ABSTRACT

Background: it is important to standardize the evaluation and reporting procedures across usability studies to guide researchers, facilitate comparisons, and promote high-quality studies. A first step to standardizing is to have an overview of how experts-based usability evaluation studies are reported across the literature. Objectives: to describe and synthesize the procedures of usability evaluation by experts that are being reported to conduct inspection usability assessments of digital solutions relevant for older adults. Methods: a scoping review of reviews was performed using a five-stage methodology to identify and describe relevant literature published between 2009 and 2020 as follows: i) identification of the research question; ii) identification of relevant studies; iii) select studies for review; iv) charting of data from selected literature; and v) collation, summary, and report of results. The research was conducted on five electronic databases: PubMed, ACM Digital Library, IEEE, Scopus, and Web of Science. The articles that met the inclusion criteria were identified, and data extracted for further analysis, including evaluators, current usability inspection methods, and instruments to support usability inspection methods. Results: a total of 3958 articles were identified. After a detailed screening, 12 reviews matched the eligibility criteria. Conclusion: overall, we found a variety of unstandardized procedures and a lack of detail on some important aspects of the assessment, including a thorough description of the evaluators and of the instruments used to facilitate the inspection evaluation such as heuristics checklists. These findings suggest the need for a consensus framework on the experts’ assessment of usability that informs researchers and allows standardization of procedures.

CCS CONCEPTS
• Applied computing, Life and medical sciences, Health care information systems;

KEYWORDS
Usability, Older adults, Scoping review, Digital solutions, Inspection methods, Heuristics, Experts’-based usability

ACM Reference Format:
Anabela G Silva, Ana Isabel Martins, Hilma Caravau, Margarida Almeida, Telmo Silva, Óscar Ribeiro, Gonçalo Santinha, and Nelson P. Rocha. 2020. Experts Evaluation of Usability for Digital Solutions Directed at Older Adults: a Scoping Review of Reviews. In 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion (DSAI 2020), December 02–04, 2020, Online, Portugal. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3439231.3439238

1 INTRODUCTION
The challenges that contemporary society faces, due to aging, are also opportunities for technological and socio-economic innovation. Older adults require security and comfort and an adequate level of support and social integration. Despite the evident decrease in the functional capacity of people as they age, a view that associates only a dependent way of life to the older population is very reductive and does not coincide with current perspectives, namely healthy and active aging, which consider the need for optimization of opportunities for social participation, health and the safety of elderly individuals, to promote their autonomy and independence [1]. These goals may be supported by new forms of care that involve the
use of technological solutions to mitigate disability and promote functioning. Digital solutions play an important role in areas like optimization and personalization of healthcare provision [2], and promotion of healthy lifestyles to minimize loneliness and social exclusion [3] [4] [5], or to improve safety, independence and confidence [2]. Digital solutions can improve the lives of the elderly at an acceptable cost, but this only occurs if technologies are adjusted to the challenges and specificities of this population. Issues such as poor digital literacy, health problems, and chronic illnesses, loss of visual and auditory acuity, or changes in fine motor skills should be considered [6] [7] as they may hinder the use of technology. To be effective, new technological developments must bring added value to those who use them, so it is essential to adapt technologies to their needs. To ensure that a digital solution is fully tailored to its users, a robust evaluation process must be considered, especially in terms of usability [8].

The ISO 9241-11 standard defines usability as the measure by which a product can be used by specific users to achieve specific objectives with effectiveness, efficiency, and satisfaction, in a context of specific use [9]. The same standard stresses that usability is dependent on the context of use, meaning that the level of usability obtained depends on the specific circumstances in which the product is used [9]. The usage context includes users, tasks, equipment (i.e., hardware and software), and the physical and social environment since all these factors can influence the usability of a system. Thus, the usability of a system corresponds to the objective of making it adapted to the body and mind of its user in a given context [10].

