"Why is the Doctor a Man?" Reactions of Older Adults to a Virtual Training Doctor

Aurora Constantin
Catherine Lai
Elaine Farrow
Beatrice Alex[*]
aurora.constantin@ed.ac.uk
clai@inf.ed.ac.uk
elaine.farrow@ed.ac.uk
balex@ed.ac.uk
University of Edinburgh
Edinburgh, UK

Johan Jeuring
J.T.Jeuring@uu.nl
Utrecht University
Utrecht, The Netherlands

Ruth Pel-Littel
Henk Herman Nap
r.pel-littel@vilans.nl
H.Nap@vilans.nl
Vilans
Utrecht, The Netherlands

[*]Also with Alan Turing Institute.

Shared Decision Making (SDM)
Definition: Shared decision making in the context of health care services is the process of a practitioner and a patient jointly choosing an appropriate medical test or treatment as a way to enable patient-centred care.

ABSTRACT

Shared decision making (SDM) is increasingly considered as the best way to reach a treatment decision in a clinical environment. However, the use of SDM in practice can be obstructed by a number of factors, such as time constraints or lack of applicability due to patient characteristics. Our project,

PrepDoc, explores how a Virtual Training Doctor (VTD) can help patients overcome some of these obstacles to experiencing effective SDM during doctor's visits. In this paper, we report on user studies conducted with 19 participants in Scotland aged 65+. The goal of these studies was to identify the reactions of this audience to the PrepDoc system, evaluate its suitability within Scotland, and elicit suggestions to improve it. Our findings revealed that the idea of empowering people to participate in SDM using a virtual agent was positively received by all participants. However, the reactions to how this idea was implemented in the PrepDoc system varied greatly across participants. Based on this, our paper outlines recommendations for enhancing the user experience with VTDs, accommodating individual differences of older adults, and accounting for the national context.

KEYWORDS

User Experience; Evaluation Study; Technology for Older Adults; Shared Decision Making; Health.

INTRODUCTION

Shared decision making (SDM) facilitates the discussion between health care professionals and patients when decisions have to be made about desired care and treatment [3]. SDM has numerous benefits (e.g., patients have increased knowledge of the options, more accurate risk perception, greater comfort with decisions [12]), and several countries have developed programs to stimulate SDM in discussions between health care professionals and patients [3]. These programs target awareness of SDM among health care professionals and patients, and sometimes offer training to health care professionals. However, SDM is not yet common practice [4]. Despite the general agreement about its benefits, uptake of SDM faces a series of barriers, such as time constraints, lack of applicability due to patient characteristics, and lack of applicability due to the clinical situation [6]. Elwyn and collaborators [3] suggest a 3-step model to support SDM which, they emphasise, "has to be built on the core skills of good clinical communication skills". Practising SDM conversations can help overcome patient fears over being seen as difficult or feeling like an unequal partner in the conversation. They can also help patients improve their health literacy, e.g., understanding of medical terms, before appointments. Such preparation can be crucial given tight time constraints in GP visits.

A number of applications have been designed for training communication skills. For example, Cláudio et al. [2] present a game for communication training and assessment of self-medication consultation skills, allowing students in Pharmaceutical Sciences to communicate with Virtual Humans (VH) in an environment that simulates realistic self-medication scenarios. Other applications have targeted the sales sector [8], and how to handle difficult passengers on public transport [1]. Zhang and Bickmore [13] describe a virtual decision coach providing guidance around prenatal testing options. However, as far as we are aware there are no existing applications designed for practising SDM in clinical environments.

Example Scenario
In one scenario a user is told that they need to undergo hip surgery, but also that they want to go on a walking holiday next month. Their task is to ensure that in the conversation with the doctor, their personal situation is discussed.

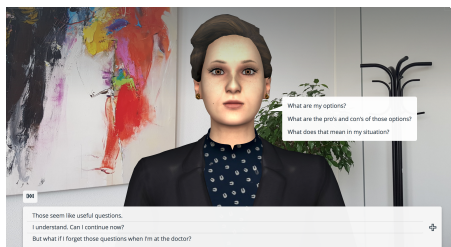
Figure 1: Daniel, the virtual training doctor

Figure 2: Sarah, the virtual assistant in the training scenarios

The goal of the PrepDoc[1] project is to explore how an online application can be designed to help patients aged 65+ to prepare for GP visits by practising SDM in conversations with a Virtual Training Doctor (VTD) using multi-modal interaction (mouse, text, audio, and speech).

