

Acc-MobileCheck: a Checklist for Usability and Accessibility Evaluation of Mobile Applications

Débora Maria Barroso Paiva

Universidade Federal de Mato Grosso do Sul
Campo Grande, MS, Brazil, 79070-900,
debora.paiva@ufms.br

and

Marisa Helena da Silva Batista

Universidade de São Paulo
São Carlos, SP, Brazil, 13566-590
mahh.isa@gmail.com

and

Luciana Aparecida Martinez Zaina

Universidade Federal de São Carlos
Sorocaba, SP, Brazil, 18052-780
lzaina@ufscar.br

and

Renata Pontin de Mattos Fortes

Universidade de São Paulo
São Carlos, SP, Brazil, 13566-590
renata@icmc.usp.br

Abstract

Mobile devices have gained more attention from the society that is using them increasingly for a variety of purposes. For complete insertion of the population in this constant digital evolution it is fundamental that mobile applications also offer access to different user profiles, regardless of their disabilities or limitations. Considering quality, productivity, and speed of application creation, there is a wide range of good development practices and evaluations. However, methods that involve usability and accessibility are still developing. The purpose of this article is to present the Acc-MobileCheck, which is a checklist of accessibility and usability for mobile devices apps, based on good software development practices and guided by Design Patterns. It aims to address difficulties that can be faced by people with hearing, visual, intellectual, or mobility impairment. Five experts and three developers of mobile apps had evaluated the Acc-MobileCheck. The conclusive results show that the checklist is usable and includes essential issues for the evaluation of accessibility and usability. The data obtained allowed a restructuring of the evaluation method developed, and the positive comments about the checklist demonstrate its adequacy to attend the demand.

Keywords: Accessibility, evaluation, mobile devices, usability.

1 Introduction

The number of people connected to the internet is continuously rising; according to the Brazilian Institute of Geography and Statistics (IBGE) [1], the number of internet users in Brazil has practically quintupled in 9 years. We can observe that this scenario is prevalent worldwide, exhibiting a constant growth in the use of the internet, stimulated by the use of mobile devices [2].

By encompassing different elements, such as people, procedures, and technologies, solutions through Human-Computer Interaction (HCI) seek to face problems in the real world and society, following the changing social demands. The advent of new alternatives for accessing information, including mobile devices, requires software designers and developers to be aware of and be prepared to offer technological means and resources to guarantee the same opportunities for people, regardless of their skills or deficiencies. Therefore, the software interface must be developed based on usability and accessibility criteria aiming at a good interaction by the final user [3, 4].

According to the ISO/IEC 25000 Standard [5], usability encompasses several sub-characteristics, for example, apprehensibility, user errors preventing, operability, interface aesthetics, and accessibility. The Standard establishes that accessibility is the extent to which a product can be used effectively, efficiently, and satisfactorily by people from a population with the broadest range of features and capabilities for achieving a specified goal in a context-specific. These two concepts are interdependent, however, there is a lack of analysis about the relationship between accessibility and usability explicitly, both concerning the web and other computational systems [6, 7], such as mobile devices.

Accessibility issues are essential to ensure and improve access for all citizens, especially people with disabilities. According to the IBGE [1], 23.9% of the Brazilian population are people with disabilities. It stated that from the age of 60, it occurs in more than half of the population. Oliveira [8] analyzed the relationship between age and the occurrence of some types of disabilities in Brazil. As reported by the author, more than 45% of the population over 55 years of age has some disability: auditory, visual, motor, cognitive, or multiple impairments. Worldwide, the WHO Press [9] highlights that around 650 million people live with a disability, which is equivalent to approximately 10% of the world population. This scenario already demonstrates the current need for mobile applications to be accessible, to ensure the inclusion of these users, and to maintain good usability for all user profiles. Therefore, our purpose is to present the Acc-MobileCheck, a checklist for evaluating accessibility and usability for apps of mobile devices based on good software development practices and guided by Design Patterns. The main contribution of this work and its differential is the Acc-MobileCheck, composed of 47 verification items and consisting of questions that look for evidence of problems of comprehension, operation, perception, and adaptation in mobile apps. It provides additional information (in relation to traditional models, such as WCAG) that supports the evaluation of usability and accessibility of mobile applications, using the well-established design patterns structure. Thus, each Acc-MobileCheck verification item is described considering the elements “Context”, “Example”, “Recommendation”, etc.

This article is presented as follows: Section 2 addresses fundamental concepts used in the research and related works; Section 3 presents the Acc-MobileCheck; Section 4 describes the Acc-MobileCheck evaluation; Section 5 presents the results obtained; and in Section 6 final considerations are made.

2 Background and Related Work

The quality of interactive software requires that evaluations be carried out during its development, emphasizing that it is more difficult and costly to fix software after implementing all of its components. These evaluations should consider quality criteria of the use, such as accessibility and usability. Furthermore, it should analyze not only from the perspective of developers but also from the final user viewpoint [10].

