

Association for Information Systems

## AIS Electronic Library (AISeL)

---

ICIS 2022 Proceedings

IS in Healthcare

---

Dec 12th, 12:00 AM

### Are Embodied Conversational Agents effective Tools for collecting Patient-reported Outcome Measures? – Towards a novel Approach in Multiple Sclerosis Care

Thure Weimann

*Technische Universität Dresden, thure.weimann@tu-dresden.de*

Hannes Schlieter

*Technische Universität Dresden, Hannes.Schlieter@tu-dresden.de*

Felix Reinsch

*Technische Universität Dresden, felix.reinsch@mailbox.tu-dresden.de*

Tjalf Ziemssen

*Carl Gustav Carus University Hospital, Technische Universität Dresden, tjalf.ziemssen@ukdd.de*

Follow this and additional works at: <https://aisel.aisnet.org/icis2022>

---

#### Recommended Citation

Weimann, Thure; Schlieter, Hannes; Reinsch, Felix; and Ziemssen, Tjalf, "Are Embodied Conversational Agents effective Tools for collecting Patient-reported Outcome Measures? – Towards a novel Approach in Multiple Sclerosis Care" (2022). *ICIS 2022 Proceedings*. 11.

[https://aisel.aisnet.org/icis2022/is\\_health/is\\_health/11](https://aisel.aisnet.org/icis2022/is_health/is_health/11)

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Are Embodied Conversational Agents effective Tools for collecting Patient-reported Outcome Measures? – Towards a novel Approach in Multiple Sclerosis Care

*Short Paper*

**Thure Weimann**

Technische Universität Dresden  
Dresden, Germany  
thure.weimann@tu-dresden.de

**Hannes Schlieter**

Technische Universität Dresden  
Dresden, Germany  
hannes.schlieter@tu-dresden.de

**Felix Reinsch**

Technische Universität Dresden  
Dresden, Germany  
felix.reinsch@mailbox.tu-dresden.de

**Tjalf Ziemssen**

Carl Gustav Carus University Hospital  
Dresden, Germany  
tjalf.ziemssen@uniklinikum-dresden.de

## Abstract

*Treating chronic diseases often involves repeated assessments from the patient's perspective to guide therapy decisions and promote quality of care. Therefore, patient-reported outcome measures (PROMs) have been established in the form of questionnaires. One promising approach for collecting PROMs are embodied conversational agents (ECAs), which have the potential to make the questionnaire completion more engaging, interactive and lower the response burden for the patient. Building on Satisficing Theory, this research-in-progress paper reports on the design and preliminary evaluation of an ECA for multiple sclerosis patients. The results indicate that such a system meets the needs of the patients and motivates a comparative study to contribute further evidence on the use and advantage of ECAs for this purpose. Based on a literature review, an evaluation approach including a research model is derived, and implications for future research are discussed.*

**Keywords:** Conversational agents, Patient-reported outcomes, Multiple sclerosis

## Introduction

In recent years, data reported by the patients themselves has become a fundamental pillar in the medical data collection (Hjollund 2019). To measure the patient's subjectively perceived health status, so-called patient-reported outcome measures, short PROMs, have been established. These are standardized and reliable questionnaires (Dawson et al. 2010). While paper-based questionnaires have been the traditional format for collecting PROMs, a change to web-based forms can be observed over the last decades. Despite several improvements (e.g., costs and data quality), common web-based forms still lack interactivity, do not provide audiovisual features to facilitate comprehension, and cannot clarify comprehension problems (Eriksen et al. 2022; te Pas et al. 2020). Modes of elicitation that might mitigate these problems are face-to-face or telephone interviews. However, these methods contradict the goal of decreased labor and costs, which were arguably one reason for switching to web-based forms. Despite the change to self-administered

digital solutions, reducing the patient's burden remains a significant challenge to achieve high response quality as well as response rates and is especially important for patients with cognitive impairments, low literacy, and a lack of computer skills (Bickmore et al. 2020; Zaratini et al. 2022).