Usability assessment is part of the interactive design, prototyping, and validation cycles, which are a very important component of the overall design process [11]. When human-computer interaction is built considering usability criteria, it enables an intuitive, efficient, memorable, effective, and pleasant interaction. The improvement in usability has several benefits, namely increased efficiency, higher productivity, reduced errors, less need for training, improved acceptance, assisting non-specialized users, and assisting users with limitations such as older adults [12] [13].

Usability evaluation can be empirical (based on data from real users) or analytical (based on the analysis by specialists of an interactive system and / or potential interactions). Empirical models include test and inquiry methods, while analytical models involve the inspection of the digital solution by experts to assess the various aspects of user interaction [14] [15].

This review focuses on inspection methods that imply the use of standards, heuristics, or guidelines by experts when performing a usability evaluation. Heuristics are an established set of principles of interface design and usability arranged on written sentences [16], having a dual-use both for creating an interface (typically used by designers and developers) and to evaluate its compliance in terms of usability (typically performed by usability evaluators) [17]. Experts have a fundamental role in both cases, as they are key contributors to all stages of development, from product design to evaluation.

The importance of usability standards is that they increase speed and decrease the cost of technological development, along with providing better consistency to the evaluated products [18]. Many inspection methods lend themselves to the inspection of user interface specifications that have not necessarily been implemented yet, which means that inspection can be performed early in the usability engineering lifecycle [19] [20]. Regardless of the development phase, it is important to standardize the evaluation and reporting of usability procedures across studies. This will contribute to high-quality usability studies, with a greater probability of identifying usability problems that will favor the quality of the products evaluated or developed. A first step to standardize is to have an overview of how the inspection procedures for assessing usability are reported throughout the literature. This scoping review of reviews aimed to describe the procedures of assessment of usability evaluation by experts that are being reported to conduct an inspection usability assessment of digital solutions relevant for older adults.

2 METHODS

This study followed a previous 5-stage framework for scoping reviews [21] [22]. These stages are: i) identification of the research question; ii) identification of relevant studies; iii) selection of relevant studies; iv) charting the data; and v) collating, summarizing, and reporting the results of the review. A scoping review of the literature was the method selected as it aims to map key concepts, summarize a range of evidence, especially in complex fields, and identify gaps in the existing literature [21], [22].

2.1 Identification of the Research Question

The research question guides the subsequent stages of the review. The research question for the present scoping review resulted from the knowledge and experience of the research team. There is an apparent lack of consensus in the academic literature regarding the procedures that should be used by experts to evaluate the usability of digital solutions. Therefore, the following research question was defined: "What are the current practices for the experts’-based evaluation of the usability of digital solutions (e.g., evaluators, methods, and instruments) relevant for the older adult population?"

2.2 Identification of Relevant Studies

A broad search strategy was used not including any reference to older adults as this might narrow the search, although this topic was subsequently used to select relevant studies. Therefore, the research expression “usability” OR “user experience” was used in the electronic search carried out in PubMed, ACM Digital Library, IEEE, Scopus, and Web of Science. Databases were searched for English language articles published between January the 1st of January 2009 and the 23rd of January 2020. The research terms were broadly defined to avoid the exclusion of studies with potential interest for analysis.

2.3 Selection of Relevant Studies

All references were imported into Mendeley software (Elsevier, North Holland), and duplicates removed. The first 300 abstracts were screened by three reviewers (HC, AGS & NPR). Differences in judgment were used to refine inclusion and exclusion criteria and were discussed until consensus was reached. Screening of the remaining abstracts was then performed by one reviewer (HC). Similarly, the first 10 full articles were screened by two reviewers.
Analyses (PRISMA) flow diagram for this scoping review is presented in Figure 1. A total of 12 reviews were included in the present scoping review. Table 1 presents the main characteristics of the included reviews (i.e., authors and year, purpose, and number of included studies).

3 RESULTS - CHARTING AND COLLATING DATA

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for this scoping review is presented in Figure 1. A total of 12 reviews were included in the present scoping review. Table 1 presents the main characteristics of the included reviews (i.e., authors and year, purpose, and number of included studies).

