

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.DOI

A systematic literature review of research-derived touchscreen design guidelines for older adults

LEYSAN NURGALIEVA¹, JUAN JARA¹, MARCOS BAEZ^{1,2}, FABIO CASATI¹, AND MAURIZIO MARCHESI¹

¹Department of Information Engineering and Computer Science, University of Trento, Trento, TN, 38123 Italy

²Department of Economics, Tomsk Polytechnic University, Tomsk, 634050 Russian Federation

Corresponding author: Leysan Nurgalieva (e-mail: leysan.nurgalieva@unitn.it).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 690962 and by the "Collegamenti" project. This work was also supported by the project "Evaluation and enhancement of social, economic and emotional wellbeing of older adults" under the agreement No. 14.Z50.310029, Tomsk Polytechnic University.

ABSTRACT

The distinct abilities of older adults to interact with touchscreen devices has motivated a wide range of contributions in the form of design guidelines, which aim at informing the design for the aging population. However, despite the growing effort by the research community, many challenges still remain in translating these research findings into actionable design guidelines, with reports hinting scant adoption or implementation issues, which ultimately hurt the development of more accessible interactive systems. In this systematic literature review we look at the research-derived design guidelines that set the foundation for design guideline compilations and standards, analyzing the aforementioned issues from the perspective of experts trying to discover, classify, and evaluate the work on the area of touchscreen design guidelines for older adults. The review analyses 52 research articles resulting in 434 research-derived design guidelines for touchscreen applications. These guidelines are analyzed using a taxonomy that considered the older adults ability evolution and the design aspects that are target of the recommendations. The results point to the use of different definition of older adults, which go as early as 55+, with design of displays and interaction styles to accommodate to vision and dexterity declines as the most prominent areas of research. However, proposed guidelines and recommendations were validated in only 15% of articles analyzed. The analysis also revealed that identifying guidelines and characterizing their focus in terms of ability declines and design aspects addressed is a demanding activity and prone to error, given the quality of reporting and details offered in research articles.

INDEX TERMS Information and communication technology, Accessibility, Ageing, Older adults, Systematic literature review.

I. INTRODUCTION

OLDER adults are turning their attention to interconnected devices as attractive means to stay in contact with family, friends, and the world around them, bringing significant benefits, especially to those who are less able to interact physically with others [1]. In particular, they turn to mobile touchscreen technology [2] as it can be more intuitive, regardless of the user's age [3]. Furthermore, such interfaces allow for an almost "complete freedom of design and interface options" [4] as they are not limited by physical

buttons or similar hardware. More intuitive interfaces can then better support the use of such devices by older adults and, therefore, increase their access to digital products and e-services [5].

However, interacting with touchscreen devices and applications poses many challenges for older adults, including usability and accessibility issues. More than for any other age group, for older adults these challenges result in frustration and anxiety [1], [6]–[9].

The specific challenges of older adults in interacting with

devices has motivated a wide range of research contributions in the form of design methods and guidelines for making devices usable and accessible for this population [10]–[12], for instance, by providing voice-activated dialing for people with limited hand dexterity [13], text entry support for older adults with severe visual decline [14], or gestures for interacting with touch based interfaces tailored for age-related motor declines [15].

Nowadays, we observe an increasing research interest in the field of human-computer interaction for older adults with an ever-growing list of research-derived guidelines published every year. Despite the potential of this research to inform the design of interactive systems, experiences validating and applying existing compilations of guidelines [16]–[21] tell us that many challenges remain in using them successfully:

- Guidelines can be confusing, or become obsolete [16], [19], [22],
- Guidelines might contradict or appear to contradict each other [16], [23],
- They can be defined with concepts difficult to understand for designers [24]–[26] (also as design guidelines and checklists become more complex over time [27]),
- The relative importance of the guidelines, and consequently which ones to enforce, might not be clear [22], [28], [29],
- Older adults form a heterogeneous population group, something that might not always be recognized by practitioners and reflected in the research [30].

As a result, despite the ongoing effort by the community, the potential of research-derived guidelines is still untapped. Indeed, previous research has also confirmed that design guidelines benefit from being revised and/or expanded by the scholarly inquiry [27], and that while being “valuable and helpful”, they are still tied to the “conditions of the study (population included, devices configuration, executed tasks)” [3], which should be taken into consideration upon deciding on applying them.

In this paper we look at the research-derived design guidelines, analyzing the aforementioned issues from the perspective of experts trying to discover, classify, and evaluate the work in the area of research-based touchscreen design guidelines for older adults.

This is the first systematic review studying these issues at this scale. Specifically, we address the following research questions (RQs):

RQ1. What are the characteristics of the older adult population and interaction design addressed by current research-derived guidelines for touchscreen? With this research question we aim at: i) analyzing the different definitions used to describe the heterogeneous older adult population, ability declines and related design support, and also ii) gaining an overview of the specific populations and design aspects addressed by current state of the art.

RQ2. What is the quality of the methods and strategies used to generate and validate the design guidelines? We

aim at assessing the process followed by the researchers in deriving the design guidelines (reliability), and the methods used to validate them (validity). By considering the strength of the evidence, we also highlight areas where more experimental research is needed.

RQ3. Which issues emerge and what effort is required in identifying and cataloging research-derived guidelines in order to make them available to the average practitioner?

As the quality of any review of design guidelines study affect the chances of guidelines being discovered and correctly interpreted, we also report on our experience in extracting and characterizing the focus of the guidelines. We assess the quality of the reporting and available details to characterise the focus of the guidelines.

This work attempts to improve access to existing touchscreen guidelines by classifying them using a fine-grained user capability model and an appropriate design taxonomy. In doing so, we make current guidelines more useful for the practitioners, letting them have a better understanding of the importance of each guideline, how reliable it is, which of the guidelines they need to enforce according to the target population and to the technology that will be used to run the application. Finally, this work aims at contributing to the research body by presenting a mapping of guidelines according to the proposed capability model, highlighting the abilities that are well covered, pointing out the gaps where more research is needed, and calling attention to individual guidelines, identifying which ones of them are well supported and could be enforced, and which ones are ambiguous, confusing, or contradicting.

In what follows we analyze the related work in accessibility of touchscreen interfaces for older adults, describe our systematic literature review and proposed taxonomy to categorizing the final set of guidelines.

II. BACKGROUND

Researchers have adopted different perspectives on the definition of “design guidelines”. Smith and Mosier [31] refer to guidelines as an encapsulation of expert judgment whose use varies depending on the user. Dix and colleagues [32] define them as the “direction for design, in both general and more concrete terms, in order to enhance the interactive properties of the system”. Stewart and Travis [33] instead refer to them as “sets of recommendations from software providers or agreed within development organizations to increase consistency of design and to promote good practice within a design process of some kind”. Informed by these definitions, in this systematic review we consider design guidelines as following: concrete recommendations that can inform interaction designers in the development of interactive software systems. We specifically focus on design guidelines derived from peer-reviewed scientific articles.

A. PRIOR WORK ON COMPILATION OF GUIDELINES

The literature has emphasized the importance of design guidelines as precise and reliable recommendations to refer to while developing technologies for older adults.

An early attempt at compiling and validating them in user studies with older adults is the work of Apted et al. [34], who describe the use of design guidelines for such touchscreen devices (in this case a tabletop), and address general ageing related challenges such as “losses in vision, cognition and motor skills”, in using them.

More recent works on synthesis and evaluation of design guidelines for a wider range of touchscreen devices present them based on the usability problems older adults face, such as searching for information, issues with gestures, or element complexity [35]. Some researchers aim at reducing “the gap between a designer’s conceptual model and a user’s mental model of the design” [12] and attempt to make guidelines more applicable for the industry, for instance, providing a checklist of prescriptive design guidelines [10].

Other studies provide general summaries of the literature on design guidelines [21], [36], [37] but do not provide a systematic analysis, such as deriving them from a qualitative empirical analysis of system and user interface (UI) requirements developed for older adults [38].

In summary, previous work focuses on various aspects of older adults interacting with touchscreen technologies, either targeting usability aspects or ageing related issues. However, there is a lack of work that addresses the diversity of the older population, recognizing its heterogeneity, instead of defining older adults solely by age or common ability declines. In particular we have not found any work that systematically cover both the variety of ageing related ability declines and related design categories of interacting with touchscreen devices.

Hence, the main motivation for conducting a systematic literature review was to critically assess the current state of the art in the field of design recommendations for the older population that experiences ageing-related ability declines. Through the analysis of current trends and gaps in designing for the heterogeneous ageing population we aim to provide a snapshot of the current state of research in this area. In addition, we also review the methods for deriving and validating the guidelines.

B. PRIOR WORK ON IDENTIFYING ISSUES IN DESIGN GUIDELINES

Previous studies point out that design guidelines can be confusing, contradictory, and obsolete (due to the advances of technology), as it happened with Web Content Accessibility Guidelines (WCAG 1.0) [16]. In their website usability tests with disabled users, Romen et al. [16] empirically validated the usefulness of using WCAG as a heuristic for website accessibility and found that “the application of WCAG alone is not sufficient to guarantee website accessibility” but they rather should be applied in combination with other recommended lists of guidelines.

A number of studies conduct literature reviews to further evaluate current research based design guidelines. For instance, Zaphiris et al. notice that design guidelines can also be “too long, general and not too specific”, which makes them difficult to interpret and apply to a user interface by designers who might not even know “when and how they can be used” [19]. Previous studies also recognize that guidelines can be defined using concepts that are unclear to designers and do not always address their needs as recipients of this research-based guidance [24], [25]. Moreover, designers do not always realize the importance of guidelines, or if they do, they do not know which ones they should enforce [28], [29].

Some guidelines might lack a clear structure and sometimes contradict each other, “creating significant accessibility problems for designers”, as observed by Newell et al. [23]. For instance, Carmien et al. recommend the use of colours, icons, and graphics in displaying information and claiming that it should be prioritized over using text for older adults with vision declines [20], while Caprani et al. warn that displaying information by “grouping menus by colour alone can also lead to difficulties. Instead it would be preferable to use text, spacing or frames” [21].

The definition of target population addressed by the guidelines can also be misleading when described by age, since nowadays older adults form a diverse group with various levels of ICT skills and abilities. Vines et al. [30] published a critical analysis of 30 years of ageing research in the HCI community, where they discuss the prevailing homogeneity of the older population group in HCI research. They found that the homogeneity is either expressed through comparisons between older and younger users or “the ways older participants are discussed in the method, findings, and discussion sections of publications”, for instance, as “retirees” or “grandparents” but without specifying the socioeconomic and cultural contexts of participants.

Studies on the evaluation of guidelines also raise questions about their quality and whether they are “consumable” not only by designers but also experts, and emphasize that it is particularly important to organize and present them carefully in order to enhance their effective use [19]. Other studies on application of general guidelines, such as that of Kim [22], report on their usability problems when “designers have trouble in accessing and retrieving relevant guidelines, thereby not being used as an integral part of the design process” and recommend to organize them in a multidimensional structure that would include both design and user factors. In line with that, Petrovčič et al. [27] confirm that design guidelines benefit from being revised and/or expanded by the scholarly inquiry. Beside being “valuable and helpful”, guidelines should be contextualized to the “conditions of the study (population included, devices configuration, executed tasks)” [3], which should also be known to the experts and taken into consideration upon deciding on applying them.

A number of studies provide various taxonomies used to categorize design guidelines [3]. However, there is no standard or commonly recognized taxonomy that could facilitate

the access to current best research-derived practices and simplify their application to the specific research and industrial projects targeting heterogeneous ageing population.

In their literature review, Petrovčič et al. [27] investigated the categories that were included in the mobile design guidelines and checklists, the most mentioned ones being related to selected visual and haptic issues (e.g., high contrast, font size, button type, button size, button positioning). By contrast, the least frequent were categories related to the elements of textual interface and to screen and menu aspects. Their results suggest that despite the increased complexity in terms of dimensions and categories, there have been many usability dimensions of age-friendly mobile phone UI design which could be better covered and, thus, warrant further development in the future. Petrovčič et al. confirm that validation in the sense of repeatability and reproducibility seems to be one of the weakest aspects of current design guidelines and checklists. In their literature review, “only half of them were validated in the original papers and even fewer used different empirical methods as a basis for establishing evidence that the guidelines had fulfilled their intended requirements”.

III. CONCEPTUAL FRAMEWORK FOR GUIDELINE CATEGORIZATION

It has been recognized that software developers and designers targeting older populations do not always have access and/or understanding of research-based design guidelines for older adults [24]. Thus, a good starting point to understand the complexities of design guidelines is to analyze them from a perspective that can inform a design task: What practical considerations should we take when designing for older adults? This requires us to consider the characteristics of the population targeted by the design guidelines, and the aspect of the interaction design that merit the recommendations.

A. USER'S CAPABILITY MODEL

To characterize older adults, we consider the aspects that define an individual's ability to interact with a system in a user's capability model. In doing so, we aim at representing the diversity of the older adult population, and have a better understanding of the guidelines that should be enforced according to a specific target, a population with its own strengths and limitations.

We structure the user's capability model via a taxonomy that presents ability categories in a function-based user modeling approach (as older adults might not have same level of abilities at a given age), like the one proposed by [39] and selected as a basis also for this work. The user profile variables proposed in [39] are connected to specific user interaction abilities and constraints, and further subdivided into perceptual, cognitive, and motor classes. In addition, during the process of guidelines categorization (and contacting guidelines' authors for a confirmation of our categorization, which will be discussed later), we found that the “hand-eye coordination” subcategory being in the “cognitive” ability category caused a lot of confusion, so it was moved into

a separate “psychomotor” category following the taxonomy for the user's capability model defined by [21]. Considering those changes, our final user's capability model resulted into the following categories:

- 1) Perceptual abilities, including vision and hearing as primary output modalities in manipulating touchscreen devices;
- 2) Cognitive abilities, such as working memory, divided attention, and information processing speed, declines of which can significantly affect user's capacity to interact with technology;
- 3) Psychomotor abilities, such as slowness and imprecision in motor control and declines in hand-eye coordination, which may make touchscreen input problematic for older adults;
- 4) Motor abilities, caused by a decrease in muscle strength and dexterity and resulting into mechanical difficulties in navigating touch based applications and devices themselves.

