

# Towards a List of Heuristics to Evaluate Smartphone Apps Targeted at Older Adults: A Study with Apps that Aim at Promoting Health and Well-being

Paula Alexandra Silva  
University of Aveiro – Portugal  
[palex@gmail.com](mailto:palex@gmail.com)

Kelly Holden  
University of Hawaii  
[kholden@hawaii.edu](mailto:kholden@hawaii.edu)

Philipp Jordan  
University of Hawaii  
[philippj@hawaii.edu](mailto:philippj@hawaii.edu)

## Abstract

*This paper presents a list of heuristics to evaluate smartphone apps for older adults. It further verifies the usefulness of the proposed list through a study performed with two groups of five expert evaluators, who inspected two popular health and fitness smartphone apps – Nike+ and Runkeeper – through heuristic evaluation. Additionally, the evaluators completed a post-evaluation survey to provide feedback to the researchers about the usefulness, strengths and gaps of the heuristics list.*

*The results of the heuristic evaluations and post-questionnaires demonstrate both a comprehensive application of the heuristics as well as an overall positive assessment of their quality and potential to identify usability issues of smartphone apps targeted at older adults.*

## Author Keywords

Heuristics; Smartphones; Older adults; Seniors; Evaluation; Heuristics evaluation, Guidelines, Mobile applications, HCI.

## 1. Introduction

Over-aging in developed countries in conjunction with an inevitable demographic change already has, and will continue to have, a significant impact on the economy of industrialized countries resulting in one of the most important challenges of mankind in the 21<sup>st</sup> century.

Now and in the future, the increasing healthcare needs for the growing fraction of older adults will strain government budgets making the management of healthcare expenses a main concern for the upcoming decades.

In parallel to the superannuation of the population, the rise of ubiquitous societies and their reliance on mobile technologies will continue to increase even more as these address needs that extend from daily living into almost every aspect of life, including personal health care and management. This

development becomes particularly important for the expanding older adult population, as this is naturally the group who is most prone towards age-related diseases and chronic conditions. This group is also the fastest growing group adopting smartphones [2].

As it is expected that smartphone apps will play an increasingly important role in fostering health and well-being [3], a suitable and inclusive mobile user interface design for older adults is essential for a successful adoption. Continuous and self-motivated physical exercise and an active lifestyle can play an important role, having not only the potential to improve perceived quality of life, health and well-being, but also to positively impact physical fitness. Even though a plethora of apps to promote, track, measure and improve physical activity exist at present [29], studies (e.g. [34]) have shown that their design is not inclusive for the specific older adults' user group. As outlined above, the disregard of this special target audience results in a substantial loss of potential benefits for older adults and society in general.

In consideration of the identified socio-demographical challenges and technological opportunities, it is crucial that smartphone apps accommodate for the characteristics of this expanding user group, making the assessment of the quality of the user interfaces designed for such user group a key topic to research.

This study addresses this research gap and provides HCI researchers and practitioners with a list of heuristics to evaluate smartphone apps for older adults. Additionally, the authors describe the research process followed, from the identification and creation of an appropriate and comprehensive list of heuristics, through to their application in a heuristic evaluation of two health-related smartphone apps, and its subsequent assessment by the evaluators. The apps chosen to further scrutinize the usefulness of the list of heuristics are Nike+ and Runkeeper, two apps that aim at promoting physical exercise.

## 2. Background

This section discusses technological opportunities and challenges of smartphone apps to improve health and well-being of older adults as well as user interface design and evaluation specifics relevant to this target audience.

### 2.1 The Potential of Smartphone Apps in Fostering Health

Health-related smartphone apps have been gaining popularity in the recent years and are mushrooming in their respective markets [29]. Application areas range from personal disease prevention and healthy living, via self-diagnosis using the device's built-in sensors through to medication compliance by the means of alarms and reminders. Overall, these apps aim at preserving and improving personal health and wellbeing. Nike+ and Runkeeper, later used in this study, specifically aim at promoting physical exercise.

Exercise is known to promote people's overall health status [14], [27], [28]. Concurrent research shows that participation in regular physical activity can substantially reduce the severity of disabilities associated with heart disease and other chronic illnesses [2]. Regular exercise for example, reduces the risk of cardiac death by 20 to 25% among people with established heart disease [3], whereas physical activity also reduces the risk of falls [4], one of the most common threats for older adults. Aside of being physically and mentally fit, regular physical exercise and activity can improve social bonding and decrease medical expenses as well, [5] and is therefore an important factor in maintaining and improving the quality of life of older adults.

Nike+ and Runkeeper are both apps that utilize the built-in sensors of the smartphone, such as the accelerometer or GPS module, to track statistics of physical activity and exercise, enabling users to not only monitor their health, but also to track progress and build towards a goal. Furthermore, these apps employ gamification principles in their user interfaces. Point systems, in the case of Nike+, NikeFuel for example, leaderboards, badges and sharing functionalities through social networks, both keep users engaged and motivated to reach individual goals and offer as well the ability to share results with their peers.

The underlying potential of these apps seems clear, however the question arises if these apps are actually designed in a way that accommodates the specific needs of older adults, a rising demographic group which would clearly benefit from their use. The fact is,

if these apps are not designed in an inclusive way, they are excluding a significant fraction of the population.

### 2.2. Designing User Interfaces for Older Adults

People change with age and as they do, psychosocial and functional changes take place, affecting vision, hearing, motion, cognition, and their relationship with themselves and others.