Building on our previous work designing a serious game to support students in higher education in practising communication skills [7] and on the 3-step model to support SDM [3] developed by Elwyn and collaborators, we designed an online application that offers users several scenarios in which they can practise SDM in conversations with a VTD [9]. The system was initially developed in Dutch and then ported to English. This paper presents the results from an evaluation study of the English-language version, carried out in Scotland. From these studies we present recommendations for enhancing the user experience with VTD systems. Our work is expected to bring a number of contributions to the CHI community. First, we identify different factors that can impact the experience of older adults (age 65+) with VTD systems for SDM. Second we outline the implications of these for the future design of VTD systems for SDM.

THE PREPDOC SYSTEM

The PrepDoc system is an online application that offers a user several scenarios in which they can practise SDM in conversations with two virtual characters (VCs): a doctor and an assistant. The initial design of the system was informed by co-creation sessions organised in Utrecht with Dutch people aged 65+, including one GP and one GP assistant.

Each scenario is completely scripted (see example in sidebar). The doctor starts the conversation, and the user proceeds, either by selecting an answer from a list of options with the mouse, or by speaking or typing their answer (Figure 1). A few questions allow free responses, but most are limited to predefined options. After each conversation with the doctor, the user interacts with the assistant (Figure 2) who highlights the main points of the conversation with the doctor and helps the user reflect on what they have learned and prepare for their own GP visit.

EVALUATION STUDY

We recruited 19 participants aged 66 to 87, using a snowball methodology. 15 participants had a university degree and two had professional qualifications. Eight (6 males) of the participants had worked in computer science (CS) and were familiar with virtual characters and dialogue systems (**CS group**); the remaining 11 (2 males) came from a variety of educational and occupational backgrounds (**no-CS group**). We were interested in understanding whether prior familiarity with CS has an impact on the perceived user experience and/or on users' reactions to the system.

Each participant worked with the PrepDoc system on a laptop, individually and at their own pace, for 60-90 minutes. The Think Aloud protocol [5] was employed while the participants were using the system. Afterwards, they completed a short online questionnaire which included a System
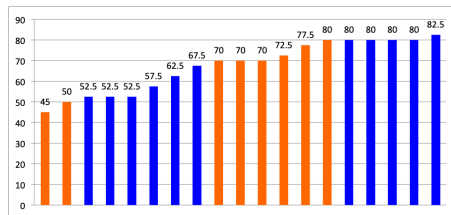
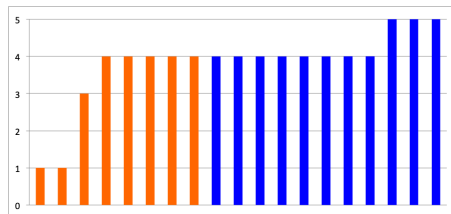**Figure 3: Individual System Usability Scale (SUS) scores from participants (CS group is shown in orange).**

Figure 4: Individual scores for perceived experience on a 5-point Likert scale, from 1 ("not good at all") to 5 ("very good"). The CS group is shown in orange.
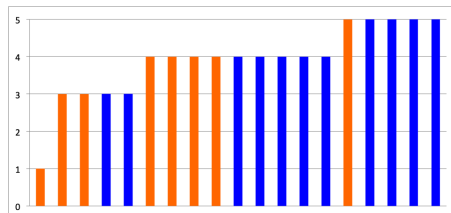

Figure 5: Individual scores for perceived usefulness on a 5-point Likert scale, from 1 ("not good at all") to 5 ("very good"). The CS group is shown in orange.

Usability Scale (SUS) questionnaire [11]. The aim of the questionnaire was to understand the overall perceived experience of using the system, and other related issues. Each participant also took part in a semi-structured interview at the end of their session. The interview questions were designed to shed more light on the user experience with the system and to collect suggestions for improvements. Two researchers from the team were present at each session to observe the participants and take notes. In addition, a camera was used to record the audio and the image of the laptop screen.

RESULTS & PRELIMINARY FINDINGS

Because the number of participants was small, we only used basic statistics for the ordinal data. To synthesise the notes and the video transcripts we employed open and axial coding [10]. The idea of empowering older adults in SDM during GP visits was unanimously praised. However, our findings revealed large differences in how participants perceived the implementation of this idea in PrepDoc.