Thereby, we adopt user's capability model combining the models presented by Caprani et al. [21] and Peissner et al. [39] with additional dimension of ability decline severity categories.

We defined three severity categories for ability decline as follows: “severe” for critical cases, such as color blindness for severe vision decline; “mild” for cases when decline could be corrected, such as minor memory problems; and “universal” specifically created for guidelines stated as fitting for both younger and older populations, e.g., providing a possibility to adjust the interaction depending on severity of specific case (“Provide a possibility to adjust the volume” [12]).

B. DESIGN TAXONOMY

We next present a design taxonomy to categorize design guidelines derived from the current literature review.

Various design taxonomies were considered, mainly coming from two sources: industry (design categories of guidelines proposed by Yahoo [1], Microsoft [2] or Android [3]), and from academia, such as the taxonomy defined by [40], which consists of categories such as trust or motivation, or [41] that includes categories such as “actions” and “objects”. However, we did not adopt the taxonomies mentioned above due to their specific focus on the interaction aspects strongly related to the devices. From another approach, traditionally, guidelines are classified in terms of abstract design principles [42], for instance, searching for information or using gestures [35]. This way of presenting guidelines may make it challenging to select the ones to apply to a specific design problem, which once again motivated us in finding a taxonomy that would provide concreteness and relevance towards covering specific design solutions.

¹<https://developer.yahoo.com/ypatterns/>

²<https://msdn.microsoft.com/library/windows/desktop/dn742479.aspx>

³<http://developer.android.com/design/index.html>

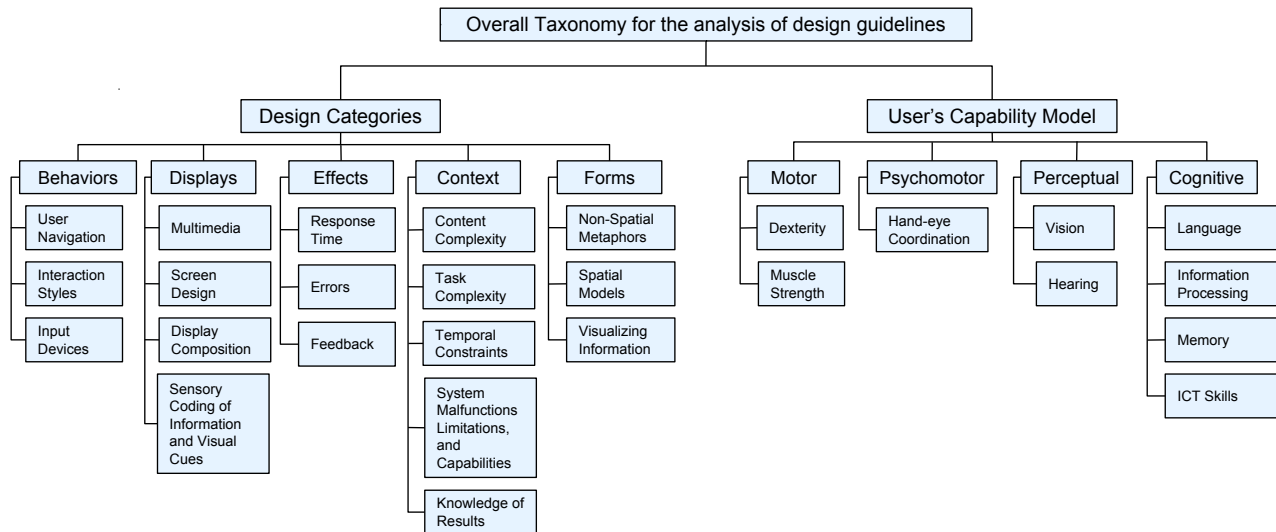


FIGURE 1: Taxonomy of the conceptual framework

The design taxonomy we chose as the most suitable reflects the view that the user interface is composed of seven fundamental components: Actions, Behaviors, Contexts, Displays, Effects, Forms, and Goals [43], which cover both design and interaction dimensions users face while using touchscreen technologies. Actions and Goals were eliminated due to not being relevant towards user interface design guidelines; *actions* refers to traditional hardware components of processing, such as computation, storage and retrieval of data, CPU, I/O, and peripherals; and *goals* represents the motivating behind the tasks performed through the interface. The final selected components are the following:

- 1) *Behaviors*, which refers to the user's interaction styles with the system, its navigation, and information input. For instance, this category includes guidelines on gestures used when using touchscreen devices or possibilities of multimodal data input;
- 2) *Context*, which refers to the settings in which the user behavior can occur and that have effect on the performance of users. Complexity of the system content and tasks related to the navigation are typical subcategories of "Context";
- 3) *Displays*, denoting the visualization of information for its own sake. Typical guidelines that belong to this category span topics such as multimedia used in the systems or composition of the content, and others related to the displaying information to users;
- 4) *Effects*, denoting feedback about the system actions as a response to the user interactions. For example, this category includes guidelines about error messages displayed to older adults;
- 5) *Forms*, that refers to models or metaphors in which actions, effects and displays are embedded, for example, in relying on familiar notions to older adults when developing touchscreen applications.

The overall taxonomy proposed and used in this work is presented in Figure 1.

C. CAPABILITY MODEL EVALUATION BY HEALTHCARE PROFESSIONALS

In order to evaluate the capability model and the distribution of the guidelines along the ability decline categories from the perspective of healthcare professionals specialized on ageing, we have conducted a focus group discussion (FGD) followed by a further expert evaluation by two geriatric medical professionals.

A focus-group discussion (FGD) was conducted in November 2016 in English with 5 geriatric medical professionals: a physiotherapist, three geriatricians and a nurse. The FGD lasted for one hour and included the following topics: categories of the most common ageing related ability declines in their practise, the issues of older patients interacting with technology, their recommendations to touchscreen applications adapted to ageing population. The focus group was given an introduction, explaining the general procedure and the importance of applying design guidelines in designing touchscreen applications for individuals experiencing ageing related ability declines. A description of the capability model and the overall categorization of the guidelines was also described.

A detailed summary of the qualitative results of the session was composed from the audio recordings by the corresponding researcher. Transcripts of the focus group were later analyzed, discussed and the findings cross-validated with a geriatrician and a geriatric nurse in the form of semi-structured interviews that lasted for 30 minutes and were also further analyzed and discussed by the first three authors of this paper. Overall, healthcare experts provided a positive evaluation of the applied capability model in a given HCI research context. They have also commented on the specific

ability decline categories that occur more often in their work practice, which we explore in more details in the discussion section.

IV. METHODS

This study has been undertaken as a systematic literature review following the guidelines described in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [44].

A. IDENTIFICATION PHASE

The first phase of our systematic literature review consisted of identifying the articles that should be reviewed. We limited our search to the databases that contain papers from conference proceedings and journals that we consider the most relevant to the areas of ageing and HCI (see Table 1 for the complete list). Additionally, we performed a wider search on the SCOPUS database, as to cover relevant work from other sources not present in the initial list.

For each database, we searched the titles and abstracts of every article for keywords related to the following topics: "older adults" (older adults, elderly, elders, ageing, aging, senior), "design guidelines" (design, guidelines, recommendations, suggestions, principles), and "touchscreen devices" (tablet, touch based, touch devices).

The search covered research articles written in English and published between January 2005 and November 2017.

B. SCREENING PHASE

In the screening phase, we evaluated, to a deeper level, which of the identified articles could contain useful content for the systematic literature review. During this phase, three researchers independently screened the content of each article and tagged it with either *Yes*, *No*, or *Maybe*, where *No* indicated that the article did not contain any useful information, and *Yes* and *Maybe* indicated that:

- The article could contain design guidelines⁴ applicable to touch devices, or
- The article could address interaction design issues targeted to older adults⁵ or
- The article includes older adults as participants in their study group.

Articles tagged with *Maybe* and disagreements were resolved in face-to-face discussions between all three researchers for the final list of included papers. Papers deriving recommendations for specific applications (and not generalizable) or that gave guidelines exclusively for hardware design were discarded.

C. ELIGIBILITY PHASE

At the eligibility phase, each article was evaluated in detail. For each article we extracted the proposed guidelines (if

⁴Researchers considered the definition of design guidelines that is explained in Section 3

⁵In this case older adults may defined either by age, for example, 65+ or 60+, or with ability declines related to aging

available) and the details of the studies that either conducted to and/or validated those guidelines. During this phase, articles were also removed from the systematic literature review if the proposed design guidelines were considered as design principles, hence, too general for this work, or confusing for the experts (coders) to interpret.

In order to keep records organized during the classification and filtering process, we applied several data management tools. As a tool for an easier collaboration and collective work, online Google spreadsheets were used to store the records obtained from reviewed articles as well as to discuss the data and make annotations. More specifically, during this phase two forms were used:

- A form to store the extracted design guidelines and the name of the article from where they came from.
- A form to store information related specifically to each selected article. Coding parameters were the following: date of publication, authors, short summary, type of ability decline and its screening methods if any, type of target touchscreen device, pre-studies (that guided the creation/definition of the proposed guidelines) and post-studies (that either applied the guidelines, or validated them) including data about subjects (size, age, percentage of females), format of user study (group or individual), mode of assessment (technical or non-technical), and presentation of the final design guidelines.

Three researchers, independently, conducted a full-text analysis of each article and extracted design guidelines from each of the selected articles and the information related to them into the 2 standardized data coding forms described above. This process resulted into a set of preliminary design guidelines that support the design of solutions that target declines of abilities of older adults.

D. INCLUSION PHASE

In this phase, we performed a qualitative analysis on the extracted information to better categorize the design guidelines and to prepare the coded data for replying to the research questions that guided our systematic literature review.

To perform the qualitative analysis we used the data collected in the two forms described earlier and the taxonomies for design categories and user's capability model shown in Figure 1.

Specifically, we address our research questions introduced in Section 1 as follows:

- For RQ1. *What are the characteristics of the target population and touchscreen interaction addressed by current research-derived guidelines?* We coded the target population either as people affected by ageing related ability declines, or as people that reached a certain age.
- For RQ2. *What is the quality of the methods and strategies used to generate and validate the design guidelines?* We extracted measurable information, i.e. metrics related to the studies conducted (materials and formal methods used, screening methods, and number

TABLE 1: Sources of selected papers

Source	Type	Database
Universal Access in the Information Society	Journal	Springer
Gerontechnology	Journal	Gerontechnology
Computers Helping People with Special Needs	Conference	Springer
Australian Conference on Human-Computer Interaction (HCI) OZCHI	Conference	The ACM Digital Library
Human-Computer Interaction INTERACT	Conference	Springer
Behaviour & Information Technology	Journal	dblp
Computer Human Interaction (CHI)	Conference	dblp
Mobile HCI	Conference	dblp
International ACM Conference on Assistive Technologies (Assets)	Conference	dblp
International Conference on Advances in Computer-Human Interaction (ACHI)	Conference	dblp
International Journal of Human Computer Interaction (IJHCI)	Journal	dblp
ACM Transactions on Computer-Human Interaction (TOCHI)	Journal	dblp
BCS conference on Human Computer Interaction	Conference	The ACM Digital Library
ACM SIGCHI Symposium on Engineering Interactive Computing Systems	Symposium	The ACM Digital Library
Human-Computer Interaction	Journal	dblp
Computers in Human Behavior	Journal	dblp
Universal Access in HCI	Conference	Springer
Computers Helping People with Special Needs (ICHP)	Conference	Springer
Assistive Technologies	Journal	The ACM Digital Library
HUMAN FACTORS	Journal	dblp
Procedia Computer Science	Journal	Elsevier
Educational Gerontology	Journal	Taylor & Francis Online
Interacting with Computers	Journal	dblp
Ergonomics in Design	SAGE	dblp
International Conference On Neural Information Processing	Conference	Springer
International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info exclusion	Conference	dblp

of participants) and formality of the methodology and number of sources in case of literature reviews.

- For RQ3. *What issues emerge and what effort is required in identifying and cataloging research-derived guidelines, as to make them available to the average practitioner?* We tagged the guidelines with keywords that represent the challenges that a reader has to surpass to extract and understand those guidelines.

The overall process of this phase was performed in three iterations.

- In the first iteration, guidelines were classified and grouped according to the ability type that was explicitly addressed by the original paper.
- In the second iteration, guidelines in each ability group were analyzed and re-evaluated based on the description text of the guidelines themselves in order to confirm if they belonged to the ability group identified in the first iteration or to another group. This was necessary as there were several articles that targeted several ability types. If a guideline was found to better fit another ability group, it was moved to that ability group. This ability group change had to be confirmed by a majority vote of internal expert agreement. Moreover, single guidelines targeting ability declines not affecting older adults were excluded in this iteration.
- In the third iteration we added the selected design taxonomy, and each individual guideline was classified as belonging to one of the design categories and subcategories described in Figure 1. Guidelines themselves

were evaluated based on the definition of “design guideline” as presented in Section 3. For example, the following text “Provide a way to exit on every screen” [20] matches our definition of guideline and was included, while the following text “Consider task complexity in navigation tasks” [45] did not conform with our guideline definition (was found to be more a design principle), and thus, was excluded. Finally, the severity category of each ability decline was added, classifying guidelines as either “severe”, “mild”, or “universal”. “Universal” decline referred to guidelines stated as fitting for both younger and older population. This iteration was also intended to identify and remove repeated guidelines from the list. By the end of the third iteration, the level of inter-rater agreement was 55% for classification of ability declines and 59% for design categories.

After this phase, we took the final list of guidelines and rephrased them into a heuristic checklist for designing accessible solutions for older adults that could be generalized and applied to different touchscreen technologies, and be easily comprehended and adopted by software developers and designers.