Technology has the potential to support the changing characteristics this expanding population experiences. This has led researchers to study a number of user interface related aspects for this special target population. These range from specific studies investigating gestural interfaces as an alternative to more conventional computer interaction methods for older adults [4] or the phenomenon of Internet adoption and aging [5], to more comprehensive studies investigating older adult's overall use of technology [9] [25] and elaborating on how to design displays for older adults [1].

Despite the reasonable maturity that mobile user interfaces already have, they still pose significant limitations and challenges due to their specific characteristics, such as the limited size of displays and their non-conventional input methods [7]. In an effort to better understand how to address these limitations, research has investigated, among others, the use and accessibility of mobile phones by and for older adults [24] and the adequate target sizes for the design of interfaces targeted at older adults [13] [15]. However, the majority of research existent so far still tends to focus on user interfaces specific to desktop or stationary computer technologies. This contrasts the already substantial, and growing reliance on mobile technology, therefore, it is important that user interface design for older adults needs to focus on these devices as well.

### 2.3 Evaluating User Interfaces for Older Adults

Coursaris and Kim [26] stated that *"increasing research on accessibility may improve the usability of products and services for often overlooked audiences"* They also present an adapted usability evaluation framework to the context of a mobile computing environment. Besides this specific framework, there are a large number of methods to assess the usability of user interfaces [19], [20], [22].

A popular method to perform usability inspections without requiring too much expertise, time and financial investment is Heuristic Evaluation [21], [22],

which has been largely applied offering positive and valid results.

Conducting a heuristic evaluation involves having a set of evaluators examine a user interface against a list of heuristics. Nielsen provides a list of 10 general Usability Heuristics for User Interface Design [23], but other lists can be defined depending on the evaluator preferences and project at hand, for example [17], [18], [6]; this study presents a similar effort.

Despite the chosen list of heuristics, the evaluators are instructed to individually inspect the user interface and, once all evaluations are completed, results are aggregated in a consolidated list of problems and corresponding violated heuristic(s). Finally, each problem is attributed a severity rating, from 0 to 4, where 0 is considered a 'cosmetic' problem and 4 a 'usability catastrophe' that is imperative to fix.

**2.3.1 Limitations of the Original 10 Heuristics of Nielsen.** As stated in the previous section, Nielsen provides a total of 10 Usability Heuristics for User Interface Design [23]. The universal character of Nielsen's list of heuristics allows for its application in the evaluation of virtually any user interface and therefore would also apply to the context of smartphone apps for older adults.

However, that list has not been created to specifically address the design of user interfaces for older adults users and mobile apps. Inevitably, a number of age-related specifics, such as the need for larger font sizes, increased contrast, bigger buttons and the need to limit the use of scroll are not sufficiently addressed and emphasized by the original 10 heuristics of Nielsen. When considering the evaluation of user interfaces targeted at older adults there is then an important opportunity for improvement.

### 3. Methodology

This research was carried out in two main stages: the first comprehending the process of finding a new comprehensive list of heuristics and a second stage in which the obtained list of heuristics is used in a heuristic evaluation by ten evaluators with the intent to verify its usefulness.

#### 3.1. Finding an Appropriate List of Heuristics

When researching the literature, no validated list of heuristics can be identified to evaluate smartphone apps targeted at older adults. In order to find an appropriate list of heuristics, a previous study by two of the authors of this paper and another author [28]

compiled the most well-known and used lists of heuristics targeted at older adults as Web users [10], [11], [12], as well as a Masters thesis that provides an initial list of smartphone evaluation heuristics for older adults [16]. This preceding study compiled a total of 153 heuristics (list available by request): 79 [11] + 36 [12] + 20 [10] + 18 [16] and resulted in a preliminary list of 35 heuristics for the evaluation of smartphone apps targeted at older adults.

This paper explains how this list was aggregated and further puts the list to scrutiny by evaluators that have not been exposed to it before. This is intended at furthering the understanding of the usefulness and appropriateness of the list.

The initially compiled list of 153 heuristics included, heuristics previously existent in the literature [10], [11], [12], [16] which were domain-specific, highly cited, and recognized. Once having aggregated all the heuristics, it was clear that they presented different levels of detail, with significant overlapping, and a limited applicability to smartphone user interfaces, mostly due to its web-specificity. The authors then decided to change the list in order to achieve a list of heuristics that would: i) fit this context, ii) be short and concise, but still concrete enough, and iii) present similar levels of detail.

In order to do so, three independent coders (two of the authors and another external coder) analyzed the list of 153 and individually coded the heuristics by similitude and by type of issue. Later, the three coders agreed upon their coding and reached consensus on determining which and how the heuristics should be grouped. Based on this decision, an aggregated, single heuristic list was created, respecting the original text or changing it in such a way that it would include the different nuances of the original heuristics in which it was based. For example, H10 – 'Write in a language that is simple, clear and adequate to the audience' was the result of the grouping of six distinct heuristics: language should be simple and clear [12]; focus the writing on audience and purpose [10]; use the users' language; minimize jargon and technical terms [10]; address your web users by using "you" [11]; choose words your web users know [11]; and, define unfamiliar terms [11].

The next two sub-sections introduce this new list of heuristics and the endeavors undertaken to verify its usefulness for the defined context: smartphone apps for older adults.