Usability. The overall SUS score (Figure 3) for all participants ($M = 67.5$, $SD = 12.39$) was very close to 68, suggesting good usability [11]. However, 6 no-CS group participants and 2 CS group participants scored the system below 68, with one giving a score of 45, which is below the threshold to be considered "acceptable". The system scored slightly better on average ($M = 68$, $SD = 12.84$) with the no-CS group than with the the CS group ($M = 67.5$, $SD = 12.59$). That may be because participants from the CS group start with higher expectations of the system.

Perceived Experience. There was also a slight tendency in the no-CS group to score PrepDoc higher than people in the CS group in terms of perceived experience with the system (Figure 4). The median and mode in both groups was 4 ($SD = 1.36$ for the CS group, $SD = 0.47$ for the no-CS group). None of the participants in the CS group scored it "very good" and two of them scored it as "not good at all". Within the no-CS group the scores were more consistent, with three participants scoring their experience "very good". This may be again the consequence of participants in CS group having higher expectations of the system, but also the fact that all of them are highly educated and have broad general knowledge. For example, a professor commented: "This is not what I expected it to be. In fact, everything is basically a presentation of things that I believe that most people know".

Perceived usefulness. There was a greater divergence in the perceived usefulness of the system between the two groups (Figure 5) The median for both groups was 4, whereas the mode was 4 for the CS group and 5 for the no-CS group ($SD = 1.20$ for the CS group, $SD = 0.75$ for the no-CS group). Only one participant in the CS group scored the system "very useful" and one participant scored it "not useful at all", whereas four people from the no-CS group found it "very useful". That is probably due to the fact that the people in the CS group already felt confident in discussions with their GPs, as they are highly educated and have excellent communication skills, evidenced by comments like, "If I were less skilled, I think that would have been useful". People in the no-CS group also tended to think that the usefulness may be related to education level, personality, and intelligence. In the interview, a

The tool is useful for showing you that you should prepare for GP visits. GP time is definitely short and being organised is important. (P10)

'What' questions are very good to help people prepare for the GP visit…they are the best bit of the whole thing. (P1)

Obviously, it's good advice to think about questions to ask a doctor and talk about to somebody else. I do that with my wife. (P5)

Several participants expressed that the system was useful to help them prepare for visits to the doctor.

I like the lesson that it's okay to ask GPs questions, and that there are alternatives. You can say "no" to a treatment. (P14)

It's very useful that it emphasises that healthcare is "a partnership" between the patient and the doctor. (P1)

The system helped some participants to realise the benefits of shared decision making.

no-CS participant stated, "I think it depends on your education and your personality and…Might be good for some people a little bit less bright".

Age- & Gender-related Stereotypes. Nine of the participants (5 from the CS group) perceived the dialogue as patronising. Some of them specifically disliked the repetition and reinforcement in the dialogue, describing it as "condescending" or "patronising", though they recognised that it may be useful for certain people. Others had stronger reactions. For example, one of the professors declared, "The trouble with this is making the assumption that older people are less intelligent, or older people are less well-informed by default". Five participants commented on the gender of the doctor, with one male participant asking "Why is the doctor a man?" – the majority of GPs in the UK are female.

Dialogue Structure. Nearly all participants (17 of 19) felt that the dialogue did not cover all possible options and/ or was not flexible enough to allow them to shape the discussion, saying, for example, "It's quite frustrating that it [the system] does not allow you to shape the conversation". Some participants considered the dialogue simplistic: "I think the system is conceptually good, but it did not cover all possibilities. I am amazed that the GP did not bring into discussion a critical issue like weight".

Individual Differences. Most participants (14 of 19; 6 of 8 in the CS group) discussed individual differences and how these should be approached within the system. For example, one participant (a retired doctor) suggested that the system should collect information about the users at the beginning. Thus, if the user is a doctor or a nurse, the VTD should know that "some of the answers are going to be coloured by that". Another participant suggested allowing the users to choose whether they would like to drive the dialogue, or prefer the VTD to do that.

National Context. Many participants (10 of 19; 4 of 8 in the CS group) noticed that the specific recommendations of the VTD are sometimes inappropriate for the UK context: "the corticosteroid injection is not common here in Scotland, but is common 'on the continent'".

Multimodal Interaction with virtual characters. There was a clear preference for using voice over clicks to avoid physical effort, with no difference between the two groups. However, because of some technical problems (e.g. the recording of open-ended answers was cut off too early), most of the participants used clicks. With regard to the VCs, participants' preferences varied. Some liked them (9 of 19; 4 of 8 in the CS group) for various reasons ("Daniel is good, he listens, but some doctors don't", "Daniel is very pleasant"). Other disliked them ("her smile is gruesome").