During each iteration of the classification and filtering process, structured questions were used to perform a guideline quality control, such as “Do you agree this guideline belongs to this ability/design category?” or “Do you agree with rephrasing the guideline text?”. Each expert had three options for the answer: “Yes” for confirming the guideline text/categorization, “No”, to indicate that the guideline

needed a review, and “Not sure” to wait on the decision of the other experts.

Disagreements were addressed by asking the third expert to classify the guidelines in question, and the final decision was reached by consensus in face-to-face discussions. These discussions also provided insights into the challenges of interpreting and understanding the guidelines as currently reported. We discuss these challenges and issues in the Results section.

As described above, analysis of coders agreement outcomes was both quantitative and qualitative. Quantitatively, three answer options were considered and calculated for the two taxonomies: design categories and user’s capability model. Qualitatively, each guideline text that had to be rephrased was discussed by at least 2 researchers to avoid the loss of the original meaning.

E. EVALUATION OF THE QUALITY OF THE PROCESSES RELATED TO THE EXTRACTED GUIDELINES

We evaluated the quality of guidelines using 2 dimensions.

- The first dimension is associated to the quality of the methodologies used to define/propose design guidelines and we refer to it as the *reliability of a design guideline*. We think this assessment is important as it can be a good indicator of how likely is a guideline to support its target population. For instance, we investigated if the guidelines were provided after conducting user studies with participants from the targeted population or by a related literature synthesis, and then tried to evaluate the quality of those methods.
- The second dimension is associated to the quality of the methodologies used to validate or test design guidelines and we refer to it as the *validity of a design guideline*. We consider to perform this assessment as it can be a good indicator of how likely is a guideline to improve the usability of a solution for its target population. For instance, we looked for the presence of studies and experiments conducted to test/validate design guidelines, e.g., check whether a guideline improves the usability of touchscreen user interfaces for older adults with a specific ability decline.

Our evaluation method was performed in two steps and, in what follows, we describe them in detail. In the first step, for each article, we classified the methodology used for deriving/proposing guidelines (to compute the *reliability*) and the methodology used to validate them (to compute the *validity*) in one of the following categories:

- *User studies* that propose (or validate) guidelines based on the results of experiments where participants interacted with user interfaces and/or prototypes;
- *Literature reviews* that aggregate design guidelines from other articles;
- *Expert evaluations* that either assess specific types of accessible applications or that describe their development process.

In the second step, we evaluated the methodologies according to their category.

For *user studies*, we considered the number of participants in the study and if the participants had to have a disability; we did collect other information related to the user studies but we did not use it for our evaluation. We chose this assessment method by following the recommendations presented in [46] where authors analyzed several research methods and suggested that the quality of a user study can already be assessed with these two variables. More specifically, authors in [46] suggest that a research study with a general population of users should have a minimum of 20 to 30 participants to be considered valid. When doing a research study that focuses on a population with disabilities (recognizing the difficulty of the recruitment for these cases) authors from the same work say that it is acceptable to have just from 5 to 10 participants.

Based on these suggestions we evaluate user studies as follows:

- User studies with participants without disabilities were scored as follows: if the number of participants was fewer than 20, the score of the study was **Low**; if the number of participants was between 20 and 30, then the score was **Good**; and if the number of participants was greater than 30, then the score was **Optimal**.
- User studies with participants with disabilities were scored as follows: if the number of participants was fewer than 5, the score of the study was **Low**; if the number of participants was between 5 and 10, then the score was **Good**; and if the number of participants was greater than 10, then the score was **Optimal**.

For *literature reviews* and *expert evaluations* we could not find unfortunately a strong method to evaluate their quality. Nevertheless, to give an assessment, we decided to look at the formality of the procedure, i.e., if the methodology followed a systematic procedure. A systematic methodology received the score **Optimal**, otherwise the methodology received the score **Low** (there was no **Good** score for these methodologies as our scoring variable was binary).

V. RESULTS

In this section we present the results of our literature review in relation to our initial research questions.

A. STUDY SELECTION AND GUIDELINE EXTRACTION

The primary search, or *Identification phase*, selected 582 papers from a set of a little more than 10K articles.

During the *Screening phase*, we evaluated the 582 articles and identified 158 articles that seemed to contain guidelines or content that could be interpreted or translated into guidelines. Later, 10 more articles were included based on the communication with authors and follow-up studies. From the excluded articles, 31 were editorial articles or duplicate works, and 393 either did not contain any guidelines or contained guidelines that were application specific or did not target touchscreen devices.

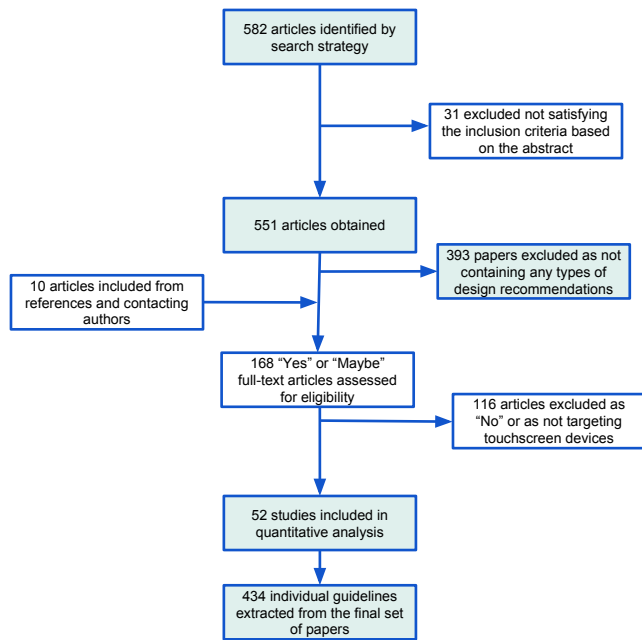


FIGURE 2: PRISMA flow diagram for this study

In the *Eligibility phase*, we performed a full-text analysis of the 168 articles with the purpose of extracting from them their corresponding guidelines. This analysis excluded 116 more articles due to: not proposing actual guidelines, or the guidelines were too general or confusing and, hence, matching more the definition of design principles rather than guidelines. This phase resulted in 52 articles marked as containing relevant guidelines for designing touchscreen applications for older adults.

Finally during *Data extraction* (“Included” in PRISMA chart), we further reviewed the final set of 52 included articles in order to extract the contained guidelines and the details of the studies that either produced the design guidelines and/or validated them. The final outcome of the literature review is **a set of 434 guidelines for designing touch-based applications for older adults**⁶ where some papers had a large number of guidelines (up to 143) and others only a few or even just one of them.

The process we followed in our study for filtering relevant papers can be seen in the PRISMA diagram in Figure 2.

B. CHARACTERIZING THE TARGET POPULATION AND INTERACTION DESIGN ASPECTS

We thus address the first part of our first research question: *What are the characteristics of the target population and touchscreen interaction addressed by current research-derived guidelines?*

1) Target population characteristics

By analyzing the articles directly related to older adults, we discovered that the target population is identified using one

of the following characteristics:

- **Chronological age** (52%): the target population belongs to an age range or is above a certain age. For this we took either the explicit definitions in the papers, or the age of participants in the reported studies.
- **Functional** (44%): the target population is affected by one or more ageing related ability declines.
- **Hybrid** (4%): the target population is defined by both of the previously defined characteristics, i.e., it belongs to an age group and is affected by one or more ability declines.

Figure 3 shows all the articles that define older adults chronologically, in ascending order.

Defining older adults by age or *chronologically*, researchers traditionally refer to official definitions such as by the World Health Organization (65 years and older) [47] or the United Nations [48] (60 years and older). In this review, we found that the *chronological* definition starts as early as age 55 [35], while just three articles [20], [49], [50] specifically defined their target population as individuals aged over 65 years and five [4], [51]–[54] – as 60+ by the United Nations.

Just one among all selected works explicitly focused on the oldest old (80+) age group [49], while other articles included the 80+ population in their studies but did not distinctly focus their research on them. We also found one particular article that addresses older workers, people in the transition age from work life to retirement [38], which in this case was considered to be the age range 55–75.

The next definition category in the table collects the articles that define older adults using *functional* characteristics, focusing on ability declines related to ageing. In this group we identified articles that tackle specific diseases related to ageing like Alzheimer’s [25] and Parkinson’s disease [55]; furthermore, there are articles that target a specific health issues that affect older adults but can also affect people of all ages, for instance, aphasia [56]. Our reviews also identified articles that address more general health problems caused by ageing, e.g., motor impairments [13], [14], [17], [57], [58], cognitive declines [59]–[61], and vision loss [50], [62]. Ten of the reviewed articles cover various ability declines within one study or set of guidelines [10], [12], [21], [63]–[69].

Finally, two articles apply a *hybrid* approach using both age and functional characteristics. Kobayashi et al. [70] recruit participants in their 60s and 70s with vision and hearing problems, and Wacharamanotham et al. [54] target older adults affected by hand tremor.

2) Targeted ability declines

In this subsection we further focus into characterizing the specific abilities that are addressed by the individual design guidelines. Their classification based on the capability model are collected in Table 2.

From Table 2 we can clearly see the asymmetry in the coverage of ability declines by extracted guidelines. The

⁶Guidelines available at the <https://design-review.mateine.org>

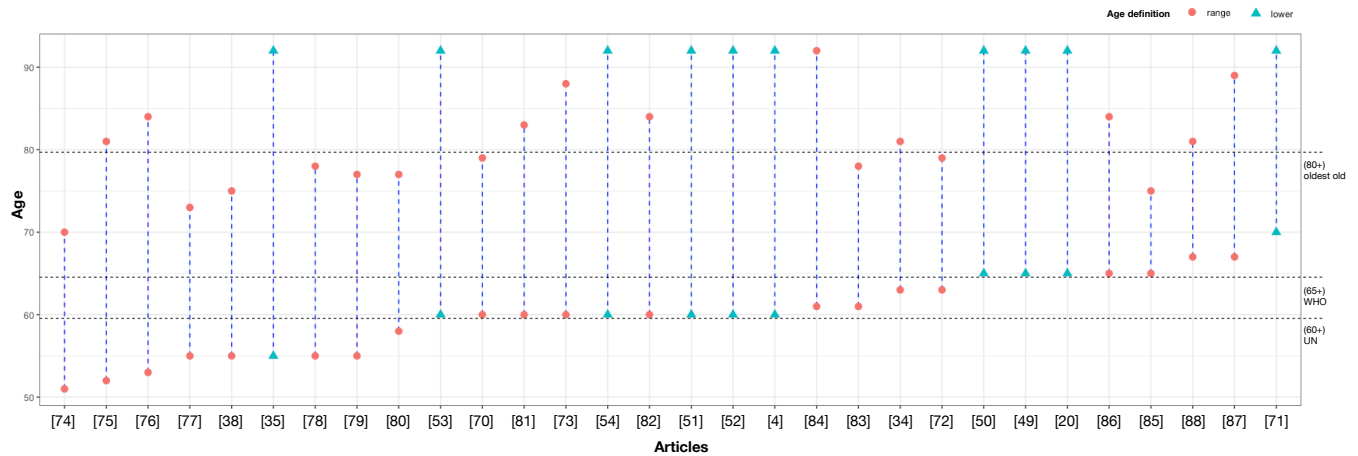


FIGURE 3: Chronological age distribution in included papers. Arrows represent unbounded age and lines represent age ranges.

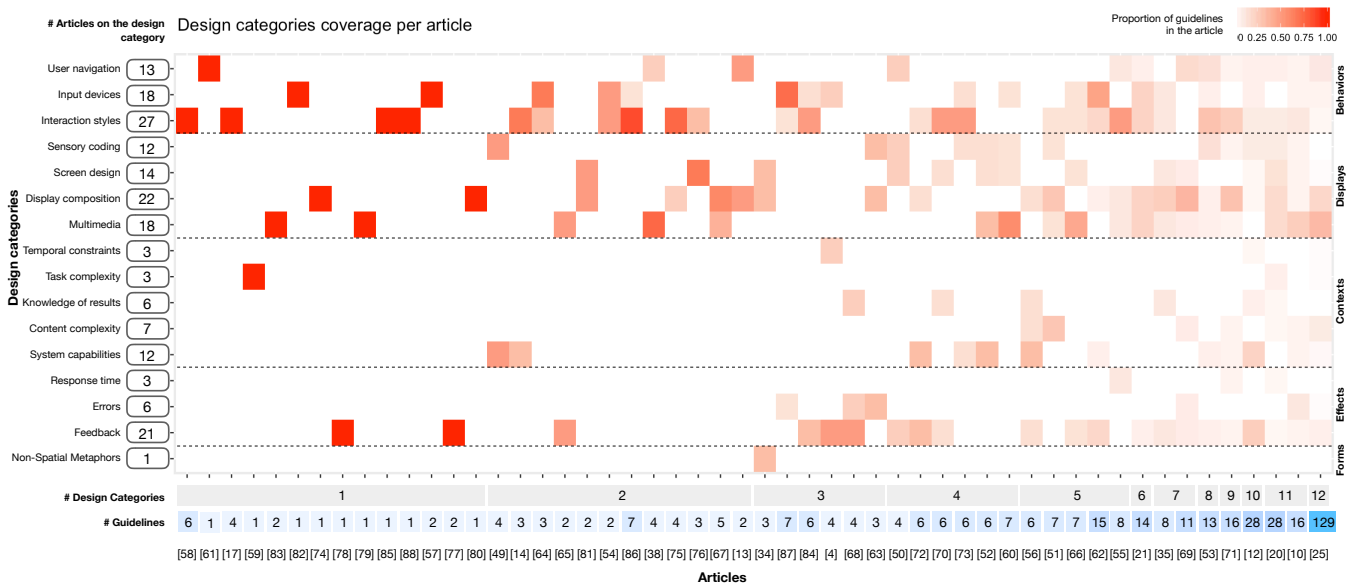


FIGURE 4: Distribution of design guidelines based on design categories and article

ability declines that are well covered correspond to **cognitive** (189 out of 434 or 43.6%) and **perceptual** (142 out of 434 or 32.8%), while **psychomotor** (58 out of 434 or 13.4%) and **motor** (45 out of 434 or 10.4%) are considerably less covered by design guidelines.

