative analysis of the experiences and behaviors of 22 individuals ($F = 73\%$, 16/22) with a mean age of 27.8, living with ADs using either a CA or WA to self-report their mental health and wellbeing in the home setting.

We decided to remove outliers during analysis of the self-report duration; defined as any value more than 2.5 SD from the global mean (GM) as a significant self-report duration in the case of the CA could arise as a result of latency in the system. In the case of the WA, this could also indicate a participant becoming distracted while completing their self-report. Seven entries from Group CA ($M = 197.4$ s, $GM = 52.2$ s) and 2 entries from Group WA ($M = 45762.5$ s, $GM = 648.1$ s) were as such removed.

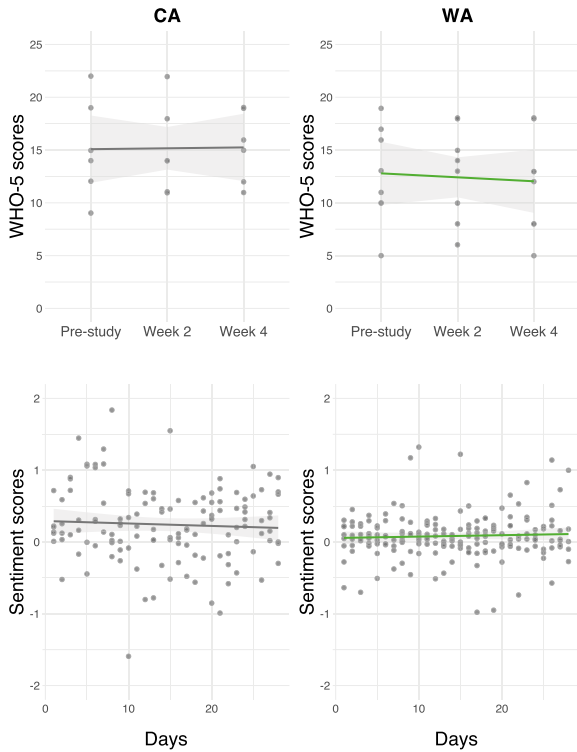


FIGURE 3. WHO-5 scores and the open-ended responses' sentiment scores.

As expected, WA participants spent a significantly higher period of time ($M = 264.6$ s, $SD = 755.5$) expressing their emotions every day than participants using the CA ($M = 47.5$ s, $SD = 26$) ($t(357) = -7.1$, $p < 0.001$).

Similarly, the number of words provided during daily self-reports by WA users ($M = 178$, $SD = 340$) was significantly higher than provided by CA users ($M = 44$, $SD = 31$) ($t(228) = -5.9$, $p < 0.001$).

Regarding users' responses to the WHO-5 questionnaire, participants were required to answer the questionnaire on the 1st, 14th, and 28th day of the study. However, several participants in both study arms responded later (CA = 45%, 5/11 and WA = 27.27%, 3/11); as a consequence of our system design, this could cause the last questionnaire to not be presented to some of those users. Participants with missing WHO-5 scores were removed from the analysis.

As shown in Figure 3, CA participants' WHO-5 scores trended slightly higher over time whereas there was a small drop in the sentiment scores of their open-ended responses. In contrast, there was a slight decline in WA participants' WHO-5 scores and little variation in the sentiment scores of their open-ended responses.

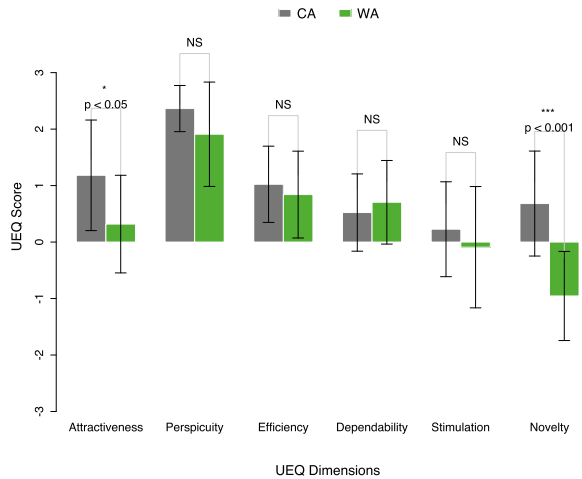


FIGURE 4. Participant's mean UEQ scores according to each dimension with error bars reflecting standard deviation. UEQ scores along the y -axis range from -3 (horribly bad) to $+3$ (extremely good). NS = Nonsignificant.