**3.1.1 A New List of Heuristics to Evaluate Smartphone Apps Targeted at Older Adults.** The process described in the previous section facilitated the creation of a comprehensive list of heuristics that includes 35 heuristics (Table 1). These were then

grouped by the coders according to their focus in the following categories: perception, cognition, dexterity, navigation, content, and visual design.

*Perception* includes heuristics that relate to limitations of the perceptual system that occur with age, such as hearing or visual acuity changes. *Cognition* contains heuristics that refer to cognitive changes that occur with age, such as the difficulty of maintaining attention or managing a large number of items by working memory. The heuristics on *Dexterity* are directly related to the difficulty of tasks created by limitations of motor skills. *Navigation* includes heuristics that are related to the understanding of the apps structure and how the user can flow through it. *Content* relates to the information and the language used in the apps. Finally, the *Visual Design* category covers aspects of design details, such as formatting and visual representations.

**Table 1. List of 35 heuristics**

O	F	Heuristic Description
<b>Cognition</b>		
H1	H1	Focus on one task at a time instead of requiring the user to actively monitor two or more tasks, and clearly indicate the name and status of the task at all times.
H2	H2	Avoid the use of interaction timeouts and provide ample time to read information.
H3	H3	Avoid the use of animation and fast-moving objects.
H4	H4	Leverage mental models familiar to older adults.
H5	H5	Reduce the demand on working memory by supporting recognition rather than recall.
H6	H6	Aim at creating an aesthetical user interface, by using pictures and/or graphics purposefully and adequately to minimize user interface clutter and avoid extraneous details.
<b>Content</b>		
H7	H7	Give specific and clear instructions and make help and documentation available. Remember that it is better to prevent an error than to recover from it.
H8	H8	Provide clear feedback and when presenting error messages make them simple and easy to follow.
H9	H9	Make sure errors messages are descriptive and use meaningful words and verbs when requiring an action.
H10	H10	Write in a language that is simple, clear and adequate to the audience.
<b>Dexterity</b>		
H11	H11	Avoid pull down menus.
H12	H12	Avoid the use of scrolling.
H13	H13	Enlarge the size of user interface elements in general; targets should be at least 14mm square.
<b>Navigation</b>		
H14	H14	Keep the user interface navigation structure narrow, simple and straightforward.
H15	H15	Use consistent and explicit step-by-step navigation.

O	F	Heuristic Description
H16	H16	Make sure that the "Back" button behaves predictably.
H17	H17	F: Support user control and freedom, allowing for alternative and flexible flows of interaction. O: Support user control and freedom.
H18	H18	Disable inactive user interface objects.
<b>Perception</b>		
H19	*	Allow users to fine tune the volume.
H20	H19	Do not rely on color alone to convey information. Be aware of color blindness.
H21	H20	Provide not only visual feedback, but also tactile and auditory.
H22	H21	Make information accessible through different modalities.
H23	H22	Use lower frequencies to convey auditory information such as confirmation tones and alerts.
H24	H23	Do not use pure white or rapidly changing contrast backgrounds.
H25	H24	Make it easy for people to change the text size directly from the screen.
H26	*	Allow users to fine-tune screen brightness and contrast.
<b>Visual Design</b>		
H27	H25	Use high-contrast color combinations of font and/or graphics and background to ensure readability and perceptibility; avoid using blue, green and yellow in close proximity.
H28	H26	Use color conservatively, limiting the maximum number of colors in use to ~four.
H29	H27	F: Make sure text uses types, styles and sizes appropriate to older adults, that is, for instance, but not exclusively: large-sized fonts, sans serif, non-condensed typefaces, non-italic, and left justified. O: Make sure text uses types, styles and sizes appropriate to older adults, for instance, but not exclusively: sans serif, non-condensed typefaces, non-italic, left justified and 12-14 point font.
H30	H28	Make links and buttons clearly visible and distinguishable from other user interface elements.
H31	H29	Make information easy to read, skim (or) and scan.
H32	H30	Group information visually (make good use of color, text, topics, etc.).
H33	H31	Allow sufficient white space to ensure a balanced user interface design.
H34	H32	Use user interface elements consistently and adhere to standards and conventions if those exist.
H35	H33	Use simple and meaningful icons.
<b>Note:</b> * - Removed from final validated list; **- Heuristic revised after evaluation; O – Original: F - Final		

### 3.2 Verifying the Proposed List of Heuristics

In order to verify the proposed list of heuristics, the authors designed a study that involved evaluating two health and fitness applications on both the iPhone iOS and Android platforms by a set of 10 evaluators - five

for one platform and five for the other, as five is the recommended number of evaluators to perform a heuristic evaluation [21]. The Android and Apple iOS platforms were chosen because they account for 82.1% of the smartphone platforms worldwide [8]. Health and fitness applications were chosen given the increasingly significant need of healthcare for older adults and the potential of health care monitoring and modification of behavior for the older adult population supported through these applications.

The study also included pre- and post-evaluation surveys, both distributed through Google Forms. The pre-evaluation survey served as a tool to vet the potential evaluators and ensure they had the right skill-set to analyze the list of heuristics. It also served to identify the devices evaluators would use in their evaluations and their familiarity with user interface design principles, older adults, mobile apps and heuristic evaluation. The post-evaluation questionnaire was meant to collect additional feedback from each evaluator about the strengths, weaknesses and gaps as well as usefulness, clarity, completeness, and appropriateness of the proposed list of heuristics.

