Qualitative Research in Evaluation. An Usability Evaluation Protocol for the Assistant on Care and Health Offline (ACHO)

Borja Rivero Jiménez(✉), David Conde Caballero, J. Jesús-Azabal, Jerónimo Luengo-Polo, Jara Bonilla-Bermejo, and Lorenzo Mariano Juárez

University of Extremadura, Cáceres, Spain
brivero@unex.es

Abstract. One of the problems associated with ageing that most concerns health professionals is low therapeutic adherence. In recent years, technological developments have appeared that can increase therapeutic adherence. To do so, it is necessary to know their usability and the possibility of use that they may have. In the framework of the project "International Institute for Research and Innovation on Ageing" (4IE), we have developed the Assistant on Care and Health Offline (ACHO), a voice assistant that provides medical appointments and medication reminders for patients. In this text we present a usability evaluation protocol for this voice assistant. We will use a multidimensional and multidisciplinary analysis in the framework of the Living Lab for the usability evaluation. Our methodology for measuring results includes three phases and different quantitative and qualitative research tools. The application of this methodology will allow us to develop a better prototype, increasing ease of use and improving the user experience.

Keywords: Terapeutic adherence · e-Health · Living lab · Voice assistant · Anthropology

1 Introduction

Increased life expectancy of a population is associated with better health conditions, but also an increase in age-related diseases [1]. In this sense, elderly patients are particularly susceptible to the phenomenon of non-adherence to medication, which can be defined as the degree to which recommendations or frequency of medication intake are met [2]. Rates of non-adherence to medication are higher in the elderly than in the rest of the population [3,4].

In recent years there has been an increase in the use of technology applied to medical health services. Technological development has opened a field of possibilities for better therapeutic adherence in patients. Some studies have appeared

J. García-Alonso and C. Fonseca (Eds.): IWoG 2020, LNBE, pp. 43–53, 2021.
https://doi.org/10.1007/978-3-030-72567-9_5

on electronic reminders using audio [5] or audiovisual devices [6] that, through reminders, facilitate therapeutic adherence.

A voice assistant is a software agent that interacts through voice activation using an intelligent speaker device [7]. One of the first studies to investigate these voice assistants applied to the health of voice assistants responded inconsistently and incompletely to a variety of questions [8]. Further studies have considered certain safety risks for patients and consumers [9]. Further work and research is needed to improve these types of devices in terms of their application to health.

Research points to the importance of adapting technology to the user experience by involving the end user in the development of the technology itself. Thus, by making adaptations according to the needs of the elderly target group, considerable increases in performance and acceptance by older people using a specific technology can be observed [10]. The "Living Labs" concept dates from the 1990s, and refers to an approach to innovation which involves a group of researchers collaborating with target users as co-creators in the development and validation of new products [11]. In this text we explain the different phases of the protocol that we followed to carry out a usability evaluation of a voice assistant for the therapeutic adherence of elderly people.

2 Objectives

This protocol is being developed as part of a larger project, "International Institute for Research and Innovation on Ageing (4IE)". The aim of this project is to develop technologies to improve the quality of life of older people in rural environments.

In this text we explain the different phases of the protocol that we followed to carry out a usability evaluation of a voice assistant for the therapeutic adherence of elderly people [12,13]. We propose an exhaustive, cyclical and multidisciplinary evaluation [14], with the following objectives:

1. To validate the correct usability of the designed prototype by observing the particular characteristics of the elderly and the context in which they live.
2. To identify problems and to develop possible guidelines for improvement.
3. To analyse the usability.
4. To involve real end-users in the process of validating the usability of the prototype.

3 The Design of Assistant on Care and Health Offline (ACHO)

The field of e-Healthcare is striving to adapt technological advances to the care of the elderly. The aim is to promote the autonomy of the elderly and thus facilitate their independent live at home [15], including personalized assistance [16]. Within this field we can find technological solutions that range from apps for mobile devices, to smartbands or clothes or what has been called Smart Home:

advanced technological systems. e-Healthcare is in full expansion driven by some programmes such as the European Union's Active Assisted Living Programme - Technology designed to improve quality of life for older people (AAL) [17]. These are interdisciplinary networks that have focused on the possibilities offered by technology and are particularly interested in voice assistants. The device consists of a home conversation interface that allows users to request and save information, as well as to perform a series of actions among which those related to health and care of the elderly are beginning to be explored [18].