User Experience

Figure 4 shows participants' subjective assessments of their self-report experiences via CA and WA according to all six dimensions of the UEQ scale. This figure indicates an overall positive self-report experience via CA. In particular, participants rated the CA significantly higher than the WA in terms of novelty ($M = 0.7$, $SD = 0.9$, $t(19) = 4.45$, $p < 0.001$) and attractiveness ($M = 1.2$, $SD = 0.9$, $t(19) = 2.19$, $p < 0.05$).

However, there were no statistically significant differences between the two systems in terms of perspicuity, efficiency, dependability, and stimulation. These results suggest then that participants experienced the practice of self-report via CA as more novel and attractive than via WA.

User Engagement

The overall rates of user engagement via CA and WA were 72.7% and 73.3%, respectively. The difference in users' engagement via CA ($M = 8$, $SD = 1.3$) and WA ($M = 8.1$, $SD = 1.41$) was not statistically significant. Only WA participants made multiple entries for the same day, which summed up to 31 entries from six participants.

Furthermore, we did not find any significant interaction (β_1) between the number of self-reports and days for CA ($\beta_1 = -0.01$) and WA ($\beta_1 = -0.07$); $t(52) = -1.27$, $p = .21$, which indicates that participants engaged with the CA to a similar extent to those users

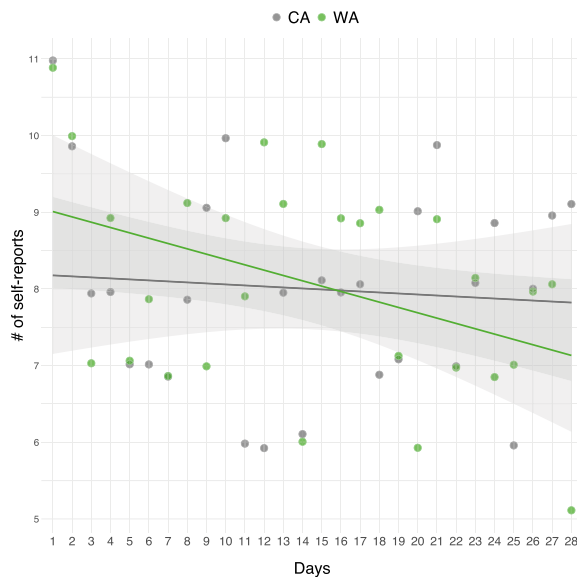


FIGURE 5. Engagement trend between CA and WA users.

of the WA, despite this comprising a new method of self-report entailing several limitations as discussed earlier. Figure 5 illustrates the trend of engagement with the system for each user group.

DISCUSSION

This work compared people living with ADs' experiences and behaviors of mental health self-report via smart speaker and WA in their home settings. Overall, participants who used smart speakers to self-report their mental health and wellbeing reported positive experiences and rated their experiences significantly higher in terms of novelty and attractiveness, compared to those participants using a WA. We also found that the difference between users' engagement with CA and WA was not statistically significant. These findings suggest that speech-enabled CAs can serve as a more novel and attractive means for the self-report of mental health and wellbeing with similar rates of engagement to a traditional web-based method.

CAs as a Novel and Attractive Means of Self-Report

Participants in this study rated self-reporting via CA as a novel and attractive experience compared to a web-based method. While this finding may seem obvious as speech is a novel mode of interaction and consists of voice features that can prove attractive (e.g., voice tone, rate of speech), one prior study has found rather contrasting results for users' self-report experiences using a CA; Quiroz *et al.* compared users' experiences

of a custom-designed Alexa Skill implementing a single open-ended question allowing users to express their emotions in a single word and asked them to take a self-test for depression or anxiety by responding to standard health questionnaires (PHQ-9, GAD-7).⁸ The authors reported that participants rarely used the system and considered the CA used during the study to lack efficiency and novelty.

One particular difference in the design of the CA implemented in the case of this study was that it allowed users to express their emotions by responding to a series of three open-ended questions for the purpose of engaging participants in more conversational interaction (see Figure 1). Perhaps then the conversational dialog design of *Sofia* might have been the key reason for users to rate this system significantly more highly; in line with prior research which indicates that users tend to hold human-to-human-like conversational expectations of CAs.¹² Therefore, although the current state of the technology may not be sophisticated enough to engage users in truly conversational interaction, our results suggest the value of this novel mode of interaction as means to support positive experiences of the self-report of mental health and wellbeing.

CAs as an Engaging Means of Self-Report

Our findings furthermore demonstrate users' consistent engagement with the CA for daily self-reports despite many limitations. The smart speakers used during this study were, in particular, required to be plugged in; therefore, rendered stationary and limiting users' access to the system. In contrast, the WA was a more accessible medium as it could also be used on mobile devices. Yet, the difference between users' engagement with these two different media was not statistically significant.