**3.2.1. Choosing which Apps to Inspect.** It was the authors' intention to choose general public apps, given it is their conviction that older adults should be 'granted access' to apps the public is using instead of dedicated applications, that may be experienced as both, ostracizing and stigmatizing. Moreover, while the quality of interaction of mainstream users is unlikely to be hindered if inclusive user interface design principles are followed, the opposite design decision will inevitably result in the exclusion of the aging population. This is a rising and important fraction of the population, who can highly benefit from general health-related apps.

To find appropriate candidates for this study, in October 2013, the authors retrieved the top-rated, free-of-charge health and fitness apps from the Google Play and iTunes store. The results were narrowed down to apps targeting physical exercise and eventually Nike+ Running by Nike, Inc. and RunKeeper – GPS tracker by FitnessKeeper, Inc. were ultimately chosen as candidates for this study.

Both these apps utilize the sensor technology embedded in smartphones to monitor its users, such as accelerometers to measure activity, GPS to track the location or distance travelled and algorithms to calculate calories burnt and steps taken for any given physical activity. Nike+ also offers the user the ability to use an external sensor in the form of a wristband, tracking pace, distance, and time while running or walking. Furthermore, these apps use progress charts, histograms of weight-loss, activity leaderboards and

other comparison opportunities, that can foster social contact and increase motivation and behavior change towards physical exercise even more.

**3.2.2 Heuristic Evaluation Procedures.** Before starting the inspection, the evaluators received an Excel spreadsheet and a list of procedures (both available on request). The excel file contained the list of 35 heuristics aggregated in the previous phase with templates for evaluators to record the problems (and corresponding violated heuristic) they had found in the user interfaces as well as the time spent on each of the evaluations.

The list of procedures advised the evaluators to individually go through the interface at least twice: initially, to get a feel for the interaction flow of the user interface; and subsequently, to focus on the specific tasks (in the next paragraph) of the user interface and identify possible problems and violations of heuristics; this is the recommended procedure for conducting a heuristic evaluation [21].

Besides evaluating the overall application, the evaluators were requested to dedicate special attention to the areas that are fundamental for the users' onboarding and for the activities considered necessary to effectively use the apps. These included: registration and login; setting a quest; and monitoring own' progress and checking own achievements.

Finally, the evaluators were directed to clearly describe each problem identified in the user interface with reference to its violated heuristic(s) and, if desired, an image demonstrating the problem. Moreover, the evaluators were encouraged to be as specific as possible and to list each usability problem separately. No directions were given to evaluators regarding the order in which evaluations should be performed, but at the end, all evaluators started by evaluating Nike+ followed by RunKeeper.

**3.2.3 Pre-Evaluation Survey: Knowing the Evaluators.** Ten evaluators were enlisted to use the new list of heuristics. The evaluators were all HCI professionals, researchers, or postgraduate students and acquaintances of the authors that were recruited based on convenience and their availability and agreement to participate. All evaluators reported having experience with heuristic evaluation, familiarity with user interface design principles, older adult user interface design needs and with the usage of mobile apps.

From the 10 evaluators five worked with the Apple iOS platform while the remaining five worked with the Android platform. Each evaluator inspected both Nike+ and RunKeeper on their respective platforms. Those evaluators working with Apple iOS used either an iPhone 5 (three evaluators) or iPhone 4s (two

evaluators) for their evaluations. Those evaluators working with the Android platform were running Android 4.1.2 Jelly Bean on a Samsung Galaxy Rush- (one evaluator), a Samsung Galaxy S3 (two evaluators), a Samsung Galaxy 4 (one evaluator), and a Nexus 4 (one evaluator).

## 4. Results

The heuristics' verification study included twenty individual evaluations performed with the proposed 35 heuristics list and a follow-up post-evaluation survey. The results of these evaluations and the post-evaluation survey are described below. For a better understanding and comprehension of this section, the authors advise the reader to have Table 1 easily accessible.

### 4.1 Evaluation Insights

After all ten evaluators finished their evaluation, the results were analyzed to provide insight on the usefulness of the overall list as a tool to support heuristic evaluation of smartphone apps targeted at older adults.

In general, the evaluators identified more problems in the Nike+ app than RunKeeper, but overall all evaluators were able to find a significant number of problems in the apps and used a diversity of heuristics to qualify the types of problems found. Nike+ averaged 25.1 identified problems per evaluator and 14.4 violated heuristics per evaluator, while RunKeeper averaged 19 identified problems per evaluator and 11.6 violated heuristics per evaluator see Table 2.

All 35 heuristics were violated and issues were identified denoting the existence of user interface design problems for the older adult user. This indicates that all heuristics are valuable for the assessment of smartphone apps. The most highly utilized heuristic across the board (highly utilized on both platforms and apps) was H29 (related to font formatting) with 24 violations on the Android platform and 26 on iPhone. Moreover, heuristics H7 (related to instructions and content), H34 (related to consistency in visual design) and H35 (related to icons in visual design) are among the top used heuristics. Please refer to Table 3 for details on the number of violations per heuristic.

For iPhone (iOS), the evaluators consistently utilized H7, H13, H14, H17, H34, and H35. Of these, H13 was the most violated heuristic. For Android, H7, H34 and H35 were also highly utilized. However, unlike iPhone's system, H12, H27, and H30 were also highly utilized on this operating system across both mobile apps. Of these highly violated heuristics, H27 was the one with most issues on the Android platforms.