CONCLUSION & RECOMMENDATIONS

VTD systems to support SDM during GP visits were highly appreciated by our participants. They found the system useful for preparing the patients for SDM during GP visits and suggestion a number of ways to extend its use. However, more research should be conducted to better accommodate needs related to ageing and individual differences in this target population, and to adapt it to the national context. Reflecting on our findings, we have the following recommendations for the field:

This system is very useful to approach sensitive topics, such as pregnancy or menopause. (P5)

A candidate scenario could be parents bringing in their children. The question is when do you interact with the child versus the parent? These are known to be difficult situations which may benefit from some exploration before the doctor's visit. (P8)

Participants offered suggestions for future scenarios that could be developed for a tool of this kind.

R1: Avoid age-& gender-related stereotypes. Focus on physical difficulties that arise with ageing rather than communication issues. Provide customisation options, so that the user can choose the gender of the VCs.

R2: Make the scenarios more flexible and allow free input. Scenario design should be better informed by user studies involving a wider variety of stakeholders and encompassing various national contexts.

R3: Incorporate a user profile (covering personality, medical knowledge, interests) to deliver content that matches the user's needs and interests. This could be achieved through a personalised and adaptive digital environment.

REFERENCES

[1] T. Bosse and C. Gerritsen. 2016. Towards Serious Gaming for Communication Training - A Pilot Study with Police Academy Students. In *INTETAIN*.

[2] A.P. Cláudio, M.B. Beatriz Carmo, V. Pinto, and A. Cavaco. 2015. Virtual Humans for Training and Assessment of Self-medication Consultation Skills in Pharmacy Students. In *Proceedings ICCSE 2015*. IEEE, 175–180.

[3] G. Elwyn, D. Frosch, R. Thomson, N. Joseph-Williams, A. Lloyd, P. Kinnersley, E. Cording, D. Tomson, C. Dodd, S. Rollnick, A. Edwards, and M. Barry. 2012. Shared decision making: a model for clinical practice. *Journal of general internal medicine* 27, 10 (2012), 1361–1367.

[4] G. Elwyn, I. Scholl, C. Tietbohl, M. Mann, A.G. Edwards, C. Clay, F. Legare, T. van der Weijden, C.L. Lewis, R.M. Wexler, and D.L. Frosch. 2013. "Many miles to go ...": a systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC medical informatics and decision making* 13, Suppl 2 S14 (2013).

[5] K. A. Ericsson and H. A. Simon. 1993. *Protocol analysis.* MIT press Cambridge, MA.

[6] K. Gravel, F. Legare, and I.D. Graham. 2006. Barriers and facilitators to implementing shared decision-making in clinical practice: a systematic review of health professionals' perceptions. *Implementation science* 1 (2006), 16.

[7] J. Jeuring, F. Grosfeld, B. Heeren, M. Hulsbergen, R. IJntema, V. Jonker, N. Mastenbroek, M. van der Smagt, F. Wijmans, M. Wolters, and H. van Zeijts. 2015. Communicate! — A Serious Game for Communication Skills. In *Proc. EC-TEL 2015 (LNCS)*, Vol. 9307. Springer, 513–517.

[8] T. J. Muller, A. Heuvelink, K. van den Bosch, and I. Swartjes. 2012. Glengarry Glen Ross: Using BDI for Sales Game Dialogues. In *Proc. AIIDE 2012*. AAAI Press, 167–172.

[9] R. Pel-Littel, H. van Zeijts, N. Schram, H. H. Nap, and J. Jeuring. 2018. A training simulation for practicing shared decision making for older patients. In *Proceedings of ICTH 2018 (Procedia Computer Science)*, Vol. 141. Springer, 287–293.

[10] Johnny Saldaña. 2015. *The coding manual for qualitative researchers.* Sage.

[11] Jeff Sauro. 2011. Measuring usability. *A Practical Guide to the System Usability Scale* 1 (2011).

[12] D. Stacey, F. Legare, K. Lewis, M.J. Barry, C.L. Bennett, K.B. Eden, M. Holmes-Rovner, H. Llewellyn-Thomas, A. Lyddiatt, R. Thomson, and L. Trevena. 2017. Decision aids for people facing health treatment or screening decisions. *The Cochrane database of systematic reviews* 4 (2017).

[13] Z. Zhang and T. Bickmore. 2018. Medical Shared Decision Making with a Virtual Agent. *Proc. IVA '18* (2018), 113–118.