In addition, while the inclusion of open-ended questions in GUI-based systems has been reported to place an additional burden on users, participants in this study primarily reported their emotional wellbeing by responding to three open-ended questions. Users' continued engagement with the CA during this study, therefore, also suggests CA's potential to overcome the reporting burden otherwise placed on users by traditional self-report systems. These results indicate CAs' potential to serve as an engaging means of mental health self-report, echoing results from prior research.^{6,7} However, users' higher rating of their self-reporting experiences via CA in terms of novelty and attractiveness also reflects the potential challenge of

sustaining the CA user engagement. Novelty and attractiveness as motivators of users' engagement¹⁹ may soon lose their impact if the CA is not also able to continuously meet users' expectations.^{12,20}

Limitations

The generalizability of these results is subject to a number of limitations. First, this study was conducted with a small sample size ($N = 22$) over a short period of time (4 weeks). Second, participants in this study used Google Nest Devices alone to self-report their emotional wellbeing via CA, and the results might, therefore, not translate to CAs deployed via other devices. Third, neither of the self-report systems (CA or WA) used in this study included a notification mechanism; and hence results might prove different with the introduction of such a feature. And finally, this study was conducted during the COVID-19 outbreak. Due to social distancing guidelines and travel restrictions, participants might have stayed home more often than normal and, hence, engaged with these systems more than they would have in a more typical context.

CONCLUSION

In this article, we have presented the results of a comparative analysis of the self-report experiences and behaviors of people living with ADs, using either a CA or WA to self-report their mental health and wellbeing in their home settings over a period of four weeks. Participants using *Sofia* rated their self-report experience as significantly novel and attractive than those who self-reported via a more traditional web form. Although there was no significant difference in the users' engagement with either platform, participants' engagement with *Sofia* remained more consistent throughout the study. These findings then suggest that CAs may serve as an engaging means to support the self-report of mental health and wellbeing.

ACKNOWLEDGMENTS

We thank our participants for their participation in this study. This project was supported by the Novo Nordisk Foundation, under Grant NNF16OC0022038, and the Copenhagen Center for Health Technology (CACHET).

REFERENCES

1. K. Doherty, A. Balaskas, and G. Doherty, "The design of ecological momentary assessment technologies," *Interacting Comput.*, vol. 32, no. 3 pp. 257–278, 2020, doi: [10.1093/iwcomp/iwaa019](https://doi.org/10.1093/iwcomp/iwaa019).
2. Y. Chen, *The Effects of Question Customization on the Quality of an Open-Ended Question*, Data, Research, and Evaluation, Nebraska Dept. Education, Lincoln, NE, USA, 2017.
3. E. Research, "The smart audio report winter 2019 from NPR and edison research," Jan. 30, 2020. [Online]. Available: <https://www.edisonresearch.com/the-smart-audio-report-winter-2019-from-npr-and-edison-research>
4. J. Y. Shin and J. Huh-Yoo, "Designing everyday conversational agents for managing health and wellness: A study of alexa skills reviews," in *Proc. 14th EAI Int. Conf. Pervasive Comput. Technol. Healthcare*, 2020, pp. 50–61, doi: [10.1145/3421937.3422024](https://doi.org/10.1145/3421937.3422024).
5. B. Kinsella and A. Mutchler, "Voice assistant consumer adoption in healthcare," Oct. 2019. [Online]. Available: <https://voicebot.ai/voice-assistant-consumer-adoption-report-for-healthcare-2019>
6. D. DeVault et al., "SimSensei kiosk: A virtual human interviewer for healthcare decision support," in *Proc. Int. Conf. Auton. Agents Multi-Agent Syst.*, 2014, pp. 1061–1068. [Online]. Available: <https://dl.acm.org/doi/10.5555/2615731.2617415>
7. R. Kocielnik, D. Avrahami, J. Marlow, D. Lu, and G. Hsieh, "Designing for workplace reflection: A chat and voice-based conversational agent," in *Proc. Des. Interactive Syst. Conf.*, 2018, pp. 881–894, doi: [10.1145/3196709.3196784](https://doi.org/10.1145/3196709.3196784).
8. J. C. Quiroz, T. Bongolan, and K. Ijaz, "Alexa depression and anxiety self-tests: A preliminary analysis of user experience and trust," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, 2020, pp. 494–496, doi: [10.1145/3410530.3414374](https://doi.org/10.1145/3410530.3414374).
9. B. Suhm, "Towards best practices for speech user interface design," in *Proc. 8th Eur. Conf. Speech Commun. Technol.*, 2003, 1–4. [Online]. Available: https://www.isca-speech.org/archive_v0/archive_papers/eurospeech_2003/e03_2217.pdf
10. A. S. Miner et al., "Assessing the accuracy of automatic speech recognition for psychotherapy," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–8, 2020, doi: [10.1038/s41746-020-0285-8](https://doi.org/10.1038/s41746-020-0285-8).
11. A. Pradhan, K. Mehta, and L. Findlater, "Accessibility Came by Accident" Use of voice-controlled intelligent personal assistants by people with disabilities," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2018, 1–13, doi: [10.1145/3173574.3174033](https://doi.org/10.1145/3173574.3174033).
12. E. Luger and A. Sellen, "Like having a really bad PA" the gulf between user expectation and experience of conversational agents," in *Proc. CHI Conf. Hum. factors Comput. Syst.*, 2016, pp. 5286–5297, doi: [10.1145/2858036.2858288](https://doi.org/10.1145/2858036.2858288).