**Table 2. Heuristic issues per evaluator**

OS	# Evaluator	Nike+		Runkeeper	
		# problems found	# heuristics violated	# problems found	# heuristics violated
iPhone	1	15	8	15	10
	2	38	13	16	9
	3	13	10	7	6
	4	9	6	10	7
	5	28	28	20	20
<b>Sum</b>		<b>103</b>	<b>65</b>	<b>68</b>	<b>52</b>
<b>Average</b>		<b>20.6</b>	<b>13</b>	<b>13.6</b>	<b>10.4</b>
Android	6	13	7	9	4
	7	11	10	11	6
	8	19	15	45	21
	9	79	22	34	10
	10	26	25	23	23
<b>Sum</b>		<b>148</b>	<b>89</b>	<b>122</b>	<b>64</b>
<b>Average</b>		<b>29.6</b>	<b>15.8</b>	<b>24.4</b>	<b>12.8</b>
<b>Total averages</b>		<b>25.1</b>	<b>14.4</b>	<b>19</b>	<b>11.6</b>

Note: aggregate count totals shown in bold.

**Table 3. Most violated heuristics**

Most violated heuristics overall	
H35 (33 violations)	
H34 (30 violations)	
H7 (27 violations)	
H29 (24 violations)	
Most violated heuristics per mobile app	
RunKeeper	Nike+
H27 (25 violations)	H6 (16 violations)
H29 (23 violations)	H8 (21 violations)
H34 (12 violations)	H13 (24 violations)
H35 (15 violations)	H29 (27 violations)
	H30 (22 violations)
	H34 (19 violations)
	H35 (18 violations)
Most violated heuristics per platform	
Android	iPhone
H7 (12 violations)	H7 (15 violations)
H12 (13 violations)	H13 (25 violations)
H27 (22 violations)	H14 (17 violations)
H29 (24 violations)	H17 (18 violations)
H30 (19 violations)	H29 (26 violations)
H34 (14 violations)	H34 (16 violations)
H35 (12 violations)	H35 (21 violations)

**Table 4. Least violated heuristics**

Least Violated Heuristics per app	
RunKeeper	Nike+
H2 (0 violations)	H28 (0 violations)
H16 (0 violations)	H23 (1 violation)
H24 (1 violation)	
Least Violated Heuristics per platform	
Android	iPhone
H11 (1 violation)	H3 (1 violation)
H33 (1 violation)	H21 (2 violations)
	H22 (2 violations)
	H18 (2 violations)

For RunKeeper on both platforms, H27, H29, H34 and H35 were the most frequent violated heuristics. Similarly for Nike+ on both platforms, H29, H30, H34, H35 were four of the most frequent violated heuristics.

The violation of heuristics as well as the diversity of violated heuristics in the list reinforces the importance of user interface design considerations for older adults and indicates the usefulness of a standardized list of heuristics, such as the one this paper introduces.

Conversely, there are several heuristics that were not utilized much, such as H2, H23, H28, H19, H11, and H33, see Table 4. Specifically H19 was underutilized in the evaluations and further issues with this heuristic are recognized in the next section resulting in its removal from the final list of heuristics.

## 4.2 Post- Evaluation Survey Results

The post-evaluation survey was distributed to the ten evaluators immediately following their evaluation submission, see Table 5.

It included eight open-ended questions for supplementary feedback on the aggregated list of heuristics. Besides inquiring the evaluators about the strengths and limitations of the list, the survey asked the evaluators if particular heuristics were (not) useful, difficult to operationalize, missing or did not fit the purpose. Finally, the evaluators were asked to rate the list of heuristics regarding usefulness, clarity, completeness, and appropriateness in four closed questions. The analysis of these four questions is addressed in a different study [30].

When asked about the **strengths** of the list of heuristics in its ability to evaluate mobile apps targeted at older adults, all evaluators indicated at least one strength. Further, two evaluators evidenced that one of the strengths was exactly the fact that this is a list specifically directed at this target user group. Six other evaluators also emphasized aspects along these lines, noting that this list prompted the specific needs of older adults. The remaining evaluators mentioned a number of different aspects: the integrity of the list, the good balance between concreteness and flexibility, and the fact that the list consisted of a good starting point for the evaluator.

When asked about the **limitations** of the list, two evaluators did not indicate any, while the remaining eight evaluators did point out some limitations. While two evaluators said the length of the list might make the list hard to handle, another indicated missing heuristics referring to the existence or absence of visual cues. Another evaluator reported difficulties using H4 and H34, because he stated lacking knowledge on mental models and standards for older adults were. Finally, one evaluator raised the theoretical question of

heuristic evaluation allowing each evaluator to define its own set of heuristics and how this particular list was better than any other one. This study would argue that this is true of any set of heuristics and thus, further analysis is not warranted.

**Table 5. Post-evaluation questions**

1.	What do you think are the <b>strengths</b> of this heuristics list in its ability to evaluate mobile apps targeted at older adults?
2.	What do you think are the <b>limitations</b> of this heuristics list in its ability to evaluate mobile apps targeted at older adults?
3.	Is/are any of the <b>heuristic(s) particularly useful</b> ? Which one(s)? Why?
4.	Is/are any of the <b>heuristic(s) particularly not useful</b> ? Which one(s)? Why?
5.	Is/are any of the <b>heuristic(s) particularly difficult to operationalize</b> ? Which one(s)? Why?
6.	Have you identified <b>problems</b> using this list of heuristics that you would have <b>overlooked</b> otherwise?
7.	Do you feel there were any <b>heuristics that did not fit the purpose</b> of evaluating mobile apps targeted at older adults? If so, which one(s)?
8.	Do you feel there were any <b>heuristics missing</b> in the list you were provided with? If so, what do you feel hasn't been addressed, what heuristic(s) is missing?