Drilling down into the subcategories of each ability, we see that for:

- **Cognitive**, the guideline coverage is relatively equally distributed among all subcategories including declines in language and information processing, reduced memory, and low ICT skills. Still, predominating ability declines are reduced information processing (68 out of 189 or 35.9%) and low or lack of ICT skills (58 out of 189 or 27.5%).
- **Perceptual**, the guideline coverage is dominated by those that support people with vision problems, e.g., blindness, color-blindness, low vision, etc. (126 out of

142 or 88.7%).

- **Psychomotor**, all the guidelines in this category (there are no subcategories) aim at helping people with hand-eye coordination problems.
- **Motor** ability decline category, which included dexterity problems and muscle strength, the guideline coverage is dominated by the ones that help people with reduced dexterity, i.e. problems in moving either their fingers and hands or arms (41 out of 45 or 91.1%).

3) Targeted design categories

To characterize guidelines by their relation to the design aspects of interactive systems, we classified the final set of guidelines in different design categories defined in the proposed taxonomy shown in Figure 1.

From Table 3, the design categories that are more affected by guidelines are **Displays** (182 out of 434 or 41.9%) and

TABLE 2: Distribution of design guidelines based on the user's capability model (N=434)

Ability	Subcategory	Example	# Guidelines	% Guidelines
Cognitive	ICT skills	"Outline the main features of the system" [71], "Ensure that error messages feedback provide mechanisms for resolving the error" [25].	52	12%
	Information Processing	"Display main information on the center of the screen" [25], "Use different colors to categorize information visually" [25].	68	15.7%
	Memory	"Always provide a 'home' button, and let users know 'where' they are" [20], "Provide reusable commands and gestures to ensure consistent interactions across applications and functions" [12].	31	7.2%
	Language	"Ensure that feedback messages is not in commando-style" [25], "Use video to facilitate understanding text" [25].	38	8.8%
Perceptual	Vision	"Use capital letters for highlight important text." [20], "Highlight the pressed button, if possible, showing a fading 'fingerprint' on the display to show the actual touch" [72].	126	29.1%
	Hearing	"Make a pause of some seconds after each spoken sentence" [25], "Provide a possibility to adjust the volume" [12].	16	3.7%
Psychomotor	Hand-eye coordination	"Increment the size of the zone round an hyperlink" [25], "Provide a cursor to show the selected information" [25].	58	13.4%
Motor	Dexterity	"Provide an adjustable delay of button response to ensure that multiple touches can be treated as one touch" [12], "Implement the ability for the interface to allow whole-handed and multifinger input. This is particularly important for individuals with hand tremors or arthritis" [73].	41	9.5%
	Muscle strength	"Address physical factors, such as the weight of the device and dexterity issues of this population, and account for limitations in the mobility of the appliance, e.g. users may not be able to use the appliance while walking" [49].	3	0.7%
	Speech	"Use acoustic models dedicated to seniors for the speech recognizer" [25].	1	0.3%

Behaviors (140 out of 434 or 32.2%), which had to do with how information and content are shown and how to interact with the system.

Following, we have **Contexts** (58 out of 434 or 13.3%), which is related to how the system shows, or reacts to, the current status of a task; **Effects** (53 out of 434 or 12.2%) that is related to the different types of feedback that a system or task should give; and finally, **Forms** (just 1 out of 434 or 0.002%) that has to do with associating tasks with known metaphors to make the tasks easier to learn and understand.

If we drill down into the subcategories of each design category, we can further say that for:

- **Behaviors:** half of the guidelines are related to the different interaction styles that users could have with the system (71 out of 140 or 50.7%), while the rest of the guidelines are divided among how users navigate through a task or system (30 out of 140 or 21.4%) and how they can input information into the system (39 out of 140 or 27.8%). Typical examples of guidelines in those categories are "provide a way to exit on every screen" [20] (user navigation) and "allow recognition of multiple voice commands at a time" [62] (input devices).
- **Displays:** Most of the guidelines are related to managing the multimedia content (81 out of 182 or 44.5%), for example, "avoid justified aligned text" [25]; and display composition, i.e. arranging the information on the screen (64 out of 182 or 35.1%): "use different colors to categorize information visually" [25].
- **Contexts:** The majority of guidelines have to do with

managing the different accessibility options and peripherals as well as the errors associated to them, i.e. "System Malfunctions, Limitations, and Capabilities" (24 out of 58 or 41.3%), for example, "display a help panel and tips about the features in the first user access" [71]. The second largest subcategory here is the complexity of the information communicated to the users (20 out of 58 or 34.4%), and an example of a related guideline is "avoid abbreviations and symbols" [25].

- **Effects:** Most of the guidelines refer to the different ways of giving interaction-based feedback to the users (43 out of 53 or 81.1%): "provide a confirmation of every completed function" [12].
- **Forms:** Just one guideline for the subcategory *Non-Spatial Metaphors*, which is the following: "Rely on familiar aspects of manipulating physical photographs. This reduces the amount to learn and remembering is easier since the user already knows how to move and share physical photographs" [34].

The heatmap in Figure 4 shows the distribution of the guidelines reported in each article, by design category. In the figure we can see how most articles propose guidelines in the areas of Interaction styles, Display composition, and Feedback. On the X axis we can also see the number of categories addressed by article, with half of the articles (26 out of 52) focusing on one to two design categories simultaneously.

4) Archetypes: Cross-dimensional analysis

Analyzing the distribution of guidelines in both design and ability categories (Figure 5), we can see some archetypes

TABLE 3: Distribution of design guidelines based on design categories (N=434)

Category	Subcategory	Example	# Guidelines	% Guidelines
Behaviours	User navigation	"Show the actual location all the time" [25], "Provide a way to exit on every screen" [20].	30	6.9%
	Interaction styles	"Avoid instant features that change with each new interaction such as filters and auto-completion" [71], "When the touch is lost during dragging, the object should stay where it has been left" [53].	71	16.4%
	Input devices	"Allow wait time setting on entering each character" [62], "Support system status determination with voice activation" [12].	39	9%
Displays	Display composition	"Use capital letters for highlight important text" [20], "Avoid transparent menus" [25].	64	14.8%
	Multimedia	"Support different types of contrast" [52], "Enable users to control the volume themselves" [25].	81	18.7%
	Screen design	"Provide shallow menus. Spread functionality across menu bar and pages" [20], "Some elderly users are less likely to notice changes of the modes and can become confused. Avoid multimode interfaces for applications as much as possible" [70].	19	4.4%
	Sensory coding of information and visual cues	"Make names of items on the screen heard as they are touched" [12], "Use universal icons along with redundant cues (e.g., color, text, and symbols)" [10].	18	4.2%
Contexts	Content complexity	"Use active voice rather than passive voice" [25], "Guide the user by means of messages in clear, objective and educational language" [71].	20	4.7%
	Knowledge of results	"Let the users know that they have successfully completed an action in the app" [68], "Provide a clear indication of progress and performance" [56].	7	1.7%
	System malfunctions, limitations and capabilities	"Make the touch screen startable in any position on the screen" [12], "Provide a training mode that teaches the user about each available gesture" [72].	24	5.6%
	Task complexity	"Allow tasks to be accomplished serially, don't force them to be done at the same time requiring cognitive switching" [20], "Provide the fewest possible of choices to users" [25].	4	1%
	Temporal constraints	"Make it possible to hold the key up to 2 s before the action will repeat" [12], "Provide users enough time to use or read content" [25].	3	0.7%
Effects	Errors	"Ensure that error messages feedback provide mechanisms for resolving the error" [25], "Support users to easily reverse their actions if they make a mistake in using the application" [68].	7	1.7%
	Feedback	"Make buttons highlighted when pressed to support correct selection" [12], "Provide a confirmation of every completed function" [12].	43	10%
	Response time	"Avoid time dependent controls" [55], "Delay and feedback when pressing a button. The touch screen buttons react too fast" [20].	3	0.7%
Forms	Non-Spatial metaphors	"Rely on familiar aspects of manipulating physical photographs. This reduces the amount to learn and remembering is easier since the user already knows how to move and share physical photographs" [34].	1	0.3%

emerging in the following areas:

- **Design of multimedia content to address vision declines** (Vision - Multimedia), with 49 guidelines that are related to how to show content so it can be comfortably and properly seen in spite of vision declines. A typical example of such intersection is to "implement 60% opacity for all highlighting to render a good contrast between the black text and background colour" [65];
- **Interaction styles to address coordination and dexterity declines** (Hand-eye coordination, dexterity - Interaction styles), with 26 and 22 guidelines that are related to how to make touchscreen interactions, such as gestures, accessible to older adults with hand-eye coordination and motor function declines. For instance, to help users to target the right spot, one should "make selections using gliding gestures for direct manipulation" [12];
- **Organizing the display composition to address information processing and vision declines** (Information processing, vision - Display composition), with 21 guidelines each that are related to how to arrange content so that it reduces the cognitive load on the users, and avoid creating unnecessary distractions for those with vision declines. For example, Ghorbel et al. recommend to "present only one message in a single interface" and "avoid using tones of similar lightness near to each other" [25];
- **Adapting content complexity to declines in language processing** (Language - Content complexity), with 16 guidelines that are related to the way information is presented and organized on a screen and how the language used on the screen matches the user's ability to comprehend it. Guidelines like "Guide the user by means of messages in clear, objective and educational language" [71] and "Express only one idea per paragraph" [25] are

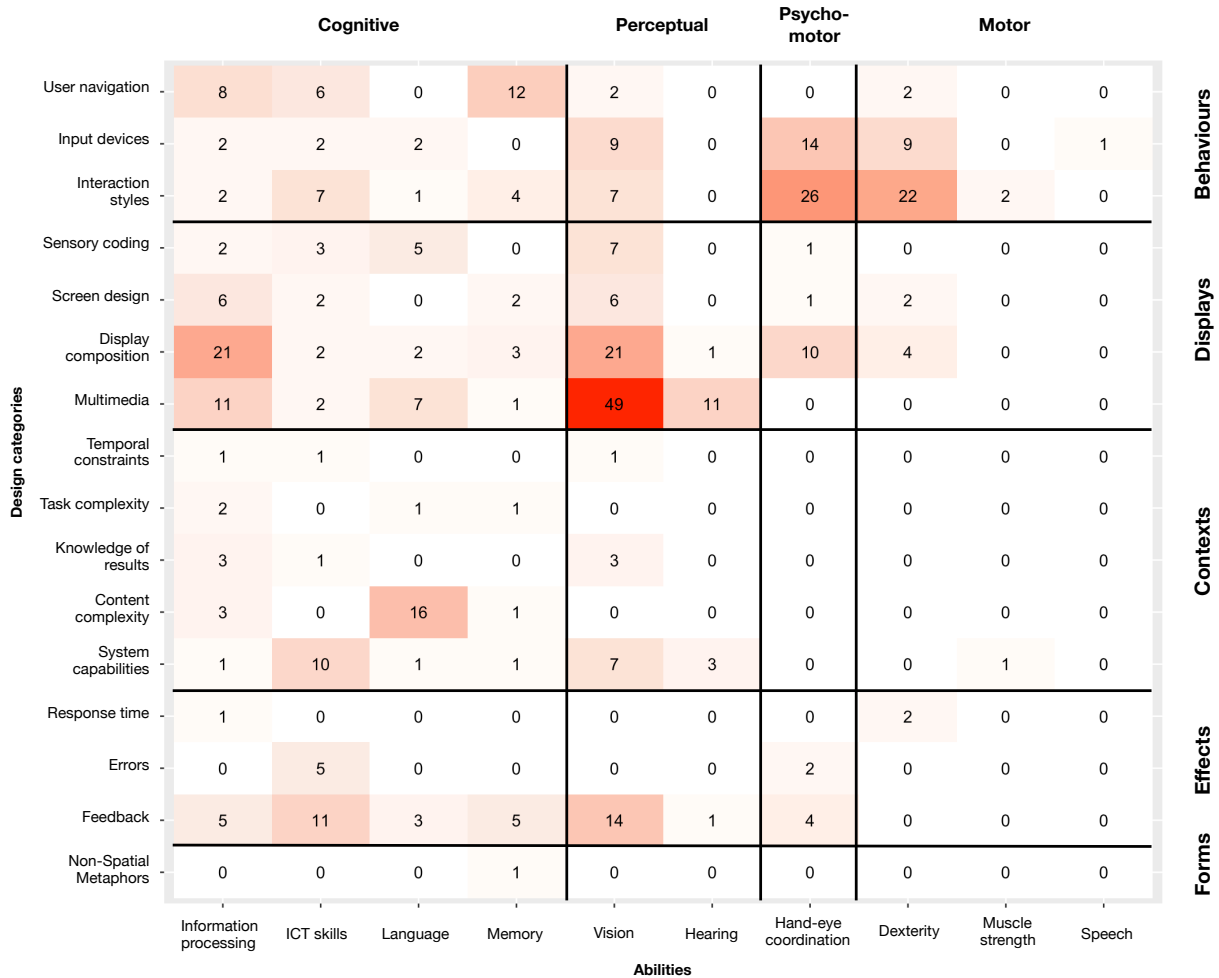


FIGURE 5: The distribution of guidelines by the design and ability categories

examples of such an intersection;

- **Design of feedback to accommodate to declines in vision, and lower ICT skills** (Vision, ICT skills - Feedback), with 14 and 11 guidelines that are related to the user's ability to perceive what is displayed on the screen, and skills and experiences in using current ICT user interfaces, which are supported by the feedback about the operations of the application in response to user behaviors. For instance, this could be addressed by using "audio confirmation to help elderly with reduced vision" [53] or by providing users with a "positive feedback icon" as "a motivator [...] and an important indicator that they are making progress" [4].
- **Designing user navigation to address memory and information processing declines** (Memory, Information processing - User navigation), with 12 and 8 guidelines that give us an insight of how a reduced ability to remember long execution tasks, and to process (complex) information in a timely manner, could be addressed by an improved navigation system. In this vein, Nunes et al. recommend to "provide clear information of current location at all times" [55], while Al-Razgan

et al. state that "the most important features should be available directly via a labeled button and not via menu navigation" [53];

- **Handling errors to accommodate lower ICT skills and coordination declines** (ICT skills, hand-eye coordination - Errors), with 5 and 2 guidelines that are related to the importance of efficient error processing in interaction with the UI and supporting users who are less experienced in using the touchscreen technologies or have hand-eye coordination difficulties, which could result into errors. This could be implemented, for instance, by ensuring "that error messages feedback provide mechanisms for resolving the error" [25] or by setting large space between commonly used buttons. "To minimize hazards and unintended actions, "Yes" and "No" touch buttons [should be] located at the farther left and right sides of the touchscreen, and other touch buttons [should be] placed on the main control panel" [10].