3.1 Evaluators

Only three out of the 12 reviews reported on the number of the evaluators involved in the usability inspection. Table 2 presents the type of technology and evaluators' number and characteristics of the included reviews. Regarding experts' characteristics, five studies do not report any, and the ones that do so are ambiguous about both the expertise level and the domain experience of those performing the evaluation. There is also a great variability on what is considered to be an "expert" as most reviews failed to provide information on how human-computer interaction or usability experts were determined. Also, in four reviews students were involved in the evaluations. Moreover, the term "expert" was used with two different meanings: in some reviews this term was used to refer to the professional with knowledge in usability and user experience; and in others to refer to the professional with knowledge on the field of use of the technology under development, i.e., if developing a technology for persons with diabetes, an expert would be, for example, a medical doctor (i.e., domain expert).

In this scoping review, seven inspection methods were identified, which are detailed in Table 3 heuristic evaluation, cognitive walkthrough, task analysis, perspective-based inspection, and guideline review, metaphor of human-thinking and systematic usability evaluation. Seven reviews refer to more than one inspection method, as is the case of the review of Hussain et al. [31] that referred to the combination of heuristic evaluation and guideline review, while five reviews reported only one inspection method, as is the case of the review of Allison et al. [23] that used only cognitive walkthrough. In addition to the inspection methods that were identified in the scoping review, others were not detected though being widely described in the literature, namely consistency inspection, and formal usability inspection.

3.2 Current Usability Inspection Methods

Heuristic evaluation was the most used usability inspection method found in this scoping review, as it was reported in 11 of the 12 studies included. The second most used method was the cognitive walkthrough, reported in half of the reviews (n=6). The third most reported method was guideline review, present in four out of 12 studies. The method perspective-based inspection was described in two studies, and the methods task analysis, metaphor of human-thinking, and systematic usability evaluations were reported by just one study.

Of the reviews that reported inspection using heuristic evaluation, nine reported generic heuristics. Nielsen’s heuristics were reported in all reviews, and one of the reviews also reported the Gerhardt-Powals heuristics [36]. Likewise, specific heuristics for different types of technology and users were reported in six of the 12 reviews. These were heuristics for gesture interaction [37] [38] [39] [40] [41] [42] [43], mobile interfaces [44],[45],[46],[47],[48],[49],[50],[51],[52], web sites [53] [54] [55] [56] [57], interactive television (iTV) [58] [59], m-commerce applications [60], mobile augmented reality [61], security in online health social networks [62], health information systems [63], egovernance [64] human-robot interaction [65], assistive robotic [66], and medical devices [67]. In addition, we also retrieved heuristics that were specific for both type of technology and users: heuristics for smartphone applications [50], mobile based applications [48], and websites [53] directed at older persons.

Regarding the reviews that reported guideline evaluations, two ISO standards were used to check its compliance in inspection evaluation. ISO 9241-11 [68], which stands for Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts, was reported in all four reviews [28] [29] [31] [32], and ISO 9126-11 [69], which stands for Software Engineering — The Product Quality, was reported in one review [28].
This scoping review of reviews aimed to synthesize current practices for the experts’ evaluation of the usability of digital solutions relevant for older adults. Results suggest that the characteristics of study evaluators are often only briefly reported, and no agreement seems to exist on the characteristics reported for each of them. Often the reader is not informed of how many experts performed the evaluation or on the expertise level and domain experience of those performing the evaluation. There is no consensus about the ideal number of experts but, according to Nielsen [19], three to five experts are generally required to carry out a usability evaluation using an inspection method. Inspection methods are complex and multifaceted, and evaluators who have usability and/or local systems expertise are critical for effective evaluation [26].

There is also a great variability on what is being considered an expert as most reviews failed to provide information on how human-computer interaction or usability experts were determined (i.e., whether it was based on formal educational years of experience, profession, or any other criteria). There are no specific guidelines on the literature regarding the characteristics that specialists should have; however it is suggested that these experts must have proven experience in human-computer interaction. They should typically be usability experts and preferably present domain expertise in the industry type of the digital solution under development. It is very important to train the evaluators, so they know exactly what they are meant to do and cover during their evaluation. Inspection evaluation is heavily dependent on the skills of experts involved in the study, thus, lack of information related to the level of expertise of the evaluator can introduce bias into the evaluation [30].