The evaluators were also asked about the heuristics they found **particularly useful**. Seven of the evaluators reported on the usefulness of the heuristics on visual design, three on the dexterity heuristics, two on the heuristics related to color; and, one evaluator on the navigation heuristics. Additionally, other heuristics were identified in isolation: H2, H5, H6, H8, H13, H20 H22, and H23.

The reasons why the evaluators indicated a given heuristic or set of heuristics as most useful were because they:

- Were straightforward to use (when referring to navigation and visual design heuristics)
- Were very specific (and necessary) to this particular type of audience, going beyond more straightforward aspects such as keeping a high contrast (~H2, H12, H22)
- Would at least ensure seniors could see the app properly, if all else was ignored (when referring to color, font size, and visual design)
- Were very important to reduce the stress (specially felt by older adults) from making an error or recovering from it (~H7 and H8)
- Allow for a desired level of flexibility while still guiding the user in the completion of tasks (~H14, H15, H17).

When asked if they had found any heuristics to be particularly **not useful**, four of the evaluators answered "No" and one other that most were appropriate. The remaining five evaluators identified H3, H6, H12, H19

(enlisted by two evaluators), H26, H27, H29, and H33 heuristics as not useful, because:

- Not all movement is detrimental (~H3)
- Despite its relevance, the word aesthetic could be misinterpreted by pretty (~H6)
- The use of scroll should be preferred for instance when weighted against a small font size (~H12)
- This feature is enabled by the operating system and the device itself, not the app (~H19, H26)
- It is questionable and discussable (~H27)
- The number of points in the font is irrelevant given the extremely wide variety of resolutions available on phones today (~H29)
- White space could also be light color (~H33).

The evaluators were also asked if any heuristic was **particularly difficult to operationalize**. One evaluator provided no answer, three answered “No”, and the remaining six indicated H2, H4 (indicated twice), H5, H7, H25, H26, H29 (indicated twice), and H34 as problematic, because:

- They are hard to understand (~H2, H5)
- They are hard to judge given the complexity of the concepts of mental models and of consistency (~H4, H34)
- Sometimes it is hard to provide instructions on every screen and the older adult is indeed able to learn after a couple of uses of the system (~H7)
- Most apps do not need this feature and when so this feature is available through the settings of the device itself (~H25, H26)
- They are hard to judge, because of the irrelevance of point measures when looking at possibly different screen resolutions (~H29).

When asked if any **problems** that would have been **overlooked** if it were not for the list of heuristics, eight evaluators said “Yes”, while one said “No”, and another provided no answer. The evaluators that said “Yes” reported that they would have overlooked the use of scroll (two evaluators), the importance of using larger interaction elements, and the need of disabling inactive user interface elements.

When asked if they felt there were any heuristics that did not **fit the purpose** of evaluating mobile apps targeted at older adults, all said “No”.

The evaluators were also asked if the list was **missing any heuristics**, to which four answered “No”, two gave no answer and four said they felt they missed heuristics that accounted for: the possibility of undo and redo; for flexibility; older adults’ memorability difficulties; for the older adults’ expertise (expert vs. novice); for the older adult’s learning curve; for the visibility of features available; and for consistency.

## 5. Discussion

When analyzing the **limitations** of the list of heuristics, two evaluators expressed concerns regarding the length of the list of heuristics. The authors acknowledge that the length of the list might be hard to handle, however, there are many facets to the unique needs of older adults and it is unlikely that they could be sufficiently addressed in a shorter list. Additionally, the results also show that there were other evaluators who would like to supply additional heuristics to the list, indicating a possible need for an even longer list. This study however, took great care in combining and refining the list in order for it to meet the needs of older adults while being as manageable as possible.

In regards to the limitations associated with visual cues provided by one evaluator, the evaluator gave no further explanation on what precisely was missing. As the heuristics list already does provide three heuristics that specifically address visual cues (H1, H5 and H30), the authors chose to not modify these any further.

The comment regarding the limitations related to ‘mental models’ and ‘consistency’ came from one of the two evaluators who self-reported little experience or knowledge on older adults and their different user interface design needs. Since other evaluators did not address this as a limitation, this study argues that even if these are complex concepts, the average heuristic evaluation expert should easily understand these.

In review of the heuristics that were indicated as **not useful**, this study considered modifying the list of heuristics and/or the text within them (Please refer to Table 1, column O and F). In consequence, H19 and H26 are not in the final list of heuristics, as volume and contrast controls are typically dealt with within the operating system of the device itself, and not typically the app. Moreover, both H19 and H26 were underutilized with H19 having a total of only five violations for both apps on both systems (Four for Nike+, two on each platform, and one for RunKeeper on Android) and H26 having a total of eight violations (two violations per app, per platform) which further indicates these two heuristics might not be useful at all.