One promising approach to address these challenges is the use of conversational agents. In general, the term “conversational agent” (CA) can be considered an umbrella term for intelligent software systems that attempt to emulate interpersonal communication (Seeger et al. 2021). By making the questionnaire completion more natural, personalized, engaging, and interactive, CAs may lower the burden on the patient and increase data quality and response rates (te Pas et al. 2020). Despite the popularity of speech-based CAs such as “Siri” or “Alexa” and a vastly growing body of research, literature on using them as dedicated survey tools is scarce (Celino and Re Calegari 2020). Concerning healthcare, CAs have been frequently studied in the role of digital therapeutics for supporting patients in their daily life or accelerating the diagnosis of diseases by supporting clinicians as end-users (Milne-Ives et al. 2020). Considering these advances, it seems surprising that research has only recently begun investigating patient-facing CAs designed for collecting PROMs in the context of chronic diseases (te Pas et al. 2020). While speech-based CAs and chatbots (i.e., text-based CAs) are considered disembodied, so-called embodied conversational agents (ECAs) are equipped with an animated and often humanoid visual representation (also called avatar) (Araujo 2018; Seeger et al. 2021). In general, the higher media richness of ECAs allows for more verbal and non-verbal social cues to be conveyed, making them potentially more effective in establishing a trustworthy relationship with the user (Qiu and Benbasat 2009; Schuetzler et al. 2018). However, there is still a lack of evidence if ECAs should be generally preferred when focusing on the disclosure and collection of sensitive data (Lind et al. 2013; Lucas et al. 2017). Therefore, we aim to investigate an ECA for collecting PROMs in the domain of multiple sclerosis, a chronic and neurodegenerative disease (Reich et al. 2018). Due to the variety of symptoms and impairments, multiple sclerosis is also referred to as “the disease of 1000 faces” which makes this context a well-suited ground for studying user interfaces designed for patients with different needs (Voigt et al. 2020). Common disease-related impairments such as fatigue, vision disturbances, or a decrease in cognitive processing speed highlight the demand for an accessible software design to lower the response burden for the patient and achieve sustained user engagement. This is of paramount importance as chronic diseases require PROMs to be collected more or less continuously throughout the lifespan of the patient. The need for novel approaches is further emphasized by the global PROM for multiple sclerosis initiative that sets digital innovations at the center of their strategic agenda (Zaratini et al. 2022). In particular, the following research question should be investigated:

**RQ:** *How effective is an ECA for collecting PROMs compared to survey methods with lower social presence?*

To address this question, the present research-in-progress paper aims to describe the state of the art regarding ECAs as survey tools in healthcare, report on the design process of an ECA for multiple sclerosis patients, and propose an evaluation approach, including a research model for future studies. The remainder of this paper is structured as follows: In the next section, we describe the theoretical background of our study and related work. Subsequently, we present the design and preliminary evaluation results of the ECA. Afterward, we derive our hypotheses, propose our research design and conclude with an outlook on future research.

## **Theoretical Background and Related Work**

### ***Embodied Conversational Agents as Survey Tools***

Recently, several studies were published that could demonstrate the effectiveness of chatbots as “virtual interviewers” compared to traditional survey methods (e.g., Celino and Re Calegari 2020; Kim et al. 2019; te Pas et al. 2020). Their use resulted in higher usability or user experience scores and thus may positively affect response quality and response rates. While these studies investigated chatbots (i.e., disembodied CAs), it seems particularly interesting what the state of the art in healthcare is regarding using ECAs. Therefore, we conducted a systematic literature review using Scopus and PubMed databases. The following search string was used in the fields title, abstract and keywords: (*"conversational agent\*" OR "conversational interface\*" OR "chatbot\*" OR "virtual agent\*" OR "virtual interviewer\*" OR "virtual human interviewer\*" OR "virtual assistant\*"*) AND (*"questionnaire\*" OR "survey\*" OR "form\*" OR "examination\*" OR "assessment\*" OR "interview\*" OR "screening\*"*) AND (*"compar\*" OR "versus" OR*

"vs." OR "equivalence"). The search was initially conducted in March 2022 and updated in August with no limitations regarding the time range. We included only peer-reviewed journal articles and conference papers written in English. Studies that did not investigate patient-facing ECAs, did not compare ECAs to a less media-rich self-administered survey method, and were not conducted in a health-related context were excluded. The context was considered health-related if either the questions asked by the ECA were similar to those asked in clinical interviews (e.g., sensitive questions such as symptoms or substance abuse) and/or the study was conducted in a therapy setting. All in all, n = 658 results were obtained (457 after duplicate removal) and initially screened for relevance based on the title, abstract, and exclusion criteria mentioned above. After the full-text screening, n = 6 articles were considered eligible, searched for forward- and backward citations (no additional papers found), and finally included in the review (see Table 1).

Reference	Study design	Key results
Pickard and Roster (2020)	Face-to-face vs. ACASI vs. ECA	Sensitive information disclosure was higher in the faceless ACASI condition
Bickmore et al. (2020)	ECA vs. Face-to-face (Study 1), ECA vs. Web-based form (Study 2)	Disclosure was higher to the ECA (Study 1) and participants were more satisfied (Study 2)
Lucas et al. (2017)	Web-based form vs. anonymized Web-based form vs. ECA	Respondents disclosed more symptoms to an ECA compared to the other conditions
Schuetzler et al. (2018)	Face-to-face vs. Web-based form vs. 2 (ECA vs. Chatbot) x 2 (low vs. high conversational relevance)	ECAs might be not better for collecting sensitive information compared to disembodied CAs
Micoulaud-Franchi et al. (2016)	Web-based form vs. ECA	Satisfaction and usability of the ECA was higher compared to a web-based form on a tablet
Lind et al. (2013)	Face-to-face vs. ACASI vs. ECA with more & less facial movement	Participants in audio-only mode (ACASI) had the highest disclosure
<b>Table 1. Overview of studies that investigated Embodied Conversational Agents (ECAs) as survey tools in a health-related context</b>		