The Assistant on Care and Health Offline (ACHO) works along the same lines as the rest of the e-healthcare projects, providing a new option in the development of intelligent environments for assisted living among elderly people. It is part of the Institute for Research and Innovation on Ageing 4IE project [19]. It is an interdisciplinary research focused on the regions of Extremadura (Spain) and Alentejo (Portugal), which is interested in knowing and describing the different problems of the elderly men and women in the area. Based on the knowledge of the reality, the aim is to validate and developed technological solutions that enable the application of innovations and new forms of care that take into account the particular characteristics of the elderly and the context in which they live.

ACHO is a voice assistant based on Snips structure [20]. Is the first voice assistant that does not store information via cloud, but keeps it locally on the device. The terminal does not need to have an internet connection. Our project, therefore, brings together on the one hand the global trend to work in assisted living environments, while on the other it takes into account the important contextual fact that access to an internet connection is not always possible depending on the type of user and the area where he or she lives. This solves two possible problems that could arise. In rural contexts such as we work in, Internet can be difficult to access due to the lack of an adequate infrastructure. And, very important: we work with older people who do not always have Internet services.

The ACHO app for smarthphone will allow the interaction between the health professional and the device in order to provide the health data that need to be remembered. The result is the participation of health professionals in a kind of "new forms of care". Some of its initial features are as follows:

1. Specification of the patient's profile. The basic information of the patient will be specified in the application along with details of the medical prescriptions and appointments always made by the health professional. This data will be stored in the application's internal database. The stored information, such as patient profiles, medical appointments and prescriptions, will only be accessible by the application.
2. Synchronisation with the voice assistant In this process the application will generate a temporary file with the information of the prescriptions. This file will be transmitted via Bluetooth with the voice assistant, which will process it and set up the corresponding reminders. The synchronisation process has been designed to avoid any possible loss of information or compromise of data security.

4 ACHO Evaluation Protocol

Scientific literature has described a good number of methodologies and tools used to ensure the quality of usability of a service [21,22]. Evaluating the usability and user experience of technology stands out as an essential step if it is to be significantly effective and meet its objectives [23], even more so when talking about older people due to the special characteristics of this age group. A number of previous products very similar to ours have not been accepted because they did not take into account this kind of issues54, so it is especially important to look for empirical evidence on how to improve the usability of different devices55.

Our multi-method approach is following some experiences that have already proved positive with devices very similar to the one proposed here [24–26]. The type of study - descriptive observation -, constitutes a usability analysis in several phases following the available evidence that advocates cyclical processes of analysis, prototyping, testing and refinement of the mechanisms of interaction with the user [23,24]. In this sense, it is important to emphasise that the passage from one phase to another is totally limiting, being impossible to access the subsequent phase without having satisfactorily overcome the previous one.

The work in this phase focuses on what is known as Living Lab [27], consisting of the strong involvement of end users in all phases of the development of prototypes. In our opinion, this method allows a more realistic validation of the environmental and holistic factors of the user, something for which the involvement of the anthropologists of the research team and their ability to interact with the users even in their own homes is fundamental. As Bevan et al. point out, the introduction of user-centred methods ensures that 'real products can be used by real people to perform their tasks in the real world' [28].

4.1 Methodology

Researchers, at the end of the evaluation process, must report the experience through the completion of different tools in order to assess various aspects of usability. Three tools have been selected.

Firstly, the System Usability Scale (SUS) [29]. It was chosen because it is in the public domain and can be freely used [30] and because it has shown great robustness and solidity in its results [24]. It is an economical tool, very simple to use and capable of being adapted for use in different situations such as the evaluation of software interfaces, web pages and applications, mobile phones, landlines, modems or voice systems (Fig. 1).

In addition to this assessment measure, two other scales recently validated by members of the Institute of Electronics and Telematics Engineering of Aveiro (Portugal) will be used to bring more consistency to the process [16]. These are the ICF-US I Scale and the ICF-US II Subscale [26]. The ICF-US I Scale allows the identification of general usability problems. The ICF-US II Scale allows the identification of possible barriers and/or facilitators, as well as identifying more specifically those elements that may require further work to improve the device.