Regarding H29, the authors consider the evaluators raised a good point when specifying the exact font point therefore the text of this heuristic is revised in the final list, no longer specifying an exact font point. H12 and H3 are kept as they originally were, since the reason presented as to why H12 was not useful was really just a comment and the text of H3 uses avoid, as opposed to, do not use. Regarding H6 and H33, the authors opted not to modify the text of the heuristics given the fact that the concepts of ‘white space’ and ‘aesthetical’ should be well understood by HCI specialists, making their comments seem unfounded.



Additionally, for H6 the term in the heuristic ‘aesthetical’ actually derives from the original Nielsen’s heuristic. Concerning H27, even if to a certain extent the authors agree that this heuristic may be discussable, the authors consider it should be kept, since the inexistence of a heuristics referring to the lack of contrast would severely hinder the interaction for older adult users.

When looking at the answers provided by the evaluators on the heuristics that they felt were **difficult to operationalize**, this study considers that these could explain why H2 and H5 were two of the least utilized heuristics (Table 4). These heuristics could possibly need revision, but as no specific reasons for not understanding the heuristics were provided by the evaluators, the authors did not know how to improve their understanding and no revision took place. H4 and H34, were considered hard to judge given the complexity of the concepts of ‘mental models’ and of ‘consistency’. Similar to H2 and H5, it is understandable that H4 received this feedback and it was underutilized in the evaluation. However, H34 was one of the most highly utilized heuristics during the evaluation. As a result, H4 and H34 were not changed, because the concepts of ‘mental models’ and ‘consistency’ should be well understood by an HCI professional. H29 was also considered hard to judge, but because of the irrelevance of point measures when looking at possibly different screen resolutions. Due to this feedback as well as the one in the question on the usefulness, this heuristic is revised in the final list. H25 and H26 were included because the evaluator argued most apps do not need this feature and this feature is available through the settings of the device itself. As was indicated above, H26 was removed from the final list. However, this study retains H25 as text size setting changes can be difficult to modify in the operation system of devices and do not always impact the text size of the app. The evaluator that included H7 argued that sometimes it is hard to provide instructions on every screen and that the older adult is indeed able to learn after a couple of uses of the system. This means this heuristic should be applied whenever judged appropriate by the evaluator, therefore the authors decided not to revise H7.

The evaluators also provided interesting comments when asked about possible **missing heuristics**. Specifically, one evaluator reported one missing one heuristic that addressed the need for flexibility. However, as an evaluator nicely points out when asked about heuristics that were particularly useful, H17 is the heuristic that address flexibility. Nonetheless, and given the fact that two evaluators felt the need for the possibility of undo and redo and for flexibility, it was decided that H17 needed revision (see Table 1 for

details). The authors considered that the concerns regarding consistency and visibility were already being addressed through H15, H34, H1, H5, and H30, and no additional changes were made. Regarding the concerns on memorability difficulties, older adults’ expertise and learning curves, the authors consider there is a wide scope of needs for older adults, and no one heuristic list could address them all. Therefore no new heuristics were added. But these concerns could provide an opportunity for further future work.

After considering the results of the evaluation and the feedback of the evaluators, a list of 33 heuristics was created (Table 1) being that the original (O column in Table 1) H17 and H29 have had its text revised in the final (F column in Table 1) version. H19 and H26 were excluded, as they were found not useful and hard to operationalize. The remaining heuristics despite their frequency of use did prove to be valuable to assess the quality of the user interfaces being evaluated.

## 6. Conclusions and Future Work

This study advances the area of inclusive design by introducing a list of 33 heuristics for evaluating smartphone apps targeted at older adults. This fills a research gap that will only increase as population ages, health-related smartphone apps utilization proliferates and smartphones become more equipped with built-in sensor and tracking technology.

This list of 33 heuristics is an evolution of an initial list of 35 heuristics created to specifically support age-related limitations, such as the demand for bigger target areas and contrast and the need to avoid the use of scroll. The research presented in this paper put the initial list to test in a heuristic evaluation. Results show that all heuristics in the list were useful in the evaluation of the apps while conducting a heuristic evaluation. Moreover, the analysis of the feedback provided by the evaluators points at the appropriateness of the list, its’ fit for purpose and its’ completeness. However, the feedback from the ten evaluators involved in the study, led to a few changes to the original list that now includes 33 heuristics.

The heuristic evaluation set up to verify the usefulness of the list of heuristics used two health and fitness apps, given the importance of physical exercise for the ageing population, but it is the authors’ conviction that this list can be used for most other types of apps. However, future work should include additional heuristic evaluations with other types of smartphone apps to complement this study. Given the wealth of possibilities, the authors encourage others to use this list and share their results and comments with them.

Still in respect to future work endeavors, the authors consider this research can be complemented with usability testing with older adults as end users, as no inspection method carried out by experts replaces user testing. Another interesting aspect to investigate are platform-related differences and how they actually impact the quality of interaction.