In this research, we additionally considered the *Design Patterns* and the *Material Design* (which address best practices for software development for mobile devices) concepts. All these elements were studied in detail, aiming to get an overview of the practical context, as presented in the following sub-sections.

2.1 Usability and accessibility assessments

There are several ways to carry out the evaluation of an interface. A useful resource during the development process is conducting automatic tests [11]. These tests are performed by an evaluator who uses a tool as a software program. That tool checks a defined set of criteria quality characteristics and generates a report with problems encountered (failures or errors due to non-compliance criteria) [10]. According to Ivory et al. [12], the use of automatic tests helps in the evaluation process, making it less time consuming, more efficient, and with reduced cost.

Similarly to accessibility, there are tools to support automatic usability evaluation on the web. However, most of them do not analyze the user's perception, cognition, and motor ability, being limited to validating the HTML syntax [13]. Thus, automatic tests are adequate and efficient to maintain the quality of the code; however, they are not appropriate to evaluate the quality of use [11, 10].

Diverse methods in the literature aim to assess the quality of use. According to Barbosa et al. [10], there are different methods for evaluating user interfaces: the investigation, observation of use, and inspection. The investigation methods usually consist of procedures supported by questionnaires and consider conducting interviews, focus groups, and field studies. In the observation of use methods, the evaluator is responsible for observing users (in their usual context or their laboratory), meanwhile collecting data on the performance.

Inspection methods characterize for assigning to evaluators the task of identifying problems that users may have when interacting with the system. In general, the inspection methods tend to be faster and less expensive to apply than other methods, as they do not always need to recruit users for opinion collection or observation sessions. During an inspection, the evaluators examine discrepancies between the interface and guidelines, in an attempt to predict the user experiences that users may come to encounter when using the interface under evaluation [14, 15]. Checklists are one type of inspection methods.

2.2 Design Patterns

Design Patterns have been adopted in the HCI domain and are known as *User Interface Design Patterns* (UIDP). There are different opinions about the necessary elements to compose the UIDP structure. An UIDP is a general and repeatable solution for a frequent usability problem in the *design* of an interface and consists of the following elements [16]: (i) name of the pattern; (ii) problem definition; (iii) solution; and (iv) consequences. Folmer [16] recommends including "Examples" of using the pattern and Van Welie and Van der Veer [17] proposed the inclusion of the topic "Reasoning".

Nudelman groups 77 UIDP for mobile applications into 14 categories, which discriminate between location and function [18]. For the present study, 12 *mobile* UIDP by Nudelman were chosen. The six categories in which the 12 selected UIDP are distributed are: (i) **Home Screen**, which aggregates presentation patterns of the application's functions and content; (ii) **Search**, which groups standards related to search, insertion, and suggestion of data by text and voice; (iii) **Ordering and Filter**, which deal with organization and filter information customizable by users; (iv) **Input Data**, which groups standards related to the insertion of data and files by users, as well as possible user interactions during the selection of these files; (v) **Forms**, which include standards for data confirmation and validation functions and (vi) **Navigation**, which deals with means of viewing application content.

2.3 Material Design

In July 2014 Google created a visual language that synthesizes the classic principles of good design called *Material Design*¹. It is "a comprehensive guide to visual design, of movement and interaction for different platforms and devices". It is also compatible with components and features available on Android, from version 5.0. The *Material Design* pattern provides a set of guidelines and tools to assist the developer, based on good user interface design. The guidelines standardize the entire graphic part of the application, like icons, colors, and animations.

The standardization in the *design* of the interface in the *Material Design* aims to allow improvements in the user experience (UX - *User eXperience*), since it increases the apprehensiveness and recognizability of the software. This is due to a set of rules for colors and common location of buttons and icons that have similar functionality, which unifies and standardizes their visual characteristics and, thus, the user does not need to relearn or guess the features in the application.

One of the goals of Material Design was to create a language to combine design principles with new resources generated by the technological advancement of devices and software, as well as allowing a unified experience across different platforms. To encompass the different requirements of these platforms and the variations, the user interface components of the language were organized in *Material Components* specific to Android, iOS, web, etc.

For example, the Hierarchy Guideline² establishes that designers should to place important actions at the top or bottom of the screen (reachable with shortcuts) and should place related items of a similar hierarchy next to each other, as presented in Figure 1.

¹<https://material.io/design>

²<https://material.io/design/usability/accessibility.html#hierarchy>

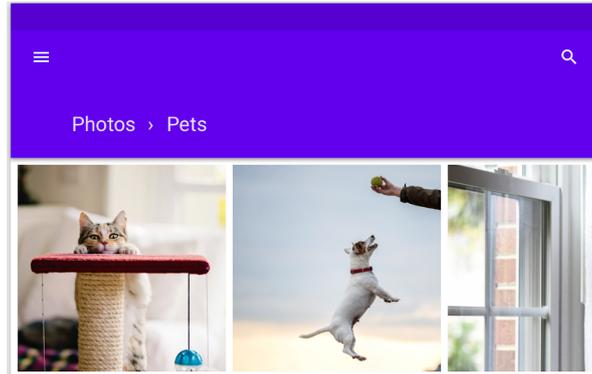


Figure 1: Example of material design guideline (<https://material.io/design/usability/accessibility.html#hierarchy>)

2.4 Related Work

Piccolo et al. [19] conducted research on visual accessibility solutions for mobile devices, through interviews and observations of blind and low vision users using mobile devices with screen readers. Based on the data, a set of nine design guidelines for the development of mobile applications was defined considering the needs of this target audience. This study contributed with details about the real needs of users using assistive technologies on mobile devices, which guided some of the verification items of Acc-MobileCheck.