Four of the included papers were published in the last five years, indicating the increasing interest in ECAs and their use for survey purposes in health-related contexts. Most papers studied the effectiveness of ECAs compared to traditional web-based forms. However, some papers also compared ECAs to face-to-face interviews or audio computer-assisted self-interviewing systems (ACASI). Primary outcomes of interest were usability or user experience metrics (e.g., ease of use, satisfaction) and self-disclosure of information operationalized as social desirability bias or the number of symptoms reported. It should be noted that only the studies by Bickmore et al. 2020, Lucas et al. 2017, and Micoulaud-Franchi et al. 2016) investigated an ECA in a therapy setting. In contrast, the other studies were conducted with healthy participants. Regarding the observation period, all studies were cross-sectional (i.e., conducted at a single point in time). Overall, the study results are mixed, with some demonstrating benefits of ECAs while others do not, and are only partially in line with those of prior studies investigating chatbots as survey tools. Concerning honest self-disclosure in surveys, the studies by Pickard and Roster (2020), Schuetzler et al. (2018), and Lind et al. (2013) suggest a negative impact of ECAs. They found that ECAs might not be preferred for collecting sensitive data (e.g., data on sexual activity or drinking behavior) and concluded that systems with less social presence might be more suitable for this task. In general, social presence can be considered as “the sense of human contact embodied in a medium” and is, therefore, higher when using ECAs compared to chatbots or traditional web-forms (Gefen and Straub 1997). In contrast to these findings, the studies by Lucas et al. (2017) and Bickmore et al. (2020) revealed that disclosure was higher when using a virtual human interviewer (ECA). Consequently, no consensus could be found in the literature on whether and under what conditions systems with high social presence (i.e., ECAs) are suitable when sensitive questions are asked. Furthermore, none of the studies investigated CAs as survey tools in the context of neurodegenerative diseases and evaluated the influence of the patient’s impairments on the interaction outcomes. This is particularly important as there might be interaction effects with the presence of disabilities of the respondent (e.g., slowdown of cognitive processing) that could affect the choice of interface (Roberts et al.

2019). Finally, current research lacks long-term studies that investigate the advantageousness of ECAs as survey tools in a chronic disease care scenario.

### ***Survey Satisficing Theory***

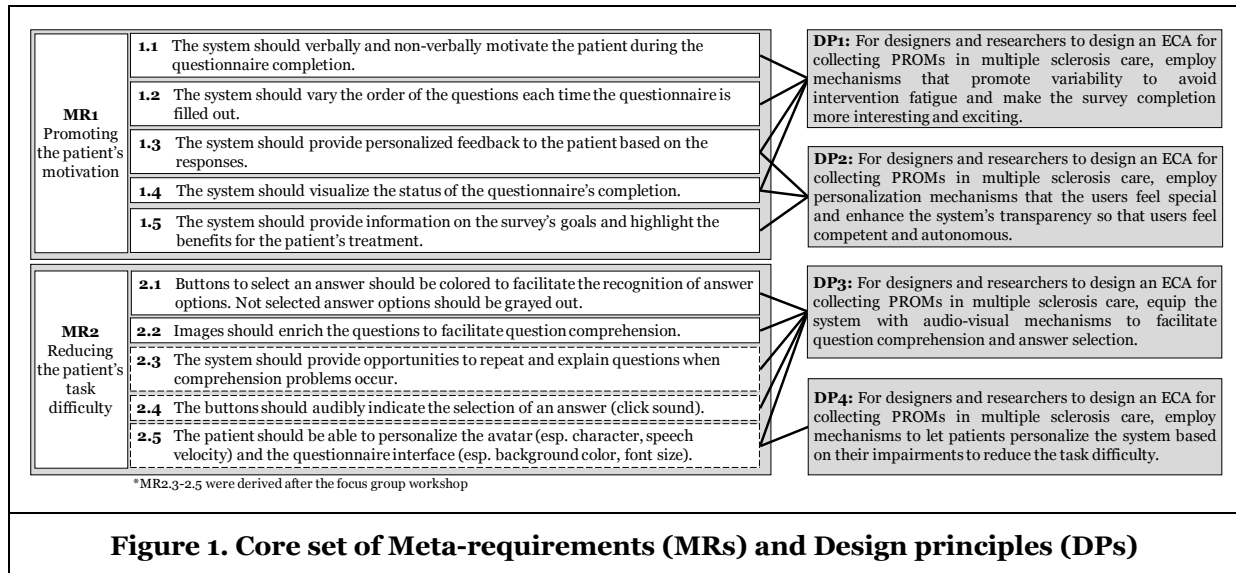
To explain the relationships between psychological and behavioral aspects when completing surveys, the theoretical framework of our study is informed by Krosnick's Satisficing Theory (1991). Therefore, instead of exploring response quality against the background of socially desirable responses as done in prior studies on ECAs as survey tools (e.g., Pickard and Roster 2020; Schuetzler et al. 2018), we focus on the so-called satisficing behavior as a cause of compromised data quality. Satisficing, in general, is a decision-making strategy to choose a sufficient and satisfying solution but not the optimal one. The term was introduced by Simon (1956) and unified the words "satisfy" and "suffice". Transferred to the context of survey methodology, satisficing explains the behavior of individuals to reduce cognitive efforts by providing sub-optimal answers. Krosnick (1991) considers satisficing as a continuum ranging from weaker forms such as simply selecting the first reasonable answer option to more potent forms such as non-differentiating (i.e., always giving the same answer), providing "not applicable" as an answer although it's not the truth or randomly choosing an answer ("mental coin flipping"). Beyond the strongest form of satisficing, nonresponding arguably lowers the respondent's cognitive burden to a minimum. The results of those strategies to lower the cognitive efforts are missing values, biased data, and thus low response quality (Roberts et al. 2019). According to Krosnick's theory, three factors promote satisficing behavior: task difficulty, the respondent's motivation, and the respondent's abilities. The latter comprises the cognitive skills, experience with the survey questions or topics, and existing attitudes (Krosnick 1991). Given that cognitive impairment is prevalent among multiple sclerosis patients (Reich et al. 2018), this patient population may be particularly prone to satisfice when completing surveys. Consequently, satisficing theory suggests tailoring the design of survey instruments to the abilities of individuals by providing mechanisms to reduce the task difficulty and increase the respondent's motivation for avoiding satisficing behavior (Schaeffer 2021).