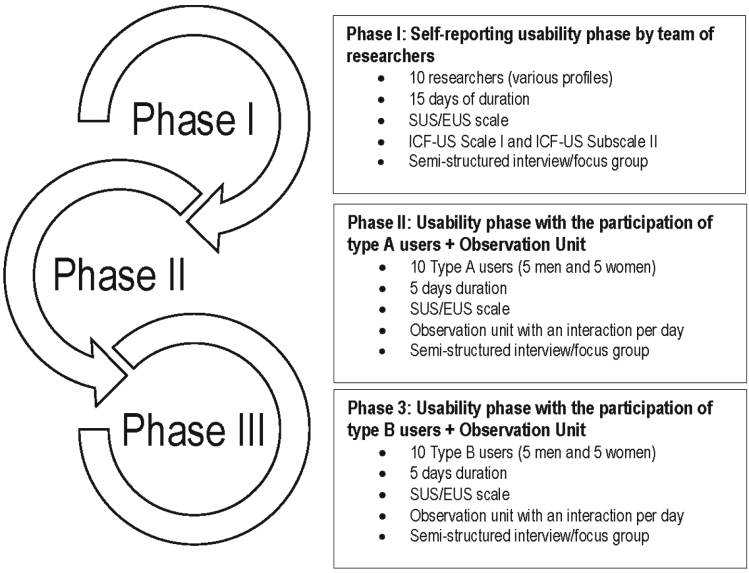

Fig. 1. Diagram of the research protocol

Previous evidence shows very positive experiences in its use, due to its use independently of the specific characteristics of the products being evaluated [24,25].

Similarly, all researchers face a semi-structured individual interview conducted by the team anthropologists. The researchers are asked to write down their impressions in a notebook. The aspects evaluated in this phase are:

- Positive or negative evaluation of the number of reminders made by the device.
- Evaluation of the way in which it is carried out: voice, tone...
- Evaluation of the name of the drug, which must be individualized.
- Feedback on consumption
- Existence of detected problems: Message saturation, message errors...
- Other subjective elements they may consider important.

4.2 Timing

The phases that we have decided to include in the design of the validation process are

1. Phase I: Self-reported usability phase by the research team, complemented by semi-structured interviews and/or Focus Groups.
2. Phase II: Usability phase with the participation of type A users + Observation Unit regarding "critical incidents" [31,32], complemented with semi-structured interviews and/or Focus Groups.

3. Phase III: Usability phase with the participation of type B users + Observation Unit regarding "critical incidents", complemented with semi-structured interviews and/or Focus Groups.

As in the initial stages of the project in which we carried out the ethnography, each and every one of the participants in the usability assessment sign an informed consent form after the actions to be carried out are made explicit. The data collection is anonymised and the participants receive all the information generated in the study. Quantitative data is stored and analysed using SPSS (Statistical Package for Social Sciences) version 22. Qualitative data is stored and categorised in the Dedalo Patform Intangible Heritage Management programme.

Phase I: Self-reported Usability by Research Team + Semi-structured Interviews. The first phase of the evaluation or analytical phase aims to determine whether our product is sustainable in terms of interface and functions. To this end, a total of 10 researchers from various knowledge profiles at the International Institute of Ageing - both Spanish and Portuguese - are evaluating the usability of the device over a period of 15 days. The researchers have been given a series of tasks through a script that they carry out at least three times a day, thus simulating the normal pattern of taking breakfast/lunch/dinner medication.

Phase 1 or the analytical phase is only overcome if

- A score on the SUS/EUS Scale above 68 is obtained in the final average of all users. This is the limit established in the scientific literature to determine the correct usability of a product [29].
- A score on the ICF-US I Scale higher than 10 is obtained in the final average of all users. This is the limit established in the scientific literature to determine the existence of a correct usability of a product [33].
- The results obtained in the ICF-US Scale II are analyzed.
- All the semi-structured interviews are carried out.
- The improvements suggested by the analysis of the instruments used are incorporated into the device. In this sense, the design changes can reduce certain errors and facilitate the usability and acceptance of the user [34,35]. A new validation cycle is not ruled out once the necessary improvements have been incorporated.

Phase II: Usability with the Participation of Type a Users + Observation Unit. In the second phase of the evaluation, information is collected on the usability and satisfaction of real users with a physical implementation of the prototype in real but controlled contexts. It is therefore an empirical model. The participants in this phase are what we have called TYPE A users: people over 65 years of age from rural areas without cognitive and/or sensory impairments, selected on a non-random basis after recommendation of suitability by professionals from the Extremadura Health System in the selected locality. The selected sample must have the capacity to detect and report possible failures,

"critical incidents" and/or problems in the interaction with the prototype. At least 10 people are selected, possibly 5 men and 5 women in different age ranges from 65 years old.