### 3.3 Instruments to Support Usability Inspection

The results of this scoping review also revealed the utilization of heuristics checklists that are a predefined set of verification points used to analyze heuristics, and against which user interface components are compared in an inspection evaluation [70]. These instruments aggregate and conjugate different heuristics for a given digital solution. This scoping review identified three checklists to support inspection methods: the Mobile-specific Heuristic Evaluation checklist [52], the Usability of Web-based Information Systems checklist (based on ISO 9241 and Nielsen’s heuristics) [71] and the MiLE+ that is based on 82 technical heuristics (i.e., 36 navigational heuristics, eight content heuristics, seven technology/performance heuristics, and 31 interface design heuristics) [72].

### Table 1: General characteristics of included reviews

| Authors | Purpose of the paper/study | Number of studies in the review |
|---------|-----------------------------|---------------------------------|
| Allison et al. [23] | Reviewing methodologies and techniques to evaluate websites; provide a framework of the appropriate website attributes that could be applied to any future website evaluations. | 69 |
| Baharuddin et al. [24] | Proposing a set of usability dimensions that should be considered for designing and evaluating mobile applications. | Not referred |
| Chuan et al. [25] | Creating a set of gesture-specific heuristics that would complement existing general usability heuristics for the design and testing of new gestural interaction. | 6 |
| Costa et al. [16] | Identifying the heuristics and usability metrics used in the literature and/or industry; based on the review results, this work presents a proposal of a set of usability heuristics focused on mobile applications on smartphones, considering the User, Task, and Context as usability factors and Cognitive Load as an important attribute of usability. | 8 |
| Ellsworth et al. [26] | Revising methods employed for usability testing on electronic health records; the aim was to evaluate methodological and reporting trends present in the current literature by investigating published usability studies of EHRs. | 120 |
| Fernandez et al. [27] | Analysing which usability evaluation methods have proven to be the most effective in the Web domain | 18 |
| Fernandez et al. [28] | Analyzing which usability evaluation methods have been employed to evaluate Web applications over the last 14 years. | 206 |
| Fu et al. [29] | Assessing the usability of diabetes mobile applications developed for adults with type 2 diabetes. | 7 |
| Hermawati & Lawson [30] | Presenting a comprehensive review of 70 studies related to usability heuristics for specific domains; the aim is to review the processes that were applied to establish heuristics in specific domains and identify gaps in order to provide recommendations for future research and area of improvements. | 70 |
| Hussain et al. [31] | Reviewing the relevant and appropriate usability dimensions and measurements for m-banking application; proposing a set of usability dimensions and measurements for m-banking evaluation | 49 |
| Lim et al. [32] | Identifying, studying, and analyzing existing usability metrics, methods, techniques, and areas in mobile augmented reality learning. | 72 |
| Yen & Bakken [33] | Reviewing and categorizing health information technology usability study methods and to provide practical guidance on health IT usability evaluation. | 346 |
Great confusion was also apparent in the use of the term "expert", as in some reviews this term refers to the professional with knowledge on the field of use of the technology under development (e.g., a physician involved in the development of a technology for monitoring diabetes) or as the professional with knowledge in usability and user experience (e.g., an human-computer interaction specialist assigned to evaluate the interface of the technology for monitoring diabetes). Both experts are valuable and have very different roles, as each one’s expertise is essential to ensure that the technology is both not only usable, but also adapted to the users’ needs.

The literature states that heuristic evaluation is one of the most used usability evaluation methods [34]. This review supports such previous finding, as this method was reported in all included reviews, except for one. One of the most important steps in an inspection evaluation is the establishment of an appropriate list of heuristics and the selection of typical tasks that should be considered during the evaluation. Heuristics can help the evaluators focus their attention on certain issues and, for that reason, it is important to use generic heuristics to ensure a global assessment of the interface. These can be complemented with specific heuristics that address concrete aspects of that type of digital solutions or type of users.