## 7. References

- [1]. Pak, Richard., McLaughlin, Anne. Designing Displays for Older Adults. Human Factors & Aging Series. CRC Press. Boca Raton. (2011).
- [2]. Nielsen Newswire: "Mobile Majority: U.S. Smartphone Ownership Tops 60%,"  
<http://www.nielsen.com/us/en/newswire/2013/mobile-majority--u-s-smartphone-ownership-tops-60-.html>. (2013).
- [3]. Slabodkin, G.: Increased device adoption will make 2013 a banner year for mHealth,  
<http://www.fiercemobilehealthcare.com/story/increased-device-adoption-will-make-2013-banner-year-mhealth/2013-01-02#ixzz2sNyKVocC>. (2013).
- [4] Hollinworth, N. Improving computer interaction for older adults. SIGACCESS Accessibility and Computing. (93), pp. 11–17. (2009).
- [5] Hill, R., Beynon-Davies, P., & Williams, M. D. Information Technology & People, 21(3), pp.244–266. doi:10.1108/09593840810896019) (2008).
- [6] Inostroza, R.; Rusu, C.; Roncagliolo, S.; Jimenez, C.; Rusu, V., "Usability Heuristics for Touchscreen-based Mobile Devices," Information Technology: New Generations (ITNG), 2012 Ninth International Conference, pp.662,667, 16-18 (April 2012)
- [7] Fiona Fui-Hoon Nah , Keng Siau , Hong Sheng, The value of mobile applications: a utility company study, Communications of the ACM, v.48 n.2, p.85-90. (February 2005).
- [8]. Lazich, R.S. (ed.): Top Operating Systems for Smartphones Worldwide, 2012 and 2016. Market Share Reporter (2013).
- [9]. Czaja, S.J., Charness, N., Fisk, A.D., Hertzog, C., Nair, S.N., Rogers, W.A., Sharit, J.: Factors Predicting the Use of Technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (create). Psychol. Aging. 21, pp.333–352 (2006).
- [10]. Chisnell, D.E., Redish, J.C.G., Lee, A.M.Y.: New Heuristics for Understanding Older Adults as Web Users. Tech. Commun. 53, pp.39–59 (2006).
- [11]. Institutes, N., & Health, O. F. (n.d.). Making Your Website Senior Friendly Tips from the National Institute on Aging and the National Library of Medicine Key Tips for Making Your Website Organizing Web Information for Older Adults.  
<http://www.nia.nih.gov/health/publication/making-your-website-senior-friendly>. (2013).
- [12]. Kurniawan, S., Zaphiris, P.: Research-derived Web Design Guidelines for Older People. Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility. ACM Press, Baltimore, MD, USA. pp. 129–135. (2005).
- [13]. Leitao, R., Silva, P.A.: Target and Spacing Sizes for Smartphone User Interfaces for Older Adults: Design Patterns Based on an Evaluation with Users. Conference on Pattern Languages of Programs. pp. 19–21 (2012).
- [14]. Merz, C.N., Forrester, J.S.: The Secondary Prevention of Coronary Heart Disease. Am. J. Med. 102, pp.573–580. (1997).
- [15]. Jin, Z.X., Plocher, T., Kiff, L.: Touch Screen User Interfaces for Older Adults: Button Size and Spacing. In: Stephanidis, C. (ed.) Universal Access in Human Computer Interaction. Coping with Diversity. pp. 933–941. ,Springer-Verlag, Berlin, Heidelberg (2007).
- [16]. Calak, P.: Smartphone Evaluation Heuristics for Older Adults, <https://atrium.lib.uoguelph.ca/xmlui/handle/10214/5610>, (2013).
- [17]. Pinelle, D., Wong, N.: Heuristic Evaluation for Games. Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems. p. 1453. ACM Press, New York, New York, USA (2008).
- [18]. Desurvire, H., Caplan, M., Toth, J.: Using Heuristics to Evaluate the Playability of Games. CHI '04 Extended Abstracts on Human Factors in Computing Systems. pp. 1509–1512. ACM Press, New York, New York, USA (2004).
- [19]. Wilson, C.: User Interface Inspection Methods: A User-Centered Design Method. Morgan Kaufmann Publishers, Waltham, MA (2013).
- [20]. Alan, D., Finlay, J.E., Abdowd, G.D., Beale, R.: Human-Computer Interaction. Prentice Hall, England (2003).
- [21]. Nielsen, J., Molich, R.: Heuristic Evaluation of User Interfaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 249–256. ACM, Seattle, Washington, USA (1990).
- [22]. Nielsen, J.: Usability Engineering. Morgan Kaufmann (1993).
- [23]. Nielsen, J.: 10 Usability Heuristics for User Interface Design, <http://www.nngroup.com/articles/ten-usability-heuristics>.
- [24] Zaphiris, P., Kurniawan, S., & Ellis, R. D. Older people and mobile phones: A multi-method investigation. International Journal of Human-Computer Studies, 66, pp.889-901. (2008).
- [25] Czaja, S. J., & Sharit, J. Age Differences in Attitudes Toward Computers. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences,53B, pp329-340. (1998).
- [26] Coursaris, C., & Kim, D. A meta-analytical review of empirical mobile usability studies. Journal of Usability Studies. 6(3), pp. 117–171. (2011).
- [27] U.S. Preventive Services Task Force: Guide to Clinical Preventive Services.. Williams and Wilkens, Baltimore, MD, USA (1996).
- [28] Silva, P., Holden, K. Nii, A. Smartphones, smart seniors, but not-so-smart apps. HCII conference proceedings. (2014).
- [29] Aitken, M., Gauntlett, C.: Patient Apps for Improved Healthcare, From Novelty to Mainstream. pp.1-65. IMS Institute for Healthcare Informatics (2013).
- [30] Silva, P., Jordan, P, Holden, K.. Something Old, Something New, Something Borrowed: gathering feedback of experts while performing heuristic evaluation with heuristics targeted at older adults. ACE 2014 Workshop on Designing Systems for Health and Entertainment: what are we missing? (2014).