Users will be trained on the actions to be carried out beforehand, and are accompanied in at least 1 of the three daily interactions foreseen by what we have called Observation Units. These are researchers who observe and evaluate the process of use and interaction in the user's context, which allows us to collect significant information to understand what changes need to be made in the environmental factor so that it can be better adapted to its users and its functionality can be improved [35]. In addition, the Observation Units are responsible for recording so-called "critical incidents" [36], that is, all situations that deviate from normality. Although there is no structure or standardised procedure for recording "critical incidents", their use has been described in the scientific literature as "very appropriate" [25,26].

These same researchers, at the end of the stipulated period of time, conduct a semi-structured interview in the same sense as indicated in Phase I, while helping people to report on usability through the SUS Scale - in this case selected because it is easier to administer. As in the previous phase, the possibility of holding Focus Groups with various users is not ruled out if the information collected is not considered sufficient. The estimated time of implementation is 5 days.

Phase 2 is considered to have been passed only if

- A score on the SUS/EUS Scale above 68 is obtained in the final average of all users. This is the limit established in the scientific literature to determine the existence of correct usability of a product.
- All semi-structured interviews are conducted.
- The improvements suggested from the analysis of the empirical evidence generated are incorporated into the device. In this sense, the design changes can reduce certain errors and facilitate the usability and acceptance of the user. A new validation cycle is not ruled out once the necessary improvements have been incorporated.

Phase III:Usability with the Participation of Type B Users + Observation Unit. Finally, the third phase of the evaluation of the pilot test aims to assess usability under normal, uncontrolled operating conditions. The so-called TYPE B users are people over 65 years old from rural areas with the only inclusion criterion of having non-critical medication according to the recommendations of the professionals of the Extremadura Health System in the selected locality.

The users, previously trained on the actions to be carried out and the objectives to be pursued, are accompanied in all the daily interactions planned by Observation Units. These researchers conduct a semi-structured interview in the same way as in the previous cases and help people to report on usability through the SUS Scale. As in the previous phases, the possibility of holding focus groups with various users is not ruled out if the information collected is not considered

sufficient. The estimated time of implementation is 5 days. Through the analysis of all the empirical material generated, we intend to measure the usability and functionality attributes of the prototype.

Phase 3 and therefore the evaluation of the usability of our prototype is considered to have been overcome only if:

- A score on the SUS/EUS Scale above 68 is obtained in the final average of all users. This is the limit established in the scientific literature to determine the existence of correct usability of a product.
- All semi-structured interviews are conducted.
- The improvements suggested from the analysis of the empirical evidence generated are incorporated into the device. In this sense, the design changes can reduce certain errors and facilitate the usability and acceptance of the user. A new validation cycle is not ruled out once the necessary improvements have been incorporated.

At the end of the validation cycle proposed here, the possibility of incorporating new functionalities is assessed, in which case the validation process would be the same as that indicated here. If this is not the case, the next phase would include an assessment of the safety of the use of the prototype with respect to the reminder when taking medicines.

5 Conclusions

This evaluation protocol will provide us in the coming months with results that will give us a reliable evaluation of the Assistant on Care and Heatlh Offline (ACHO). The evaluation of the usability of health technologies is essential, especially for those technologies that will be used by older people. To counteract certain difficulties in evaluation, we want to have technology evaluation tools that allow us to know the reality of use in their context. In future publications we will offer the concrete results and the difficulties encountered in the evaluation of this voice assistant.

Acknowledgment. This work was supported by the 4IE+ project (0499_4IE_4_PLUS_4_E) funded by the Interreg V-A España-Portugal (POCTEP) 2014–2020 program.