In this scoping review, generic and specific heuristics were retrieved. Regarding generic heuristics, Nielsen’s heuristics [35] appeared in every single review that addressed heuristic evaluation,
workload, in this scoping review only three heuristic checklists which has led to the growing establishment of instruments to sup-

which suggests that this is considered the gold standard of heuristic evaluation. More specific heuristics were reported for different types of technologies, such as mobile interfaces, gestural interaction, and web sites. We also retrieved heuristics that were specific for both type of technology and users, in this case, older adults. This type of heuristics considers not only the attributes of the type of technology but also the characteristics of the type of users. In the case of older adults, issues such as lower digital literacy, the need for larger fonts and buttons, or difficulties in touch interaction should be considered.

Inspection methods present difficulties in terms of application, which has led to the growing establishment of instruments to sup-

port inspection using heuristics: the heuristics checklists. These tools are attempts to objectify inspection using heuristics. In prac-

tice, these instruments aggregate and conjugate different heuristics for a given digital solution, for example, a checklist of heuristics for mobile phones. The use of checklists can support the verification of heuristics since they allow to detail each heuristic in more exhaustive aspects and thus considerably facilitate the evaluation. The use of heuristics checklists also simplifies the analysis and results in interpretation, helping to prioritize the need to solve usability problems. Despite the advantages of using checklists reported in the literature, which include reducing memory load, errors, and workload [70], in this scoping review only three heuristic checklists were reported.

The checklists were used as a practical design support tool, or as an evaluation support tool used to suggest necessary areas for interface redesign and it can be used throughout the design process in evaluating multiple design alternatives [73]. For example, using usability checklists enhances the effectiveness and efficiency of heuristic evaluation [74] and it has been found that using a checklist leads to the identification of 90% of usability problems. These findings contrast with a previous report [75] suggesting that heuristic evaluation (without a checklist) typically do not predict more than 30 to 50% of usability problems. These results suggest that the use of checklists might improve the traditional heuristic evaluation technique [75], however, according to the results of this scoping review, this practice does not seem to be disseminated. Despite not being reported in the 12 reviews included in this scoping review, some checklists are widely used in the development and evaluation of specific technologies, such as the checklist for mobile phone user interfaces [74], and the accessible smartphone interface design heuristics checklist [65]. There is also a checklist specific for older adults, the Touch-based Mobile Heuristics Evaluation for elderly people for evaluating the usability of touch-based mobile phones [48].

This research study presents some limitations that are directly related to the typology of reviews, such as the absence of assessment of the quality of the included reviews and the quantitative summary of findings [76]. Also, this is an area where a large amount of publications takes place as conference proceedings and we did not search specifically for these. Nevertheless, it is likely that by including mostly reviews published in journals that these are more comprehensive documents, since conference proceedings tend to have lower word counts for included papers. Also, the judgment made to decide whether a manuscript was on a product or technology that could be of use for older adults was a subjective judgment made by the authors and could have biased the results towards the field of health.

5 CONCLUSION

Overall, we found a lack of a detailed description of the evaluators that are central to the inspection methods application. The inspection method most reported was heuristic evaluation, but even this one has drawbacks, namely the low investment in the application of instruments that facilitate its use, such as heuristics checklists. These findings suggest the need for a consensus framework on the experts’ assessment of usability that informs researchers and allows a standardization of procedures.

Usability inspection evaluation is an essential component of the usability evaluation of digital solutions; however, it cannot be con-

sidered in isolation or stay from the usability test evaluation that involves real end-users in the evaluation sessions. Ideally, evaluation with experts should take place before testing with real users, so that the biggest usability problems are discovered and addressed before they prevent participants from discovering harder to spot workflow specific issues. Furthermore, using usability inspection methods allows a combination of methods of different nature, which is considered good practice in terms of usability evaluation. Even though inspection methods find many usability problems that are not identified by users testing, it is also the case that it may miss some problems that can only be found by users. In addition, evaluators are probably likely to overlook usability problems if the system is highly domain-dependent and they have little domain expertise [77]. Eliminating usability errors before usability testing allows the testing to reveal more unique and subtle usability concerns. In this sense, inspection evaluation is complementary to user testing and, therefore, fundamental for the real adaptation of the technology to the end-users and the context of use.