### **Design and Preliminary Evaluation of the ECA**

To design the digital health intervention, we followed a design science research methodology and first derived a core set of seven meta-requirements (MRs). As suggested by Bradbury et al. (2014), we included different views (health professionals and patients) that are both crucial for the successful implementation and adoption of the intervention in clinical practice. Therefore, we first conducted several group interviews with three health professionals involved in multiple sclerosis care (one physician and two research associates). In a brainstorming session, every expert was asked how the system should be designed based on the individual experience and knowledge of the needs of multiple sclerosis patients. Afterward, the individual ideas were discussed and summarized in a consensus-building process among the experts. The resulting core set of MRs was then clustered to requirements that address the patient's motivation and requirements regarding reducing the task difficulty. Based on this core set derived from expert interviews, we designed a first prototype, which was then used to validate and extend the MRs in a focus group workshop with patients. Figure 1 depicts an overview of the elicited MRs.

For promoting the patient's motivation, the system should encourage questionnaire completion (e.g., "You're already halfway done. That's great!") (MR1.1), the order of the question should be varied when a patient fills out a questionnaire multiple times (MR1.2), the system should provide personalized feedback based on the responses (MR1.3) and visualize the status of completion (MR1.4). For the provision of personalized feedback, we distinguish two approaches. First, based on deviations detected between responses at earlier points in time (if the questionnaire is filled out multiple times), specific questions could be asked back (e.g., "Four weeks ago, you stated no mood disabilities. What happened?"). Second, after completing the questionnaire, the patient could receive feedback on what the responses mean regarding the course of the disease. Furthermore, the system should provide information on the goals of the PROM assessment and highlight the benefits for the patient's treatment to enable more competent decisions (MR1.5). To reduce the task difficulty, buttons should be colored (MR2.1), and questions enriched by images (MR2.2). Based on this initial set of MRs, we derived three design principles (DP1-3) based on the recommended structure by Gregor et al. (2020) and instantiated them in a prototype. The entire questionnaire application was implemented using the gaming engine "Unity" and extends a web portal for multiple sclerosis patients described by Voigt et al. (2020). For the avatar, we selected a young female

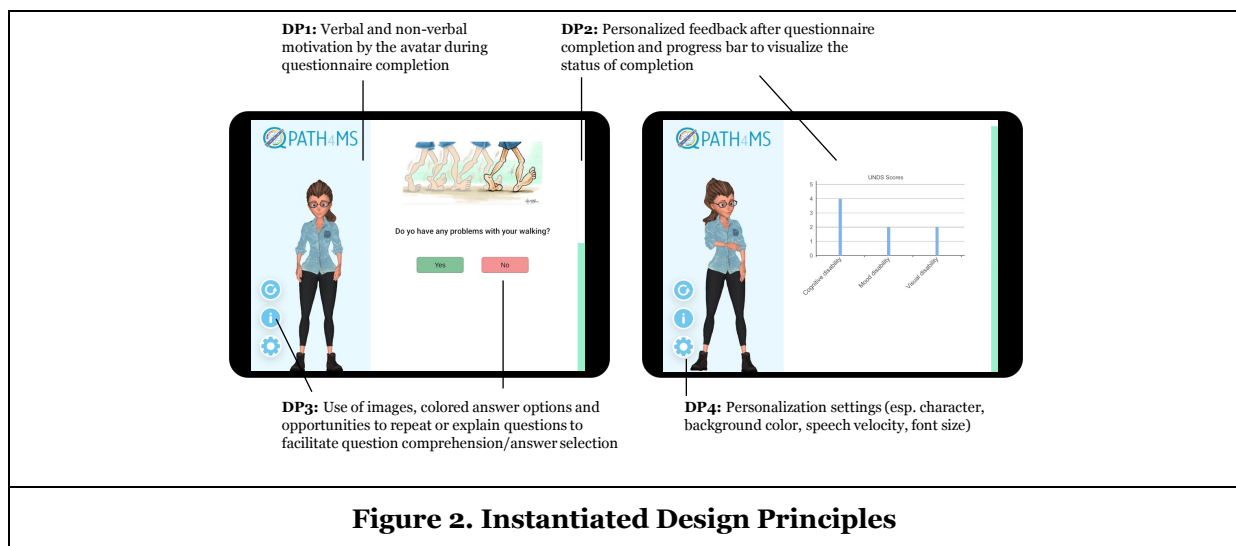
character that was preferred by the users in a prior study of our research group where different avatar designs were investigated in an obesity context (Weimann et al. 2022).