13. C. Topp, S. D. Østergaard, S. Sùndergaard, and P. Bech, "The WHO-5 well-being index: A systematic review of the literature," *Psychother. Psychosomatics*, vol. 84, no. 3, pp. 167–176, 2015, doi: [10.1159/000376585](https://doi.org/10.1159/000376585).
14. Accessed: Nov. 24, 2021. [Online]. Available: <https://cloud.google.com/dialogflow/docs>
15. Accessed: Nov. 24, 2021. [Online]. Available: <https://en.nvk.dk/rules-and-guidelines/act-on-research-ethics-review-of-health-research-projects>
16. B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Proc. Symp. Austrian HCI Usability Eng. Group*, 2008, pp. 63–76, doi: [10.1007/978-3-540-89350-9_6](https://doi.org/10.1007/978-3-540-89350-9_6).
17. K. Doherty and G. Doherty, "Engagement in HCI: Conception, theory and measurement," *ACM Comput. Surv. (CSUR)*, vol. 51 no. 5, pp. 1–39, 2018, doi: [10.1145/3234149](https://doi.org/10.1145/3234149).
18. T. Rinker, "Package 'sentimentr'," 2017. Accessed: Nov. 24, 2021. [Online]. Available: <https://cran.r-project.org/web/packages/sentimentr/readme/README.html>
19. H. L. O'Brien and E. G. Toms, "What is user engagement? A conceptual framework for defining user engagement with technology," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59 no. 6, pp. 938–955, 2008, doi: [10.1002/asi.20801](https://doi.org/10.1002/asi.20801).
20. M. Cho, S. S. Lee, and K. P. Lee, "Once a kind friend is now a thing: Understanding how conversational agents at home are forgotten," in *Proc. Designing Interactive Syst. Conf.*, 2019, pp. 1557–1569, doi: [10.1145/3322276.3322332](https://doi.org/10.1145/3322276.3322332).

RAJU MAHARJAN is currently a postdoctoral researcher at the Technical University of Denmark, DTU Management, Kongens Lyngby, Denmark. His research interests include the design of conversational agents to support the self-report of mental health and wellbeing. He is the corresponding author of this article. Contact him at rajm@dtu.dk.

DARIUS A. ROHANI is currently a postdoc at the Psychiatric Center Copenhagen, Rigshospitalet, and the Technical University of Denmark, Department of Health Technologies. In 2021, he received the Lundbeck Postdoc research grant to continue his work in mobile recommender systems to support patients with depressive disorders. Contact him at daroh@dtu.dk.

KEVIN DOHERTY is currently a postdoctoral human–computer interaction researcher with the Technical University of Denmark (DTU HealthTech), Kongens Lyngby, Denmark. His research interests include interdisciplinary design research, mental health and psychotherapy, user engagement and self-report methods, care ethics, and mobile interaction design. Contact him at kevdo@dtu.dk.

PER BÆKGAARD is currently an associate professor with the Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU Compute), Kongens Lyngby, Denmark. He heads the Human-Centered AI M.Sc. education with DTU. His research interests include human–computer interaction, user experience, cognition, and applied machine learning. Contact him at pgba@dtu.dk.

JAKOB E. BARDAM is currently a professor in digital health with the Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark. He is the maintainer of the open-source CARP Mobile Sensing Framework (<https://carp.cachet.dk/cams/>). His research interests include software architecture for mobile and ubiquitous computing, mobile sensing, and human–computer interaction. Contact him at jakba@dtu.dk.