All the other areas present less than 20 guidelines, with several having none at all (see the related heatmap in Figure 5).

C. RELIABILITY AND VALIDITY OF CURRENT GUIDELINES

In this subsection we address our second research question: *What is the quality of the methods and strategies used to generate and validate the design guidelines?* As detailed in Section IV-E, to estimate the **reliability**, we evaluate the studies carried out in each of the included papers used to derive/propose design guidelines. For the **validity** we evaluate the studies used to validate design guidelines they proposed.

1) Reliability of design guidelines

After analyzing the studies from each of the included articles by using the methodology presented in Section IV-E, we computed the following reliability scores:

- **Optimal**: we found that 31% of the included studies are ranked in this level, from which 3 are literature reviews, 1 is an expert evaluation, and 12 are user studies (7 of them recruited participants with disabilities).
- **Good**: we found that 23% of the included studies are ranked in this level. All of them are user studies from which 3 required participants with disabilities.
- **Low**: we found that 46% of the included studies are ranked in this level, from which 10 are literature reviews, 4 are expert evaluations, and 10 are user studies (none of them require participants with disabilities).

Table 4 summarizes the types of studies, adopted by the research articles analysed, for producing the design guidelines.

TABLE 4: Categories of studies producing design guidelines

Type of study	Subcategory	References
Literature reviews	Non-systematic reviews	[10], [20], [21], [25], [34], [35], [53], [57], [64], [69]
	Systematic literature reviews	[51], [60], [62]
User studies	Ability-based user studies	[12]–[14], [17], [50], [54], [55], [58], [59], [61]
	User studies with older adults (age based)	[4], [38], [49], [52], [65], [70]–[88]
Expert evaluations	Expert agreement	[56], [63], [67], [68]
	System description	[66]

The reliability score of each article is passed down automatically to the guidelines extracted from that article. Figure 6 (left) shows the distribution of our reliability score for all the extracted guidelines. More than half of the articles relied on good or optimal methods, although this translated into less than half of good and optimal guidelines.

Figure 7 shows the distribution of the reliability score according to our capability model. We can see that only in the psychomotor category more than half of the guidelines received a good reliability score or better, for all other categories, more than half of the guidelines received a low reliability score.

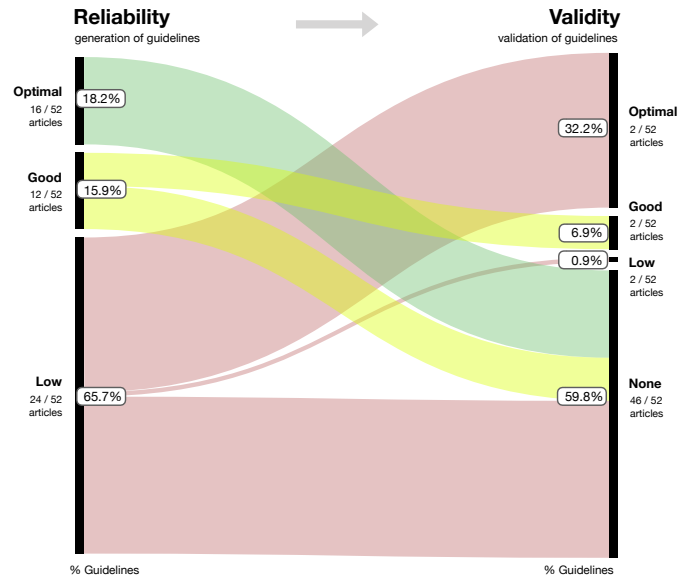


FIGURE 6: Reliability and Validity scores of all extracted guidelines. The flow between the dimensions represents the association between the quality for producing and validating the design guidelines

Reliability score of design guidelines by ability declines

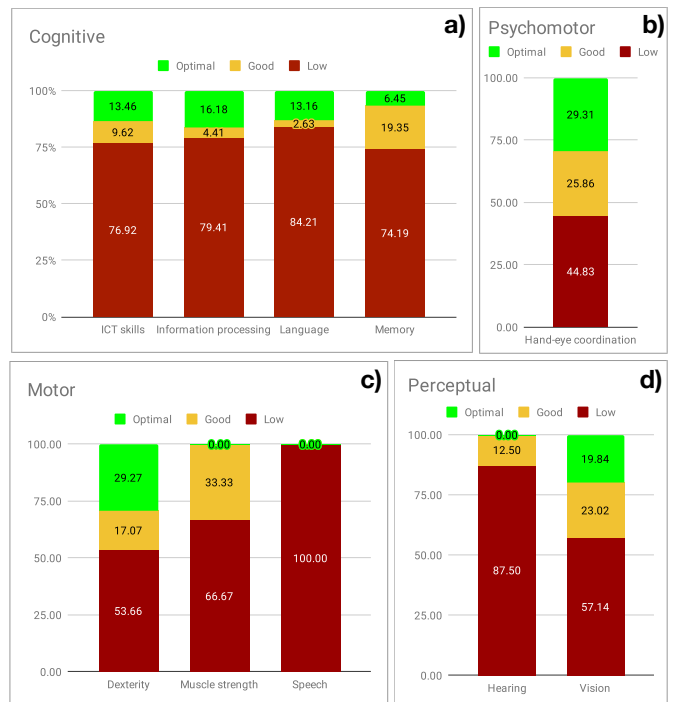


FIGURE 7: Reliability score distribution according to our capability model.

2) Validity of design guidelines

Unfortunately, only 11.5% (6 out of 52) of the included articles provided validation studies of their proposed guidelines, while the rest 88.5% did not validate their findings.

From the group of articles with no guideline validation, two studies attempted some testing: Rodrigues et al. [84] performed a computer simulation to evaluate the guidelines, but did not involve end users; Ruzic et al. [10] used their proposed guidelines in a software development process but without further testing the software with the target population.

The studies that actually validated their proposed guidelines ([12], [13], [25], [34], [69], [78]) usually did it by applying them in the development of an application and later testing it with the target population.

After analyzing the 6 studies that performed guideline validation, we computed the following validity scores:

- **Optimal:** we identified 2 articles that fall in this level [25], [69]. Both of them target users with disabilities.
- **Good:** we identified 2 articles that fall in this level [12], [13], and both of them target users with disabilities.
- **Low:** we identified 2 articles that fall in this level [34], [78], targeting older users with no disabilities.

The validity score of each article is passed down automatically to the guidelines extracted from that article. Figure 6 shows the distribution of our validity score for all the extracted guidelines.

Figure 8 shows the distribution of the validity score according to our capability model. In this case the distribution of the psychomotor category is the one that has the least validated guidelines together with motor (consider that speech is just 1 guideline). In the cognitive category, almost half of the guidelines are validated and most of them received an optimal validity score. Finally we have the perceptual category where more than half of the guidelines are validated and the majority of them received an optimal validity score.

For the interested reader, we included in the appendix (Table 5) the list of included articles that performed a user study for either proposing guidelines or for validating them, with the list of participants of the study, how many were considered older adults, and the score that we gave to the study based on our scoring method described in section IV-E

D. IDENTIFICATION AND CATALOGING OF DESIGN GUIDELINES

In this subsection we address our third, and last, research question: *What issues emerge and what effort is required in identifying and cataloging research-derived guidelines, as to make them available to the average practitioner?*

1) Identifying guidelines

Thus, to address this question and give an indication of the effort required in the process of *identifying and extracting* design guidelines, we describe the guidelines in terms of how they were reported, and how salient they were:

Validity score of design guidelines by ability declines

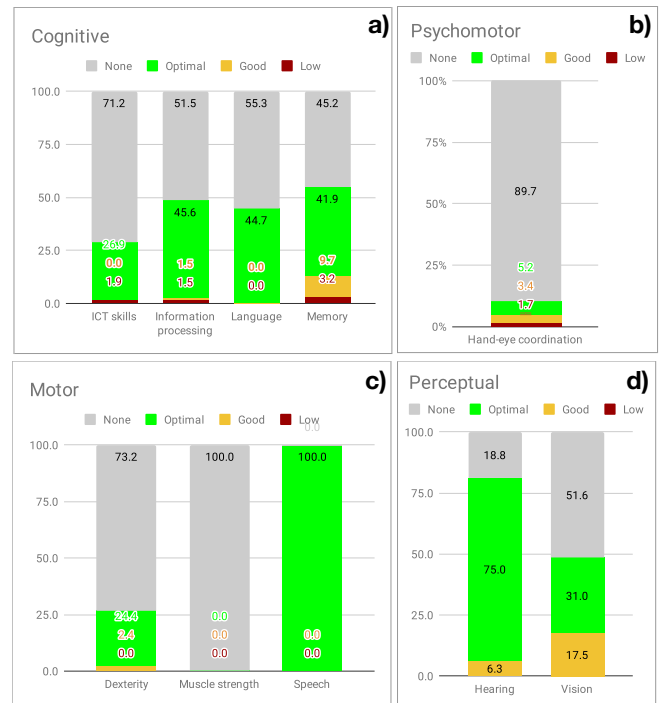


FIGURE 8: Validity score distribution according to our capability model.

- Almost 67% of the included articles **present guidelines in a clear and structured format**, presenting them explicitly in checklists, making them easy to identify, or at least narrowing the search to a section of the article;
- Other 33% **present guidelines in an unstructured format**, writing them in the form of discussions, using unclear formulation (sometimes simply incomprehensible), and consequently, making it difficult to assess whether certain findings could be indeed defined as design guidelines.

Identifying and extracting design guidelines when they were not structured was, generally, a complicated task.

This task was performed independently by three researchers, to reduce bias in what is still a subjective process, which was followed by face-to-face discussions in order to resolve disagreements consisted in approximately 17.4% for all the analyzed articles. Disagreement was calculated based on whether an article contained design guidelines, which was straight forward if they were presented in a structured way, otherwise the first three authors had to interpret the contribution to evaluate if it represented a guideline.

2) Cataloging guidelines

Cataloging guidelines can inform practitioners of the focus and target of design guidelines, but doing so requires experts to characterize them based on a reference taxonomy. To provide insights on the required effort and potential issues in

cataloging guidelines, we report on the classification process involving our design taxonomy.

The guideline classification process was more challenging, as it required interpreting the limited guideline text and associating it to one of the categories from the chosen taxonomies. In this case we had a 55% interrater agreement on ability categories, and 59% on design categories.

To further avoid ambiguity, the original authors of the articles were contacted for feedback on the identified guidelines, to confirm (or correct) our guideline classification and rephrasing. This resulted in 65% answer rate. Original authors either confirmed or commented on the classification and guideline phrasing, few of them asked for more detailed information about our work. There were also cases of authors that requested to re-word their initial findings and make some corrections, which we did.

E. FOCUS GROUP AND EXPERT EVALUATION

We conducted a focus group discussion with healthcare professionals specialized on aging (geriatric care) in order to assess our capability model, and get their value judgment on the changing abilities that require more attention.

Experts started with an overview of their practice and the challenges they most commonly face. With respect to physical ability declines, experts commented that older adults they treat are generally affected by physical musculoskeletal declines (usually due to arthritis), visual and hearing declines, hand tremors, and mobility difficulties (walking, stairs climbing, maneuvering obstacles). Their patients are usually 80+ (old-old), homebound, and do not actively participate in community/social life. As for the most common cognitive declines, healthcare professionals mentioned memory problems, executive function difficulties (planning, organizing), and low ICT skills. Experts also indicated that in absence of a major critical health event, e.g., a heart attack or a stroke, physical declines appear first (and patients might stay cognitively active until very old age), however, if there is a cognitive decline, physical declines unavoidably affect older adults as a consequence.

From their condition as experts, and based on their experience, the participants commented that the capability model covered the most important changing abilities. However, they added the following points to consider:

- Aging related ability declines are symptoms of common diseases that appear in older age;
- They emphasized the importance of using screening tools when addressing ability declines and recruiting older population groups, especially the oldest old (80+);
- As diabetes is a common disease in older age, reduced touch sensitivity is a common decline among older adults.

An important takeaway is that declines are not always independent, but are oftentimes manifestations of conditions that affect more than one ability. This indicates that to make guidelines really actionable, design guideline compilations

and repositories should provide “profiles” of typical conditions that would facilitate the mapping to changing abilities and guidelines. Our capability model provides the building blocks to build such a profile-based discovery.

Finally, the experts were asked to rank the ability declines from our capability model based on their clinical geriatric experience, which resulted in the following list of the most relevant declines for each ability category. The most relevant **cognitive** declines are:

- 1) Reduced memory;
- 2) Problematic information processing;
- 3) Low or absent ICT skills;
- 4) Language processing problems.

The most relevant **perceptual** declines are:

- 1) Various vision declines;
- 2) Hearing declines.

The most relevant **motor** declines are:

- 1) Reduced muscle strength;
- 2) Dexterity problems;
- 3) Speech declines.

We pick up on this ranking later to discuss different aspects of our findings.