Concerning the avatar's voice, we used the Microsoft Azure Text-to-Speech service (i.e., speech synthesis). The prototype was then evaluated in a focus group workshop with  $n = 6$  patients ( $n = 2$  females) in different stages of the disease and with different disabilities caused by the disease. At the beginning of the workshop, the general idea of a novel approach for collecting PROMs was introduced to the patients. In the next step, each patient interacted with the application deployed on a 10-inch tablet and filled out a subset of an adapted German version of "Guy's Neurological Disability Scale" (GNDS) (Sharrack and Hughes 1999). The used scale contains 78 questions and assesses 13 disabilities associated with multiple sclerosis (e.g., cognitive, visual, upper, and lower limb disabilities). Most of the GNDS questions are being answered by "yes" or "no", with additional questions being asked to each disability depending on the patient's answer (branching logic). For example, if the patient states to have no walking problems, no further questions are being asked related to lower limb disabilities. Before starting the questionnaire, the avatar introduced herself, "Hey, I'm Lea. May I ask you some questions?" and briefly explained the benefit for the patient in terms of more targeted medical consultation and treatment. After questionnaire completion, the patient obtained personalized feedback from the avatar on the calculated scores of the GNDS dimension. Both features were implemented to increase the transparency of the system's data collection and evaluation (see DP2). For evaluating the impressions of the system, we used the user experience questionnaire (UEQ) (Laugwitz et al. 2008) that was filled out afterward by the patient (paper-based). Additionally, we asked for qualitative feedback on the personal impression, areas of improvement, and if the patient could imagine using the system in the long term. After the participants had interacted with the system and completed the evaluation questionnaire, a 15-minute group discussion was held. Based on the results of the focus-group workshop, the MRs and DPs were refined, extended, and implemented (MR2.3-2.5). Figure 2 depicts how the derived DPs were instantiated in the prototype.

Generally, the patients liked the ECA-guided questionnaire completion, and all stated that they could imagine using such a system in the long term. The average UEQ ratings exceeded the mark of +0.8 in all six dimensions (from -3 to +3) indicating a positive evaluation (attractiveness: +1.8, perspicuity: +2.4, efficiency: +1.6, dependability: +1.8, stimulation: +1.5 and novelty: +1.4). The standard deviation was below 1 in all dimensions. Notably, the average perspicuity rating was the highest, suggesting that the system was easy to use and get familiar with. Although this approach for collecting PROMs was novel for this patient group, the average novelty rating was the lowest compared to the other dimensions. Previous experiences with avatars (e.g., in video games) might explain this finding. The qualitative feedback further highlights the positive impressions, e.g., "Great concept, promises a lot", "The avatar has a pleasant voice but could speak a little bit slower" or "good audio qualities, good that questions are read aloud". Despite the synthetic nature of the avatar's voice, the voice was still evaluated positively and pleasant. Further, the patients liked that the questions were read aloud and found this facilitation when answering the

questionnaire. This was further reflected in the group discussion, where the participants stated that they could also imagine using an audio-only system without an avatar (i.e., ACASI). The patients perceived the personalized feedback on the different GNDS dimensions helpful and found this system's transparency important. However, the group discussion also revealed that the patients wished to have dedicated personalization options where the character of the avatar (esp. gender and style), the speech velocity, the font size, and the background color could be adjusted (MR2.5). Particularly for patients facing a slowdown of cognitive processing and visual disturbances, the opportunity to change the speech velocity and background color was perceived as an important feature. Additionally, some patients asked for buttons where the questions could be repeated or re-explained when comprehension problems occur (MR2.3) and buttons that audibly indicate (click sound) the selection (MR2.4). Half of the patients criticized the system forcing them to wait until the entire question was read aloud before allowing them to move on to the next question. Indeed, there are pros and cons to this approach. On the one hand, the system ensures that the patient at least had the opportunity to understand the whole question, and artificial breaks of the avatar's voice are avoided. On the other hand, the duration of the questionnaire completion is relatively strictly defined, which could decrease the patient's motivation. However, for further evaluation, we decided to continue using this approach.