References

1. WHO: World report on ageing and health. Technical report, World Health Organization, Geneva, Switzerland (2015). https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf;jsessionid=182A9CD45B7035C521E9DC1A8BCDE127?sequence=1
2. Pisano González, M.M., González Pisano, A.: La modificación de los hábitos y la adherencia terapéutica, clave para el control de la enfermedad crónica. Enfermería Clínica **24**(1), 59–66 (2014). https://doi.org/10.1016/j.enfcli.2013.10.006

3. Smith, D., Lovell, J., Weller, C., Kennedy, B., Winbolt, M., Young, C., Ibrahim, J.: A systematic review of medication non-adherence in persons with dementia or cognitive impairment. PloS One **12**(2), e0170651 (2017). https://doi.org/10.1371/journal.pone.0170651
4. Haynes, R.B., Ackloo, E., Sahota, N., McDonald, H.P., Yao, X.: Interventions for enhancing medication adherence. Cochrane Database Syst. Rev. (2), CD000011 (2008). https://doi.org/10.1002/14651858.CD000011.pub3
5. Andrade, A.S.A., et al.: A programmable prompting device improves adherence to highly active antiretroviral therapy in HIV-infected subjects with memory impairment. Clin. Infect. Dis. Official Publ. Infect. Dis. Soc. Am. **41**(6), 875–882 (2005). https://doi.org/10.1086/432877
6. Christensen, A., et al.: The impact of an electronic monitoring and reminder device on patient compliance with antihypertensive therapy: a randomized controlled trial. J. Hypertension **28**(1), 194–200 (2010). https://doi.org/10.1097/HJH.0b013e328331b718
7. Chung, A.E., Griffin, A.C., Selezneva, D., Gotz, D.: Health and fitness apps for hands-free voice-activated assistants: content analysis. JMIR mHealth uHealth **6**(9), e174 (2018). https://doi.org/10.2196/mhealth.9705
8. Miner, A.S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., Linos, E.: Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern. Med. **176**(5), 619–625 (2016). https://doi.org/10.1001/jamainternmed.2016.0400
9. Alagha, E.C., Helbing, R.R.: Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform. **26**(1) (2019). https://doi.org/10.1136/bmjhci-2019-100075
10. Thilo, F.J., Bilger, S., Halfens, R.J., Schols, J.M., Hahn, S.: Involvement of the end user: exploration of older people's needs and preferences for a wearable fall detection device - a qualitative descriptive study. Patient Prefer. Adherence **11**, 11–22 (2017). https://doi.org/10.2147/PPA.S119177
11. Almirall, E., Wareham, J.: Living labs: arbiters of mid- and ground-level innovation. Technol. Anal. Strateg. Manage. **23**(1), 87–102 (2011). https://doi.org/10.1080/09537325.2011.537110
12. Jesús-Azabal, M., Rojo, J., Moguel, E., Flores-Martin, D., Berrocal, J., García-Alonso, J., Murillo, J.M.: Voice Assistant to Remind Pharmacologic Treatment in Elders. In: García-Alonso, J., Fonseca, C. (eds.) Gerontechnology. IWOG 2020, pp. 123–133. Springer, Cham (2020)
13. Moguel, E., Jesus Azabal, M., Flores-martin, D., Berrocal, J., García-Alonso, J., Murillo, J.M.: Asistente de voz para el recordatorio de tratamiento farmacológico. In: Jornadas de Ingeniería del Software y Bases de Datos (JISBD). Cáceres (2019). hdl.handle.net/11705/JISBD/2019/078
14. Hoppe, A.: Technological stress-mental strain of younger and older users if technology fails. In: Wichert, R., Eberhardt, B. (eds.) Ambient Assisted Living, pp. 17–29. Springer, Heidelberg (2011)
15. Ennis, A., Rafferty, J., Synnott, J., Cleland, I., Nugent, C., Selby, A., McIlroy, S., Berthelot, A., Masci, G.: A smart cabinet and voice assistant to support independence in older adults. In: International Conference on Ubiquitous Computing and Ambient Intelligence, pp. 466–472 (2017). https://doi.org/10.1007/978-3-319-67585-5_47
16. European Commission: Shaping Europes digital future. Technical report, Brussels (2020)