VI. DISCUSSION

In this section we discuss the implications of our results, the topics that emerged, and what we learned during this systematic literature review.

The last decade of research on touchscreen design guidelines form an extensive body of valuable recommendations that target a wide range of services and technologies being adapted to ageing related declines. The works included in this review address important and critical questions of making touchscreen devices usable by older adults through various design recommendations. However, the compilation process uncovered some shortcomings in terms of coverage, formulation, poor structuring, and reliability of findings. We discuss those issues below.

A. CHARACTERISTICS OF THE OLDER ADULT POPULATION AND INTERACTION DESIGN (RQ1)

1) Different chronological definitions, targeting general rather than specific populations, and an emerging functional focus
Two approaches to defining the target population were identified in this review: chronological and functional definitions.

We observed articles that adhered to official chronological definitions, setting the starting age as early as 55. Interestingly, the target was dominated by the younger end, with only 8 out of 30⁷ articles starting at 65+. There was also a majority of articles addressing population ranges spanning more than 20 years, which points to rather wider ranges of the target population.

The issue with these rather general definitions is that older adults conform to a heterogeneous group [3], [30], [89], [90],

⁷The number of articles that reported on conducted user studies involving older population

where differences in functional abilities can greatly vary from the young to the oldest old, as declines tend to accelerate with age [91], [92]. For the same reason, focusing on younger populations has the effect of leaving out individuals that are more likely to benefit from the implementation of proper guidelines, and therefore of ICT.

Older adults within the same age cohort can also be expected to be different, as declines can be moderated by many factors such as level of physical activity, social connections, education, presence of disabilities, among others [91]–[94]. Thus, taking a chronological view only is a limited way of describing the older adult population.

Nearly half of the articles we reviewed, however, adopted a functional approach, addressing specific ability changes – or adapting interactions to the special abilities of the population – instead of generalizing groups by age. In taking this perspective, the resulting guidelines have the potential of talking more precisely to individuals, and supporting design approaches centered in user abilities (e.g., Wobbrock et al. [95]).

The takeaway message here is that defining more precisely the target population is paramount to having guidelines that can effectively guide practitioners in the design process, avoiding some of the stereotyping discussed in the literature [96]. Guidelines cannot be expected to be “universal” and generalize to the entire older adult population but rather cater to different, possibly smaller, groups of individuals. This should be clear to software designers and developers who do not always have a realistic picture of their target population when they refer to older adults [96], usually treating them as a homogeneous group that is affected by a set of physical and cognitive declines [89].

2) Design guidelines archetypes covering the most important ability and interaction design dimensions

The guideline categorization process revealed that each design parameter is not exclusive to just one ability, and neither is one ability determined by a single design dimension. Confirming and extending this observation, guideline distribution clearly shows that design of multimedia content and display composition (layouts) are crucial elements in addressing perceptual (vision and hearing) declines, as well as reduced cognitive (information processing) abilities. Another topic that emerged in guideline distribution analysis is the importance of efficient interaction styles and input techniques to address dexterity and ageing related hand-eye coordination changes. As for the cognitive declines in general and low ICT skills in particular, beside efficient interaction support, implementing appropriate user navigation, error and feedback handling were the most prominent.

Previous research has also raised the importance of proper design of displays in touchscreen devices for older adults. In their literature review, Petrovčič et al. [27] investigated the categories that were included in the mobile design guidelines and checklists, they found that the most frequently mentioned categories were related to selected *visual and haptic issues*

(e.g., high contrast, font size, button type, button size, button positioning), which aligns with our observations and emphasizes the importance of appropriate visualization of information as well as providing adapted interaction opportunities for users with ageing related ability declines, such as reduced vision.

Research on interaction styles emerged as another hot topic in this review. Although the *direct* input enabled by touchscreen devices are known to be more accessible than the *indirect* input provided by the traditional mouse and keyboard interactions [97], some type of gestures can still be more problematic than others for older adults. Indeed, Motti and colleagues [37] tell us in their literature review of interaction techniques that effectiveness of touchscreen interactions depend on the skills and background of the user but also on the configurations of the devices and specific technique used. Our review shows that efforts are well focused in this area, especially when it comes to addressing coordination and dexterity declines.

The organization of both display composition and user navigation, given the changes information processing abilities, are prominent archetypes in our analysis that have also been identified as important by previous research. Wildenbos et al. [98] investigated the usability issues encountered by older adults using mHealth apps, and identified that the most severe issues were related to unnatural navigation through the App, which was affected by slower cognitive performance but also technology anxiety, longer learning time and speed of performance. The same work also refers the importance of designing feedback to accommodate to lower ICT skills, an archetype identified in this review. According to Wildenbos and colleagues, the usability issues related to ‘forgiveness and feedback’ are affected by technology anxiety and low computer literacy, which translates into “Errors” and “Feedback” in our design taxonomy (see Figure 1). In their more recent work [99], the same authors also connect usability issues related to ‘Errors’ and ‘Efficiency’ to vision declines.

Looking at the coverage of perceptual ability declines, it is worth noticing that in this review there are only a handful of works that address specifically *hearing problems*, which are usually discussed in works related to general accessibility but there is no research specific for them. This could be explained by the focus on visual feedback and multi-modal interactions. Auditive interactions are used normally for notifications, and these can be replaced with other types of feedback, haptic or visual. Most of the proposals that address hearing declines are related to adding captions to videos and providing text alternatives to audio information. Other ability declines that have a very low coverage are *speech and muscle strength* (related to the motor category). The lack of guidelines for declines in speech has a similar explanation than for hearing; the lack of guidelines for declines in muscle strength could be explained by the lack of interactions that require grip (or a similar) type of interactions. Thus, in the context of touchscreen interfaces (including mobile phones and tablets), most of the research is focused on gesture-based interactions, with

conversation-based or natural interactions not very present in this medium.

B. QUALITY OF THE METHODS AND STRATEGIES USED TO GENERATE AND VALIDATE THE DESIGN GUIDELINES (RQ2)

As detailed in section V, we evaluated the quality of the procedures used to generate and validate the design guidelines in the included articles. These consisted of experiments and user studies with older adults [12], as well as analyses of findings with comparisons to an existing body of work. We start discussing first the findings related to the methods used for generating and validating guidelines, and then we discuss the quality of the particular guidelines, with a particular focus on the guidelines that address the most relevant ability declines that affect older adults (according to the experts from our focus group).

For the evaluation of user studies, we used the approach proposed in [46] and defined an objective and unbiased methodology. Unfortunately, for literature reviews and expert recommendations we could not find any methodological approach on how to evaluate these type of studies. Thus, we evaluated them assessing whether they followed the method of systematic literature review or not.

1) Good research on creation, more needed in validation

Taking in consideration all the methods that we found for proposing design guidelines (user studies, literature reviews, and expert evaluations) we have found that the quality was “good” or better (in the scale low, good, optimal) in a little more than half of the selected works (54%).

The average number of older adults recruited in studies to later derive guidelines consisted of 23 participants, if we consider the overall number of participants (including younger age groups), the average number of participants raises to 31. A little more than half of the user studies (19 out of 34) used some type of screening method to identify the presence of ability declines in older adults: In 5 of them participants were recruited directly from facilities for specific ability declines; in 4 of them participants self-reported their ability declines, e.g., lack of ICT skills; and in 10 of them participants were screened using validated methods like the Snellen eye chart (for measuring visual acuity), the Mini-Mental State Examination (for measuring cognitive impairment), and spiral drawing (for measuring hand tremor).

On the contrary, the validation of proposed guidelines has been stated only in about 11.5% of works, which represents a rather disappointing trend. This finding raises awareness of the need of further experimental investigations in order to determine the trustworthiness and efficacy of existing guidelines and providing an operational framework for a reliable generation of design recommendations.

We could say that the “low” quality of methods for user studies can be explained, or justified, by the recruitment difficulties of studies explicitly with older adults as has been mentioned earlier [27], [100].

2) Good quality of validated guidelines but still more validation needed

From the point of view of individual guidelines, we can see in Figure 6 that around 70% of guidelines have low reliability. This indicates that the studies with a good reliability or better (around 50% of them), unfortunately, did not propose many guidelines. If we do the same analysis for validity, we see that around 60% of guidelines are not validated. However, almost all of the validated guidelines have a good or better validity score. Furthermore, despite that only 11% of the included articles validated their proposed guidelines, they amount to almost 40% of all the validated guidelines, which means that each article validated a large number of guidelines.

Taking in consideration the most relevant ability declines (according to the experts from our focus group) we can see (Figure 7) that for: **cognitive** declines, around 75% of guidelines have a low reliability, this applies to all sub-categories; **perceptual** declines, around 60% of guidelines for vision (the most relevant ability decline for this category) have a low reliability; **motor** declines, around 70% of guidelines for muscle strength (the most relevant ability decline of this category) have a low reliability; and **psychomotor** declines, around 45% of guidelines have a low reliability. This is the only category (with hand-eye coordination as the only sub-category) where more than half of the guidelines have a good or better reliability. This is an indicator that there is a lot of room to improve the methods used to propose guidelines, especially the ones that address the most relevant declines.

A similar analysis for validity shows us a slightly different picture as not all guidelines have been validated, so for: **cognitive** declines, we have that around 50% of guidelines for the most relevant declines (memory and information processing) are validated and most of them have good or better validity. The other categories, ICT skills and Language, have around 30% and 50% of validated guidelines respectively, with most of them having good or better validity, **perceptual** declines, around 50% of guidelines for vision (the most relevant ability decline for this category) are validated and most of them have good or better validity. Most of guidelines for the hearing sub-category are validated but we need to consider that there are in total 16 of them (most of them with optimal validity), **motor** declines, none of the guidelines for muscle strength (the most relevant ability decline of this category) are validated. The dexterity sub-category has only a little more than 25% of validated guidelines and the speech category misguides as it has only 1 guideline, which is validated, and **psychomotor** declines, we have less than 10% of guidelines with validation. This is an indicator that more work is needed to validate proposed design guidelines, especially for the categories with less validation.

A last remark, we consider studies to validate guidelines slightly more important than studies to propose them, as the validation helps to confirm the utility of the proposed guidelines.

C. IDENTIFICATION OF RESEARCH-DERIVED GUIDELINES AND THEIR AVAILABILITY TO PRACTITIONERS (RQ3)

The process of identifying and cataloging research-derived guidelines included evaluation of the guideline visibility and usability, the effort required in extracting them and characterizing their focus. Based on the inferred guidelines quality discussed earlier, we next discuss the potential of guidelines to be discovered and correctly interpreted by practitioners.

1) Reporting can make identification of guidelines demanding, and lead to missed opportunities

One of the outcomes of our evaluation is that the guideline extraction process was straightforward just for half of them: the half where guidelines were clearly stated in the article. For the rest of the papers, guideline identification and extraction required much more effort and time, as guidelines were presented as experiment outcomes, future recommendations, and observations. As a consequence this process can be lengthy (as experimented by the authors), potentially prone to error, and may lead experts to overlook relevant guidelines – as reported in the results section, guidelines overlooked by a first expert were identified by the second one.

Previous research acknowledges that the identification of guidelines or the effort necessary to recognize and extract them from each selected paper highly depends on the way they are presented, as well as the skills of experts in identifying them [26]. Making findings and contributions difficult to identify and consume indicate a missed opportunity, as this prevents the uptake of recommendations by the larger community. The use of standard reporting formats for reporting, and the development of knowledge bases could help address this issue and benefit the whole community.

2) Challenging identification of the guideline's precise focus

One of the findings that emerged during the guideline classification process was the complexity of the process. Due to the limited guideline text and context, the level of detail provided in the articles, and the way they were reported, made the task of interpreting them and identifying their precise focus time consuming and challenging. It required multiple iterations of discussions among the authors and even contacting the authors of original articles. Another important issue that became evident from our review and analysis is the lack of a common validated categorization framework that could be universally adopted to classify existing research-based design guidelines.

Traditionally, the clarity of the guideline application purpose (usability) was investigated by either identifying the usability problems they target or by comparing guidelines with each other to detect the ones more useful [22]. In both cases, a general requirement of the guidelines is to be usable for designers to build efficient interfaces according to them, and for that it is particularly important that the design guidelines are easily accessible, clearly indicate their focus and addressed usability problems.

These findings complement those mentioned earlier and emphasize not only the need of having easier and clearer access to the best design practices for developing touchscreen applications for older adults, but also the need of a more structured approach in their categorization and validation.

As an attempt to target that issue, the final list of included papers and respective guidelines was depicted in a repository (<https://design-review.mateine.org>) as a collection of guidelines derived from our review. We believe that similar approaches to categorize new guidelines and make them available will benefit future contributions in this area. We also see its potential in allowing researchers and developers to apply and consult the guidelines while developing touchscreen applications or conducting studies for and with older adults.

One more thing that could help guidelines in being more usable is the indication of the technology for which they can be applied. We found that 37.3% of all articles state that their guidelines target medium “touchscreen devices” without specifying the exact type of them. From these, we have that 33.5% of them define these “touchscreen devices” as “mobile touchscreen devices” or sometimes as “smart-phones”. The rest of articles either specify a type of touch-based device, like tablet (11.8%) or tabletop (5.9%), or state that touchscreen devices were also included in their study together with other input devices [38], [51], [74]. The reason for this lack of specificity of devices could be due to the generality of some interactions, like visual interactions, that are touchless and thus, can be applied to several types of devices, or could be defined as device agnostic interactions.

VII. CONCLUSIONS

We performed a systematic literature review with identification of research trends on the topic of touchscreen design guidelines for older adults and gaps to be covered. Guidelines derived from the literature formed a list, which could be applied while developing inclusive touchscreen applications. In this process, we addressed three relevant research questions.

First, the target population aimed to be supported by the application of design guidelines was defined using different chronological definitions. We observed that they targeted general rather than specific populations and contained an emerging functional focus. As for the design dimension, we identified guideline archetypes covering the most important ability and interaction design dimensions, such as design of multimedia content to address vision declines or adapting content complexity to declines in language processing.