## Evaluation Approach

To evaluate the ECA's effectiveness, we aim to compare the system in the next step to other survey modes within a larger patient population (planned sample size,  $n = 162$ ). Therefore, we will conduct a cross-sectional experiment first and aim to contribute to a longitudinal study in a subsequent step. The cross-sectional study will take place at the multiple sclerosis center of a university hospital as part of the regular patient visit (1600 multiple sclerosis patients under current control). In the longitudinal study, the patients will use the questionnaire application in the home environment. Overall, we will investigate three different survey modes with varying degrees of social presence manipulated by their media-richness: A) ECA-guided questionnaire (high social presence), B) audio computer-assisted self-interviewing (ACASI), and C) a web-based form (low social presence). The ACASI condition corresponds to the system in Figure 2 without the avatar, whereas the third condition (WF) has neither an avatar nor audio enrichment but still colored buttons and images. In the cross-sectional study, patients will be randomly assigned to one of six groups, allowing for all possible combinations of orders. We will use the GNDS as a questionnaire implemented in the three different systems. At the beginning, data on the socio-demographic background, time living with the disease, and digital literacy are captured for subgroup analysis. Only after the first condition the measures described in the following paragraphs, with the exception of the subjectively preferred mode, are collected using a traditional web-based form (without any additional features). The second and third conditions will present a reduced subset of the GNDS to avoid patients becoming bored and exhausted over the experiment but still enable a subjective judgment of which mode was most preferred at the end. For the longitudinal study, we randomly assign the patients to one of the three conditions (ECA, ACASI, or WF).

Krosnick’s Satisficing Theory (1991) defines the primary constructs of interest in our study. In accordance with related studies (e.g., Celino and Re Calegari 2020; te Pas et al. 2020), user experience is considered as a central concept of our evaluation approach. Similarly to the preliminary evaluation, we will therefore use the UEQ. The scale structure of the UEQ is classified into pragmatic quality (efficiency, perspicuity, dependability), hedonic quality (stimulation, novelty), and a valence dimension (attractiveness) combining both dimensions and serves as an overall judgment (Laugwitz et al. 2008). Considering that the pragmatic quality refers to “do-goals” (i.e., system support to accomplish a certain task) and the hedonic quality to “be-goals” (support of intrinsic motivation, e.g., being special, competent, and autonomous), the UEQ constructs reflect the task difficulty and the respondent’s motivation (Hassenzahl 2008). In addition, we consider the time to complete the questionnaire as a measure of the respondent’s motivation (measured by the system). However, the relationship between time to motivation is ambiguous. On the one hand, the respondent’s motivation tends to decrease as longer the questionnaire completion takes (Krosnick 1991). On the other hand, a longer completion time may also indicate higher response quality when the respondent takes more time to think about the questions thoroughly and can be positively correlated with the respondent’s motivation (Gummer and Roßmann 2015). Arguably, the time to complete could serve in both directions as an indicator of the respondent’s motivation and should be interpreted along with further motivation measures. Concerning the task difficulty, we will also measure the perceived effort using the corresponding sub-scale of the Survey Participation Inventory (SPI) (Brüggen et al. 2011). The respondent’s abilities impaired by the disease will be directly measured via the GNDS. We will use the willingness to answer voluntary questions by disclosing therapy-relevant information to operationalize satisficing behavior. Therefore, we will also use the GNDS that includes a set of questions related to the sexual function but first asks the patient, “Do you mind if I ask you about this?” and gives “yes”, “no”, and “not applicable” as answer options. We consider the answers “no” and “not applicable” as a form of satisficing behavior for bypassing the additional questions.

Based on prior research revealing that increasing social presence positively affects constructs related to do- and be-goals (e.g., Gefen and Straub 1997; Qiu and Benbasat 2009), we hypothesize that the ECA yields to significantly higher levels of the respondent’s motivation (H1a), lower task difficulty (H2a) and thus less satisficing behavior (H3a) than the ACASI condition. Likewise, the ACASI yields to significantly higher levels of the respondent’s motivation (H1b), lower task difficulty (H2b), and less satisficing behavior (H3b) than the WF condition. Further, we hypothesize that the relationship between system and motivation (H4) and task difficulty (H5) is moderated by the respondent’s abilities. Figure 3 summarizes the research model of our study. For secondary outcome analysis (exploratory), we will also measure the attractiveness (UEQ scale), self-reported response quality (SPI scale), response rate (only in the longitudinal study) and directly ask for the preferred system type. Given that social presence may induce trust and thus affect the perceived usefulness and self-disclosure (Qiu and Benbasat 2009; Schuetzler et al. 2018), we will measure this construct using the scale by Schrepp and Thomaschewski (2019). Last but not least, social presence will be measured based on the items by Gefen and Straub (1997) to check if the manipulations were successful.