17. The international classification of functioning, disability and health as a conceptual model for the evaluation of environmental factors. Procedia Comput. Sci. **14**, 293–300 (2012). https://doi.org/10.1016/j.procs.2012.10.033
18. Pradhan, A., Mehta, K., Findlater, L.: "Accessibility came by accident": use of voice-controlled intelligent personal assistants by people with disabilities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI 2018, pp. 1–13. ACM Press, New York (2018). https://doi.org/10.1145/3173574.3174033
19. Garcia-Alonso, J., Berrocal, J., Rivero Jiménez, B., Conde Caballero, D., Mariano Juárez, L., Murillo, J.M., Mendes, D., Fonseca, C., Lopes, M., Perez-Vereda, A., Canal, C.: Instituto Internacional de Investigación e Innovación del Envejecimiento. In: Jornadas de Ingenieria del Software y Bases de Datos, JISBD 2018 (2018)
20. Coucke, A., Saade, A., Ball, A., Bluche, T., Caulier, A., Leroy, D., Doumouro, C., Gisselbrecht, T., Caltagirone, F., Lavril, T., Primet, M., Dureau, J.: Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. CoRR **abs/1805.10190** (2018). http://arxiv.org/abs/1805.10190
21. Martins, A.I., Queirós, A., Rocha, N.P., Santos, B.S.: Avaliação de usabilidade: uma revisão sistemática da literatura. RISTI-Rev Ibérica Sist E Tecnol Informação, pp. 31–43 (2013)
22. Best, M.L., Smyth, T.N.: Global/Local Usability: Locally Contextualized Usability in the Global South BT - Global Usability, pp. 9–22. Springer, London (2011). https://doi.org/10.1007/978-0-85729-304-6_2
23. Martins, A., Queirós, A., da Rocha, N.: Usability evaluation of products and services: a systematic review. In: E-Soc (2013)
24. Martins, A.I., Rosa, A.F., Queirós, A., Silva, A., Rocha, N.P.: European Portuguese validation of the system usability scale (SUS). Procedia Comput. Sci. **67**, 293–300 (2015). https://doi.org/10.1016/j.procs.2015.09.273
25. Jegundo, A.L., Dantas, C., Quintas, J., Dutra, J., Almeida, A.L., Caravau, H., Rosa, A.F., Martins, A.I., Queirós, A., Rocha, N.P.: Usability evaluation of a virtual assistive companion. In: World Conference on Information Systems and Technologies, pp. 706–715 (2019)
26. Martins, A.I., Queirós, A., Silva, A.G., Rocha, N.P.: ICF based usability scale: evaluating usability according to the evaluators' perspective about the users' performance. In: Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, pp. 378–383 (2016)
27. Dekker, R., Franco Contreras, J., Meijer, A.: The living lab as a methodology for public administration research: a systematic literature review of its applications in the social sciences. Int. J. Public Adm. **43**(14), 1207–1217 (2020). https://doi.org/10.1080/01900692.2019.1668410
28. Bevan, N.: European usability support centres: support for a more usable information society. In: European Telematics: Advancing the Information Society. Proceedings of TAP Annual Concertation Meeting, pp. 50–65 (1998)
29. Brooke, J.: Others: SUS-a quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
30. Sauro, J.: A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. LLC, Denver (2011)
31. Ivory, M.Y., Hearst, M.A.: The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. (2001). https://doi.org/10.1145/503112.503114

32. Tsiourti, C., Ben Moussa, M., Quintas, J., Loke, B., Jochem, I., Lopes, J.A., Konstantas, D.: A virtual assistive companion for older adults: design implications for a real-world application. In: Bi, Y., Kapoor, S., Bhatia, R. (eds.) Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016. IntelliSys 2016. LNNS (2018). https://doi.org/10.1007/978-3-319-56994-9_69
33. Martins, A.I., Queirós, A., Rocha, N.P.: Validation of a usability assessment instrument according to the evaluators' perspective about the users' performance. Univers. Access Inf. Soc. **19**(3), 515–525 (2020). https://doi.org/10.1007/s10209-019-00659-w
34. Kalimullah, K., Sushmitha, D.: Influence of design elements in mobile applications on user experience of elderly people. Procedia Comput. Sci. **113**, 352–359 (2017). https://doi.org/10.1016/j.procs.2017.08.344
35. Holden, R.J., Campbell, N.L., Abebe, E., Clark, D.O., Ferguson, D., Bodke, K., Boustani, M.A., Callahan, C.M.: Usability and feasibility of consumer-facing technology to reduce unsafe medication use by older adults. Res. Soc. Adm. Pharm. RSAP **16**(1), 54–61 (2020). https://doi.org/10.1016/j.sapharm.2019.02.011
36. Gitlin, L.N.: The role of social science research in understanding technology use among older adults. Self-care in later life, pp. 142–169 (1998)