Second, we evaluated the process of guideline extraction to investigate their accessibility and clarity, which proved to be a challenging task. We introduced the capability model and the design taxonomy that we used to categorize the final list of research based design guidelines derived from the literature review. With this contribution, we present the areas related to the touchscreen interaction of older adults that are covered more than others, bring attention to their uneven distribution, and indicate the potential gaps that could benefit from future research.

Third, we analyzed the methodology that associated papers adopted in producing and validating design guidelines. By this, we aimed at making the guidelines more useful for designers and developers, supporting them in their understanding of the relevance of each guideline and its validity. Our findings point out to the need for validating existing design guidelines and for increasing the quality of guideline generation.

The question that remains is: “Is there a need for more research in areas that are lacking design guidelines?” or, maybe, by the nature of the type of touchscreen interactions there is no actual need for them. This question becomes even more important, as designers have to prioritize and choose wisely considering the possible compromises and trade-offs [101], but unfortunately, there is little guidance on how to choose and apply available guidelines [60].

As an additional contribution of our review, we provide a collection of research guidelines for touchscreen application targeted at older adults, and a reference taxonomy that could help in analyzing and characterizing guidelines.

The collection of guidelines could indicate which guidelines are validated and how they are distributed in covering abilities declines in the heterogeneous older adult population. This could help developers and designers in understanding better the abilities to be considered in the design. It can also help the community in identifying areas not currently covered by design guidelines, and motivate researchers to reproduce and validate existing findings.

As an attempt to address that need, the final list of included papers and respective guidelines are depicted in a repository (<https://design-review.mateine.org>) as a collection of guidelines derived from our review. We believe, it would allow researchers and developers to apply and consult the guidelines while developing touchscreen application or conducting studies for and with older adults, and has a potential to become a repository to submit new guidelines and make them available for future contributions in this area.

VIII. FUTURE WORK

Regardless of all the design recommendations addressing touchscreen devices that were extracted from the research performed in the last decade under the scope of applications for older adults, there are still gaps in this field.

We see the work on improving the reporting and characterisation of research-derived guidelines as a direction where much work is needed. It is also clear that a general effort towards validating and reproducing the findings from the community will help in providing more solid and actionable research guidelines. A third direction we find promising is that of developing tools that can facilitate the process of discovering design guidelines and embedding them in the design process. We are currently pursuing these lines of research, building on the foundation of this work.

IX. LIMITATIONS

Classification limited to available information. The classification of guidelines was performed based on the information provided in the research articles, and discussions between the researchers. Given the limitations in the reporting styles, the classification might not correspond to the actual focus intended by the authors. We addressed this limitation by collecting feedback from the authors, but the information was not available in all cases.

Guideline assessment limited to the methods used in the research articles. The assessment heuristic employed in this paper was based on previous literature, limited to the methods used to produce or validate the guidelines. This gives us an indication of the process used in the research articles, but it does not guarantee that the resulting guidelines are indeed of good quality.

ACKNOWLEDGMENT

This project has received funding from the EU Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690962. This work was also supported by the “Collegamenti” project funded by the Province of Trento (l.p. n.6-December 13rd 1999)

REFERENCES

- [1] Y. Barnard, M. D. Bradley, F. Hodgson, and A. D. Lloyd, “Learning to use new technologies by older adults: Perceived difficulties, experimentation behaviour and usability,” *Computers in Human Behavior*, vol. 29, no. 4, pp. 1715–1724, 2013.
- [2] S. J. Parker, S. Jessel, J. E. Richardson, and M. C. Reid, “Older adults are mobile too! identifying the barriers and facilitators to older adults’ use of mhealth for pain management,” *BMC geriatrics*, vol. 13, no. 1, p. 43, 2013.
- [3] P. Gorce, V. Nadine, and L. Motti, “Interaction techniques for older adults using touchscreen devices: a literature review from 2000 to 2013,” *Journal d’Interaction Personne-Système*, vol. 3, 2017.
- [4] N. L. Hill, J. Mogle, E. Colancecco, R. Dick, J. Hannan, and F. V. Lin, “Feasibility study of an attention training application for older adults,” *International journal of older people nursing*, vol. 10, no. 3, pp. 241–249, 2015.
- [5] A. Kuerbis, A. Mulliken, F. Muench, A. A. Moore, and D. Gardner, “Older adults and mobile technology: Factors that enhance and inhibit utilization in the context of behavioral health,” *Mental Health and Addiction Research*, 2017.
- [6] B. D. Carpenter and S. Buday, “Computer use among older adults in a naturally occurring retirement community,” *Computers in Human Behavior*, vol. 23, no. 6, pp. 3012–3024, 2007.
- [7] B. O’Mullane, B. Bortz, and R. B. Knapp, “Design challenges for internet interfaces for older people and widget-based, touch screen solution architecture,” *Gerontechnology*, vol. 9, no. 2, p. 239, 2010.
- [8] J. Coelho and C. Duarte, “Socially networked or isolated? differentiating older adults and the role of tablets and television,” in *Human-Computer Interaction*. Springer, 2015, pp. 129–146.
- [9] K. G. Vroman, S. Arthanat, and C. Lysack, “Who over 65 is online?? older adults’ dispositions toward information communication technology,” *Computers in Human Behavior*, vol. 43, pp. 156–166, 2015.
- [10] L. Ruzic, S. T. Lee, Y. E. Liu, and J. A. Sanford, “Development of universal design mobile interface guidelines (udmig) for aging population,” in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2016, pp. 98–108.
- [11] L. R. Kascak, E. Y. Liu, J. A. Sanford et al., “Universal design (ud) guidelines for interactive mobile voting interfaces for older adults,” in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2015, pp. 215–225.

- [12] N. Mi, L. A. Cavuoto, K. Benson, T. Smith-Jackson, and M. A. Nussbaum, "A heuristic checklist for an accessible smartphone interface design," *Universal access in the information society*, vol. 13, no. 4, pp. 351–365, 2014.
- [13] E. Corbett and A. Weber, "What can i say?: addressing user experience challenges of a mobile voice user interface for accessibility," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2016, pp. 72–82.
- [14] S. Trewin, C. Swart, and D. Pettick, "Physical accessibility of touchscreen smartphones," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 2013, p. 19.
- [15] K. S. Niksirat, C. Silpasuwanchai, Z. Wang, J. Fan, and X. Ren, "Age-related differences in gross motor skills," in *Proceedings of the International Symposium on Interactive Technology and Ageing Populations*. ACM, 2016, pp. 109–118.
- [16] D. Rømen and D. Svanæs, "Validating wcag versions 1.0 and 2.0 through usability testing with disabled users," *Universal Access in the Information Society*, vol. 11, no. 4, pp. 375–385, 2012.
- [17] H. Nicolau, T. Guerreiro, J. Jorge, and D. Gonçalves, "Mobile touchscreen user interfaces: bridging the gap between motor-impaired and able-bodied users," *Universal access in the information society*, vol. 13, no. 3, pp. 303–313, 2014.
- [18] J. Abascal, M. Arrue, I. Fajardo, N. Garay, and J. Tomás, "The use of guidelines to automatically verify web accessibility," *Universal Access in the Information Society*, vol. 3, no. 1, pp. 71–79, 2004.
- [19] P. Zaphiris, M. Ghiawadwala, and S. Mughal, "Age-centered research-based web design guidelines," in *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 2005, pp. 1897–1900.
- [20] S. Carmien and A. G. Manzanares, "Elders using smartphones – a set of research based heuristic guidelines for designers," in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2014, pp. 26–37.
- [21] N. Caprani, N. E. O'Connor, and C. Gurrin, *Touch screens for the older user*. InTech, 2012.
- [22] H. Kim, "Effective organization of design guidelines reflecting designer's design strategies," *International Journal of Industrial Ergonomics*, vol. 40, no. 6, pp. 669–688, 2010.
- [23] A. F. Newell and P. Gregor, "Design for older and disabled people—where do we go from here?" *Universal Access in the Information Society*, vol. 2, no. 1, pp. 3–7, 2002.
- [24] C. M. Law, J. S. Yi, Y. S. Choi, and J. A. Jacko, "Are disability-access guidelines designed for designers?: do they need to be?" in *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. ACM, 2006, pp. 357–360.
- [25] F. Ghorbel, E. Metais, N. Ellouze, F. Hamdi, and F. Gargouri, "Towards Accessibility Guidelines of Interaction and User Interface Design for Alzheimer's Disease Patients," in *ACHI 2017, Nice, France, March 2017*, p. 6.
- [26] M. Hertzum, "Images of usability," *Intl. Journal of Human-Computer Interaction*, vol. 26, no. 6, pp. 567–600, 2010.
- [27] A. Petrovčič, S. Taipale, A. Rogelj, and V. Dolničar, "Design of mobile phones for older adults: An empirical analysis of design guidelines and checklists for feature phones and smartphones," *International Journal of Human-Computer Interaction*, vol. 34, no. 3, pp. 251–264, 2018.
- [28] A. F. Newell, "Older people as a focus for inclusive design," *Gerontechnology*, vol. 4, no. 4, pp. 190–199, 2006.
- [29] C. Ponsard, P. Beaujeant, and J. Vanderdonckt, "Augmenting accessibility guidelines with user ability rationales," in *IFIP Conference on Human-Computer Interaction*. Springer, 2013, pp. 579–586.
- [30] J. Vines, G. Pritchard, P. Wright, P. Olivier, and K. Brittain, "An age-old problem: Examining the discourses of ageing in hci and strategies for future research," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 22, no. 1, p. 2, 2015.
- [31] S. L. Smith and J. N. Mosier, *Guidelines for designing user interface software*. Mitre Corporation Bedford, MA, 1986.
- [32] A. J. Dix, J. E. Finlay, and G. D. Abowd, "Beale rusel," *Human-Computer Interaction*, 2004.
- [33] T. Stewart and D. Travis, "Guidelines, standards, and style guides," in *The human-computer interaction handbook*. L. Erlbaum Associates Inc., 2002, pp. 991–1005.
- [34] T. Apted, J. Kay, and A. Quigley, "Tabletop sharing of digital photographs for the elderly," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 781–790.
- [35] V. L. Claypoole, B. L. Schroeder, and A. D. Mishler, "Keeping in touch: Tactile interface design for older users," *ergonomics in design*, vol. 24, no. 1, pp. 18–24, 2016.
- [36] J. Johnson and K. Finn, *Designing User Interfaces for an Aging Population: Towards Universal Design*. Morgan Kaufmann, 2017.
- [37] L. G. Motti, N. Vigouroux, and P. Gorce, "Interaction techniques for older adults using touchscreen devices: a literature review," in *Proceedings of the 25th Conference on l'Interaction Homme-Machine*. ACM, 2013, p. 125.
- [38] F. Boll and P. Brune, "User interfaces with a touch of grey?—towards a specific ui design for people in the transition age," *Procedia Computer Science*, vol. 63, pp. 511–516, 2015.
- [39] M. Peissner, D. Häbe, D. Janssen, and T. Sellner, "Myui: generating accessible user interfaces from multimodal design patterns," in *Proceedings of the 4th ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 2012, pp. 81–90.
- [40] A. Sutcliffe, "Collaborative requirements engineering: bridging the gulfs between worlds," in *Intentional Perspectives on Information Systems Engineering*. Springer, 2010, pp. 355–376.
- [41] S. Treu, "Computer applications and tasks," in *User Interface Design*. Springer, 1994, pp. 61–83.
- [42] A. Hochberg, J.-H. Hafke, and J. Raab, *Open I Close: Windows, Doors, Gates, Loggias, Filters*. Birkhäuser Verlag AG, 2010, vol. 1.
- [43] M. H. Chignell, "A taxonomy of user interface terminology," *ACM SIGCHI Bulletin*, vol. 21, no. 4, p. 27, 1990.
- [44] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "ReprintâĀĀpreferred reporting items for systematic reviews and meta-analyses: the prisma statement," *Physical therapy*, vol. 89, no. 9, pp. 873–880, 2009.
- [45] Q. Li and Y. Luximon, "A comparative study between younger and older users on mobile interface navigation," in *Proceedings of the ACHI 2017: The Tenth International Conference on Advances in Computer-Human Interactions*, 2017.
- [46] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.
- [47] W. H. Organization et al., "The global strategy and action plan on ageing and health. 2016," 2016.
- [48] N. Unies, *World population prospects: the 2017 revision: key findings and advance tables*. UN, 2017.
- [49] B. B. Neves, R. L. Franz, C. Munteanu, R. Baecker, and M. Ngo, "My hand doesn't listen to me!: Adoption and evaluation of a communication technology for the 'oldest old'," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 1593–1602.
- [50] H. N. Kim, T. L. Smith-Jackson, and B. M. Kleiner, "Accessible haptic user interface design approach for users with visual impairments," *Universal access in the information society*, vol. 13, no. 4, pp. 415–437, 2014.
- [51] J. Coelho and C. Duarte, "A literature survey on older adults' use of social network services and social applications," *Computers in Human Behavior*, vol. 58, pp. 187–205, 2016.
- [52] R. P. M. Fortes, G. A. Martins, and P. C. Castro, "A review of senescent's motivation in the use of tactile devices," *Procedia Computer Science*, vol. 67, pp. 376–387, 2015.
- [53] M. S. Al-Razgan, H. S. Al-Khalifa, M. D. Al-Shahrani, and H. H. AlAjmi, "Touch-based mobile phone interface guidelines and design recommendations for elderly people: a survey of the literature," in *International Conference on Neural Information Processing*. Springer, 2012, pp. 568–574.
- [54] C. Wacharamanatham, J. Hurtmanns, A. Mertens, M. Kronenbuerger, C. Schlick, and J. Borchers, "Evaluating swabbing: a touchscreen input method for elderly users with tremor," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 623–626.
- [55] F. Nunes, P. A. Silva, J. Cevada, A. C. Barros, and L. Teixeira, "User interface design guidelines for smartphone applications for people with parkinson's disease," *Universal Access in the Information Society*, vol. 15, no. 4, pp. 659–679, 2016.
- [56] D. M. Martin, R. Laird, F. Hwang, and C. Salis, "Computerized short-term memory treatment for older adults with aphasia: feedback from clinicians," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 2013, p. 44.
- [57] O. Polacek, A. J. Sporka, and P. Slavik, "Text input for motor-impaired people," *Universal Access in the Information Society*, vol. 16, no. 1, pp. 51–72, 2017.