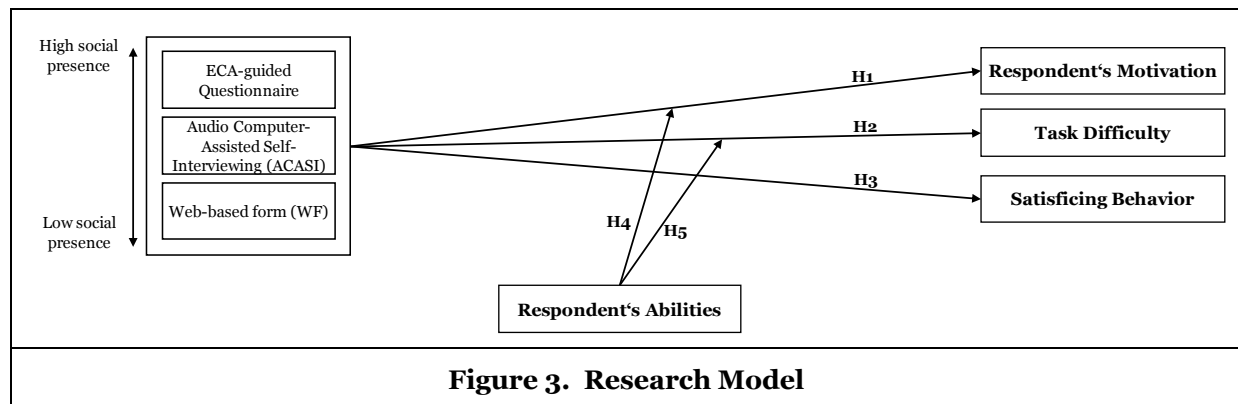


Figure 3. Research Model

## Conclusion

Our research might contribute further evidence on using ECAs in the medical domain and their potential suitability for collecting PROMs. In particular, novel approaches that can maintain sustained engagement



and high response quality over the long term could be another step towards the notion of digital biomarkers and further drive personalized medicine. Additionally, such approaches promise improved care processes and could thus contribute to decreased healthcare costs. The present study contributes with a validated core set of MRs and DPs derived from expert interviews and a patient workshop to enrich the knowledge base on how these systems could be designed. Given the diversity of neurological impairments faced by multiple sclerosis patients, the derived design knowledge and expected contributions may also be transferred to other disease contexts, such as Parkinson's disease or stroke rehabilitation. The preliminary evaluation in a small patient group could further demonstrate the usefulness of the ECA that needs to be proven in a comparative setting with a larger study population. Therefore, the present paper describes a two-step evaluation approach consisting of a cross-sectional and a subsequent longitudinal study. Since chronic diseases involve repeated PROM assessments, more studies are needed to investigate ECAs for this purpose in the long run.

## Acknowledgements

This work is part of the EFRE-funded research project QPath4MS. We thank all partners and supporters for enabling this project's contribution to science and healthcare practice.

## References

- Araujo, T. 2018. "Living up to the Chatbot Hype: The Influence of Anthropomorphic Design Cues and Communicative Agency Framing on Conversational Agent and Company Perceptions," *Computers in Human Behavior* (85), pp. 183–189.
- Bickmore, T., Rubin, A., and Simon, S. 2020. "Substance Use Screening Using Virtual Agents: Towards Automated Screening, Brief Intervention, and Referral to Treatment (SBIRT)," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, Virtual Event Scotland UK: ACM, October 20, pp. 1–7.
- Bradbury, K., Watts, S., Arden-Close, E., Yardley, L., and Lewith, G. 2014. "Developing Digital Interventions: A Methodological Guide," *Evidence-Based Complementary and Alternative Medicine* (2014), pp. 1–7.
- Brüggen, D. E., Wetzels, M., Ruyter, K. D., and Schillewaert, N. 2011. "Individual Differences in Motivation to Participate in Online Panels: The Effect on Reponse Rate and Reponse Quality Perceptions," *International Journal of Market Research* (53:3), pp. 369–390.
- Celino, I., and Re Calegari, G. 2020. "Submitting Surveys via a Conversational Interface: An Evaluation of User Acceptance and Approach Effectiveness," *International Journal of Human-Computer Studies* (139), p. 102410.
- Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., and Carr, A. J. 2010. "The Routine Use of Patient Reported Outcome Measures in Healthcare Settings," *BMJ* (340:jan18 1), pp. c186–c186.
- Eriksen, J., Bygholm, A., and Bertelsen, P. 2022. "The Association between Patient-Reported Outcomes (PROs) and Patient Participation in Chronic Care: A Scoping Review," *Patient Education and Counseling*, pp. 1852–1864.
- Gefen, D., and Straub, D. W. 1997. "Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model," *MIS Quarterly* (21:4), p. 389.
- Gregor, S., Chandra Kruse, L., and Seidel, S. 2020. "Research Perspectives: The Anatomy of a Design Principle," *Journal of the Association for Information Systems* (21:6), p. 2.
- Gummer, T., and Roßmann, J. 2015. "Explaining Interview Duration in Web Surveys: A Multilevel Approach," *Social Science Computer Review* (33:2), pp. 217–234.
- Hassenzahl, M. 2008. "User Experience (UX): Towards an Experiential Perspective on Product Quality," in *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine on - IHM '08*, Metz, France: ACM Press, p. 11.
- Hjollund, N. H. I. 2019. "Fifteen Years' Use of Patient-Reported Outcome Measures at the Group and Patient Levels: Trend Analysis," *Journal of Medical Internet Research* (21:9), p. e15856.
- Kim, S., Lee, J., and Gweon, G. 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*, Glasgow Scotland Uk: ACM, May 2, pp. 1–12.
- Krosnick, J. A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in