ACKNOWLEDGMENTS

SHAPES – Smart and Health Ageing through People Engaging in Supportive Systems – is funded by the Horizon 2020 Framework Programme of the European Union for Research Innovation. Grant agreement number: 857159 - SHAPES – H2020 – SC1-FA-DTS – 2018-2020.

REFERENCES

[1] World Health Organization, “Active Ageing: A Policy Framework. A contribution of the World Health Organization to the second United Nations World Assembly on Aging,” Geneva, 2002.
[2] F. Wagner, J. Basran, and V. D. Bello-Haas, “A Review of Monitoring Technology for Use With Older Adults,” J. Geriatr. Phys. Ther., vol. 35, no. 1, pp. 28–34, 2012.
[3] A. Gilson, D. Dodds, A. Kau, M. Potteiger, and J. H. Ford II, “Using Computer Tablets to Improve Moods for Older Adults With Dementia and Interactions With Their Caregivers: Pilot Intervention Study,” JMIR Form. Res., vol. 3, no. 3, p. e14530, Sep. 2019.
[4] C. Leone, J. S. L. Lim, A. Stern, J. Charles, S. Black, and R. Baecker, “Communication technology adoption among older adult veterans: the interplay of social and cognitive factors,” Aging Ment. Health, vol. 22, no. 12, pp. 1666–1677, Dec. 2018.
[5] W. A. Rogers and T. L. Mitzner, "Envisioning the future for older adults: Autonomy, health, well-being, and social connectedness with technology support," Futures, vol. 87, pp. 133–139, Mar. 2017.
[6] T. Holthe, L. Halvorsrud, D. Karterud, K.-A. Hoel, and A. Lund, "Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: a systematic literature review," Clin. Interv. Aging, vol. 13, pp. 863–886, May 2018.
[67] J. Zhang, T. Johnson, V. Patel, 2003, "Using usability heuristics to evaluate patient safety of medical devices," Elsevier.

[68] ISO, "ISO 9241-11: International standards for HCI and usability standards related to usability," 1998.

[69] ISO, "ISO/IEC 9126-1 Standard, Software Engineering – Product Quality – Part 1: Quality Model," 2001.

[70] Usability Body of Knowledge, "Checklist | Usability Body of Knowledge." [Online]. Available: http://usabilitybok.org/checklist. [Accessed: 02-Jun-2020]

[71] A. Oztekin, A. Nikov, 2009, "UWIS: An assessment methodology for usability of web-based information systems," Elsevier.

[72] L. Triacca, A. Inversini, and D. Bolchini, "Evaluating web usability with MiLE+," in Proceedings - Seventh IEEE International Symposium on Web Site Evolution, WSE 2005, 2005, vol. 2005, pp. 22–29.

[73] N. Mi, L. A. Cavuoto, K. Benson, T. Smith-Jackson, and M. A. Nassbaum, "A heuristic checklist for an accessible smartphone interface design," Univers. Access Inf. Soc., vol. 13, no. 4, pp. 351–365, Oct. 2014.

[74] Y. G. Ji, J. H. Park, and M. H. Yun, "A usability checklist for the usability evaluation of mobile phone user interface," Int. J. Hum. Comput. Interact., vol. 20, no. 3, pp. 207–231, 2006.

[75] L. C. Law and E. T. Hvannberg, "Complementarity and convergence of heuristic evaluation and usability test: A case study of UNIVERSAI brokerage platform," in ACM International Conference Proceeding Series, 2002, vol. 31, pp. 71–80.

[76] Z. Munn, M. D. J. Peters, C. Stern, C. Tufanaru, A. McArthur, and E. Aromataris, "Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach," BMC Med. Res. Methodol., vol. 18, no. 1, p. 143, Dec. 2018.

[77] H. Desurvire, J. Kondziela, and M. E. Atwood, "What is gained and lost when using methods other than empirical testing," in Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems - CHI '92, 1992, p. 125.