- [58] T. Guerreiro, H. Nicolau, J. Jorge, and D. Gonçalves, "Towards accessible touch interfaces," in Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility. ACM, 2010, pp. 19–26.
- [59] M. Huenerfauth, L. Feng, and N. Elhadad, "Comparing evaluation techniques for text readability software for adults with intellectual disabilities," in Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility. ACM, 2009, pp. 3–10.
- [60] J. Borg, A. Lantz, and J. Gulliksen, "Accessibility to electronic communication for people with cognitive disabilities: a systematic search and review of empirical evidence," *Universal Access in the Information Society*, vol. 14, no. 4, pp. 547–562, 2015.
- [61] J. Sevilla, G. Herrera, B. Martínez, and F. Alcántud, "Web accessibility for individuals with cognitive deficits: A comparative study between an existing commercial web and its cognitively accessible equivalent," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 14, no. 3, p. 12, 2007.
- [62] R. J. P. Damaceno, J. C. Braga, and J. P. Mena-Chalco, "Mobile device accessibility for the visually impaired: problems mapping and recommendations," *Universal Access in the Information Society*, pp. 1–15, 2018.
- [63] R. Calvo, A. Iglesias, and L. Castaño, "Evaluation of accessibility barriers and learning features in m-learning chat applications for users with disabilities," *Universal Access in the Information Society*, vol. 16, no. 3, pp. 593–607, 2017.
- [64] S. Bachl, M. Tomitsch, C. Wimmer, and T. Grechenig, "Challenges for designing the user experience of multi-touch interfaces," in Proc. workshop on engineering patterns for multi-touch interfaces, 2010.
- [65] A. Komminos, E. Nicol, and M. D. Dunlop, "Designed with older adults to support better error correction in smartphone text entry: The maxieyboard," in Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. ACM, 2015, pp. 797–802.
- [66] R. Miñón, F. Paternò, M. Arrue, and J. Abascal, "Integrating adaptation rules for people with special needs in model-based ui development process," *Universal Access in the Information Society*, vol. 15, no. 1, pp. 153–168, 2016.
- [67] R. Calvo, F. Seyedarabi, and A. Savva, "Beyond web content accessibility guidelines: Expert accessibility reviews," in Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. ACM, 2016, pp. 77–84.
- [68] C. N. H. Ljilja Ruzic and J. A. Sanford, "Design and evaluation of mobile interfaces for an aging population," in Proceedings of the ACHI 2017: The Tenth International Conference on Advances in Computer-Human Interactions, 2017.
- [69] L. Muskens, R. van Lent, A. Vijfinkel, P. van Cann, and S. Shahid, "Never too old to use a tablet: designing tablet applications for the cognitively and physically impaired elderly," in International Conference on Computers for Handicapped Persons. Springer, 2014, pp. 391–398.
- [70] M. Kobayashi, A. Hiyama, T. Miura, C. Asakawa, M. Hirose, and T. Ifukube, "Elderly user evaluation of mobile touchscreen interactions," *Human-computer interaction-INTERACT 2011*, pp. 83–99, 2011.
- [71] R. X. E. de Almeida, S. B. L. Ferreira, and H. P. Soares, "Recommendations for the development of web interfaces on tablets/ipads with emphasis on elderly users," *Procedia Computer Science*, vol. 67, pp. 140–149, 2015.
- [72] S. Harada, D. Sato, H. Takagi, and C. Asakawa, "Characteristics of elderly user behavior on mobile multi-touch devices," in IFIP Conference on Human-Computer Interaction. Springer, 2013, pp. 323–341.
- [73] A. M. Piper, R. Campbell, and J. D. Hollan, "Exploring the accessibility and appeal of surface computing for older adult health care support," in Proceedings of the sigchi conference on human factors in computing systems. ACM, 2010, pp. 907–916.
- [74] W. A. Rogers, A. D. Fisk, A. C. McLaughlin, and R. Pak, "Touch a screen or turn a knob: Choosing the best device for the job," *Human Factors*, vol. 47, no. 2, pp. 271–288, 2005.
- [75] Q. Gao and Q. Sun, "Examining the usability of touch screen gestures for older and younger adults," *Human factors*, vol. 57, no. 5, pp. 835–863, 2015.
- [76] Z. X. Jin, T. Plocher, and L. Kiff, "Touch screen user interfaces for older adults: button size and spacing," in International Conference on Universal Access in Human-Computer Interaction. Springer, 2007, pp. 933–941.
- [77] W.-C. Tsai and C.-F. Lee, "A study on the icon feedback types of small touch screen for the elderly," *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, pp. 422–431, 2009.
- [78] A. B. Naumann, I. Wechsung, and J. Hurtienne, "Multimodal interaction: A suitable strategy for including older users?" *Interacting with Computers*, vol. 22, no. 6, pp. 465–474, 2010.
- [79] L. Wang, H. Sato, P.-L. P. Rau, K. Fujimura, Q. Gao, and Y. Asano, "Chinese text spacing on mobile phones for senior citizens," *Educational Gerontology*, vol. 35, no. 1, pp. 77–90, 2008.
- [80] S. Vetter, J. Bützler, N. Jochems, and C. M. Schlick, "Fitts' law in bivariate pointing on large touch screens: Age-differentiated analysis of motion angle effects on movement times and error rates," in International Conference on Universal Access in Human-Computer Interaction. Springer, 2011, pp. 620–628.
- [81] G. R. Reddy, A. Blackler, D. Mahar, and V. Popovic, "The effects of cognitive ageing on use of complex interfaces," in Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction. ACM, 2010, pp. 180–183.
- [82] A. L. Smith and B. S. Chaparro, "Smartphone text input method performance, usability, and preference with younger and older adults," *Human factors*, vol. 57, no. 6, pp. 1015–1028, 2015.
- [83] I. Darroch, J. Goodman, S. Brewster, and P. Gray, "The effect of age and font size on reading text on handheld computers," *Human-Computer Interaction-INTERACT 2005*, pp. 253–266, 2005.
- [84] É. Rodrigues, M. Carreira, and D. Gonçalves, "Enhancing typing performance of older adults on tablets," *Universal Access in the Information Society*, vol. 15, no. 3, pp. 393–418, 2016.
- [85] A. Murata and H. Iwase, "Usability of touch-panel interfaces for older adults," *Human factors*, vol. 47, no. 4, pp. 767–776, 2005.
- [86] H.-T. Chang, T.-H. Tsai, Y.-C. Chang, and Y.-M. Chang, "Touch panel usability of elderly and children," *Computers in Human Behavior*, vol. 37, pp. 258–269, 2014.
- [87] H. Nicolau and J. Jorge, "Elderly text-entry performance on touchscreens," in Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility. ACM, 2012, pp. 127–134.
- [88] K. M. Afroza Sultana, "Target selection difficulties of older adults with small touch screen devices," in CHI'17 Workshop on Designing Mobile Interactions for the Ageing Populations. ACM, 2017.
- [89] D. Kuh, "A life course approach to healthy aging, frailty, and capability," *The Journals of Gerontology: Series A*, vol. 62, no. 7, pp. 717–721, 2007.
- [90] S. J. Czaja and C. C. Lee, "The impact of aging on access to technology," *Universal Access in the Information Society*, vol. 5, no. 4, p. 341, 2007.
- [91] T. A. Salthouse, "When does age-related cognitive decline begin?" *Neurobiology of aging*, vol. 30, no. 4, pp. 507–514, 2009.
- [92] A. S. Buchman, P. A. Boyle, R. S. Wilson, D. A. Fleischman, S. Leurgans, and D. A. Bennett, "Association between late-life social activity and motor decline in older adults," *Archives of internal medicine*, vol. 169, no. 12, pp. 1139–1146, 2009.
- [93] F. Landi, A. M. Abbatecola, M. Provinciali, A. Corsonello, S. Buscattini, L. Manigrasso, A. Cherubini, R. Bernabei, and F. Lattanzio, "Moving against frailty: does physical activity matter?" *Biogerontology*, vol. 11, no. 5, pp. 537–545, 2010.
- [94] K. Yaffe, A. J. Fiocco, K. Lindquist, E. Vittinghoff, E. M. Simonsick, A. B. Newman, S. Satterfield, C. Rosano, S. Rubin, H. Ayonayon et al., "Predictors of maintaining cognitive function in older adults the health abc study," *Neurology*, vol. 72, no. 23, pp. 2029–2035, 2009.
- [95] J. O. Wobbrock, S. K. Kane, K. Z. Gajos, S. Harada, and J. Froehlich, "Ability-based design: Concept, principles and examples," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 3, no. 3, p. 9, 2011.
- [96] D. Hawthorn, "Interface design and engagement with older people," *Behaviour & Information Technology*, vol. 26, no. 4, pp. 333–341, 2007.
- [97] E. Wood, T. Willoughby, A. Rushing, L. Bechtel, and J. Gilbert, "Use of computer input devices by older adults," *Journal of Applied Gerontology*, vol. 24, no. 5, pp. 419–438, 2005.
- [98] G. Wildenbos, L. W. Peute, and M. W. Jaspers, "A framework for evaluating mhealth tools for older patients on usability," in MIE, 2015, pp. 783–787.
- [99] G. Wildenbos, L. Peute, and M. Jaspers, "Aging barriers influencing mobile health usability for older adults: A literature based framework (mold-us)," *International journal of medical informatics*, 2018.
- [100] L. Mody, D. K. Miller, J. M. McGloin, M. Freeman, E. R. Marcantonio, J. Magaziner, and S. Studenski, "Recruitment and retention of older

- adults in aging research,” *Journal of the American Geriatrics Society*, vol. 56, no. 12, pp. 2340–2348, 2008.
- [101] S. S. Z. W. A. O. X. R. C. Silpasuwanchai, J. Jokinen, “Exploring the design space of text-entry interfaces for aging users,” in *CHI’17 Workshop on Designing Mobile Interactions for the Ageing Populations*. ACM, 2017.

TABLE 5: Quality of methods used in the articles that include user studies either in production or validation of guidelines

Article	Production of guidelines			Validation of guidelines		
	Type of study	No. participants	Score	Type of study	No. participants	Score
[25]	Non-systematic literature review	N/A	Low	Experiments with older adults	24 Alzheimer's patients	High
[34]	Non-systematic literature review	N/A	Low	Experiments with younger and older adults	12 younger and 12 older adults	Low
[69]	Non-systematic literature review	N/A	Low	Experiments with older adults	14 older adults with disabilities	High
[78]	Experiments with younger adults	62 younger adults	Low	Experiments with younger and older adults	15 younger and 15 older adults	Low
[12]	User studies with diverse age group	3 younger and 3 older adults	Medium	Experiments with diverse age group	6 younger and 3 older adults	Medium
[86]	Experiments with older adults and children	30 children and 30 older adults	High	N/A	N/A	N/A
[74]	Experiments with younger, middle-aged, and older adults	40 younger and 40 older adults	High	N/A	N/A	N/A
[85]	Experiments with young, middle-aged, and older adults	13 younger, 13 middle-aged, and 19 older adults	Low	N/A	N/A	N/A
[84]	Baseline study with younger adults followed by a study with older adults	20 younger and 20 older adults	Medium	N/A	N/A	N/A
[73]	Usability study with older and younger adults	10 younger and 20 older adults	Medium	N/A	N/A	N/A
[82]	Experiments with younger and older adults	25 younger and 25 older adults	Medium	N/A	N/A	N/A
[52]	Experiments with younger and older adults	20 younger and 20 older adults	Medium	N/A	N/A	N/A
[88]	Experiments with younger and older adults	16 younger and 20 older adults	Medium	N/A	N/A	N/A
[80]	Experiments with younger and older adults	15 younger and 15 older adults	Low	N/A	N/A	N/A
[71]	Experiments with younger and older adults	5 younger and 5 older adults	Low	N/A	N/A	N/A
[83]	Experiments with younger and older adults	12 younger and 12 older adults	Low	N/A	N/A	N/A
[59]	Experiment with diverse age group	20 adults, intellectual disabilities	High	N/A	N/A	N/A
[14]	Experiment with diverse age group	16 adults, dexterity impairments	High	N/A	N/A	N/A
[58]	Experiment with diverse age group	15 tetraplegic adults	High	N/A	N/A	N/A
[81]	Experiment with diverse age group	37 younger and older adults	Medium	N/A	N/A	N/A
[75]	Experiments with younger and older adults	40 younger and 40 older adults	High	N/A	N/A	N/A
[13]	Experiment with diverse age group	9 participants, motor impairments	Medium	Experiment with diverse age group	5 participants with motor impairments	Medium
[77]	Experiments with older adults	45 older adults	High	N/A	N/A	N/A
[76]	Experiments with older adults	66 older adults	High	N/A	N/A	N/A
[38]	Experiments with older adults	41 older adults	High	N/A	N/A	N/A
[54]	Experiments with older adults	10 older adults	Low	N/A	N/A	N/A
[79]	Experiments with older adults	12 older adults	Low	N/A	N/A	N/A
[4]	Experiments with older adults	9 older adults	Low	N/A	N/A	N/A
[65]	Workshops with older adults	N/A	Low	N/A	N/A	N/A
[70]	User studies with older adults	20 older adults	Medium	N/A	N/A	N/A
[72]	User studies with older adults	21 older adult	Medium	N/A	N/A	N/A
[87]	User studies with older adults	15 older adults	Medium	N/A	N/A	N/A
[49]	User studies with older adults	5 older adults	Low	N/A	N/A	N/A