- Surveys,” *Applied Cognitive Psychology* (5:3), pp. 213–236.
- Laugwitz, B., Held, T., and Schrepp, M. 2008. “Construction and Evaluation of a User Experience Questionnaire,” in *Symposium of the Austrian HCI and Usability Engineering Group*, Springer, pp. 63–76.
- Lind, L. H., Schober, M. F., Conrad, F. G., and Reichert, H. 2013. “Why Do Survey Respondents Disclose More When Computers Ask the Questions?,” *Public Opinion Quarterly* (77:4), pp. 888–935.
- Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., and Morency, L.-P. 2017. “Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers,” *Frontiers in Robotics and AI* (4), p. 51.
- Micoulaud-Franchi, J.-A., Sagaspe, P., de Sevin, E., Bioulac, S., Sauteraud, A., and Philip, P. 2016. “Acceptability of Embodied Conversational Agent in a Health Care Context,” in *Intelligent Virtual Agents* (Vol. 10011), Lecture Notes in Computer Science, D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, and A. Leuski (eds.), Cham: Springer International Publishing, pp. 416–419.
- Milne-Ives, M., de Cock, C., Lim, E., Shehadeh, M. H., de Pennington, N., Mole, G., Normando, E., and Meinert, E. 2020. “The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review,” *Journal of Medical Internet Research* (22:10), p. e20346.
- te Pas, M. E., Rutten, W. G. M. M., Bouwman, R. A., and Buise, M. P. 2020. “User Experience of a Chatbot Questionnaire Versus a Regular Computer Questionnaire: Prospective Comparative Study,” *JMIR Medical Informatics* (8:12), p. e21982.
- Pickard, M. D., and Roster, C. A. 2020. “Using Computer Automated Systems to Conduct Personal Interviews: Does the Mere Presence of a Human Face Inhibit Disclosure?,” *Computers in Human Behavior* (105), p. 106197.
- Qiu, L., and Benbasat, I. 2009. “Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems,” *J. Manage. Inf. Syst.* (25:4), USA: M. E. Sharpe, Inc., pp. 145–182.
- Reich, D. S., Lucchinetti, C. F., and Calabresi, P. A. 2018. “Multiple Sclerosis,” *New England Journal of Medicine* (378:2), (D. L. Longo, ed.), pp. 169–180.
- Roberts, C., Gilbert, E., Allum, N., and Eisner, L. 2019. “Research Synthesis: Satisficing in Surveys: A Systematic Review of the Literature,” *Public Opinion Quarterly* (83:3), Oxford University Press US, pp. 598–626.
- Schaeffer, N. C. 2021. “Presidential Address. Far From Ordinary Questions: Task Difficulty, Motivation, and Measurement Practice,” *Public Opinion Quarterly* (84:3), pp. 798–812.
- Schrepp, M., and Thomaschewski, J. 2019. “Design and Validation of a Framework for the Creation of User Experience Questionnaires,” *International Journal of Interactive Multimedia and Artificial Intelligence* (5:7), p. 88.
- Schuetzler, R. M., Giboney, J. S., Grimes, G. M., and Nunamaker, J. F. 2018. “The Influence of Conversational Agent Embodiment and Conversational Relevance on Socially Desirable Responding,” *Decision Support Systems* (114), pp. 94–102.
- Seeger, A.-M., Pfeiffer, J., and Heinzl, A. 2021. “Texting with Humanlike Conversational Agents: Designing for Anthropomorphism,” *Journal of the Association for Information Systems* (22:4), p. 8.
- Sharrack, B., and Hughes, R. A. 1999. “The Guy’s Neurological Disability Scale (GNDS): A New Disability Measure for Multiple Sclerosis,” *Multiple Sclerosis Journal* (5:4), pp. 223–233.
- Simon, H. A. 1956. “Rational Choice and the Structure of the Environment,” *Psychological Review* (63:2), pp. 129–138.
- Voigt, I., Benedict, M., Susky, M., Scheplitz, T., Frankowitz, S., Kern, R., Müller, O., Schlieter, H., and Ziemssen, T. 2020. “A Digital Patient Portal for Patients With Multiple Sclerosis,” *Frontiers in Neurology* (11), p. 400.
- Weimann, T., Fischer, M., and Schlieter, H. 2022. *Peer Buddy or Expert? – On the Avatar Design of a Virtual Coach for Obesity Patients*, presented at the Hawaii International Conference on System Sciences.
- Zaratin, P., Vermersch, P., Amato, M. P., Bricchetto, G., Coetzee, T., Cutter, G., Edan, G., Giovannoni, G., Gray, E., Hartung, H. P., Hobart, J., Helme, A., Hyde, R., Khan, U., Leocani, L., Mantovani, L. G., McBurney, R., Montalban, X., Penner, I.-K., Uitdehaag, B. M. J., Valentine, P., Weiland, H., Bertorello, D., Battaglia, M. A., Baneke, P., and Comi, G. 2022. “The Agenda of the Global Patient Reported Outcomes for Multiple Sclerosis (PROMS) Initiative: Progresses and Open Questions,” *Multiple Sclerosis and Related Disorders* (61), p. 103757.