Investigating how visually impaired people interact with mobile devices, Kim et al. [20] also proposed a set of implications (in the form of heuristics) to guide the development of accessible mobile applications. Usability tests were carried out on a native mobile application with users at different levels of visual loss. Among the usability problems found, five relevant implications were identified in the design of Assistive Technology resource interfaces for people with visual impairments: satisfying affect from auditory sense, understanding the process of visual-spatial information, providing consistent and simply-structured UI layout, increasing configurable settings, and improving speech performance of screen reader. These heuristics provided us with specific details that were used in the Acc-MobileCheck proposal, mainly about the UI layout and configurable settings.

Nathan et al [21] developed a model for evaluating mobile applications for people with hearing impairments. Initially, interviews were conducted with these users while using a mobile application to collect information about their needs and challenges. The results obtained with users were compared to requirements mentioned in studies on hearing impairment found in a systematic review. Thus, a quality model with six dimensions of usability (efficiency, satisfaction, learnability, effectiveness, understandability, and accessibility) was defined for application evaluations for people with hearing impairment. This work was relevant for highlighting the differences in the types of problems that are treated in Acc-MobileCheck, especially when the verification items question the barriers of people with hearing impairment.

A recent work [22] reports on usability tests with 10 participants, 6 blind and 4 users with normal vision, using 4 sites and their respective 4 native mobile applications. The evaluation resulted in a total of 514 problems found, 105 by users with normal vision and 409 by blind users, the most frequent and most serious being found by the last group. The results show critical problems that need to be addressed when developing mobile applications and sites. Thus, the work raised important points that guided the creation of Acc-MobileCheck to check if the questions and definitions elaborated met accessibility problems raised by visually impaired users (alternatives for images, identification of buttons (clickable elements), etc.).

Considering mobile evaluation, Eler et al. [23] introduced the idea of using automated test generation to explore the accessibility of mobile apps. They presented the MATE tool (Mobile Accessibility Testing), which automatically verifies apps by performing different checks for accessibility issues related to visual impairment. Moura et al. [24] provided an API that automatically runs tests, analyzing whether the interfaces of an application conform to accessibility rules and offer tips for improving accessibility. Patil et al. [25] presented a tool for evaluation of application accessibility in Android environment. According to the authors, the Enhanced UI Automator Viewer tool is the more powerful tool to inspect a UI component of mobile application and the color contrast feature. It intends to identify best color combination for foreground and background to make Android applications accessible for users with low vision and color blindness.

The need for an evaluation method that could indicate, in a more objective way, usability and accessibility problems was the main motivation for the elaboration of the Acc-MobileCheck. In particular, the Acc-MobileCheck encompasses the strengths of researches carried out about evaluation, mobile development, UIDP, and includes practical information on accessibility.

Table 1: Alexa's selected websites

ID	Position in Alexa ranking	URL	Functionality
1	1.	Google.com	Search
2	2.	Youtube.com	Entertainment
3	3.	Facebook.com	Social network
4	4.	Baidu.com	Search
5	6.	Amazon.com	Online shopping
6	7.	Wikipedia.org	Information
7	9.	Twitter.com	Social network
8	17.	Linkedin.com	Social network
9	19.	Instagram.com	Social network
10	36.	Netflix.com	Entertainment

3 Acc-MobileCheck

The decision to study the characteristics of UIDP refers to the fact that they are patterns that describe solution for recurring usability problems in the user's interaction with the interface, although allowing different implementations. Thus, by identifying them in the interface, just by comparing them with their definition in the UIDP, it can be verified if the correct implementation was carried out.

In addition, *Material Design* was also studied, as it is a good practice of current mobile design, which brings most of the design recommendations for the interface elements (size, spacing, color, location, touch function, etc.) applicable to different situations in the user interface, defined in a structure similar to a mobile UIDP. The user interacts with the content of an application in different ways, whether on mobile or desktop. Thus, there are differences in the UIDP of each interaction device, even between different operating systems for the same device. Therefore, the analysis of UIDP for all variables it would become impracticable. In this case, we focused on native apps for mobile devices, with Android operating system.

Considering the 77 UIDP proposed by Nudelman [18], we have selected 12 of them. Our choice was based on an in-depth study in which ten selected mobile apps, available on the market, served as a basis for searching the most frequently UIDP used and ordinary for people [26].

From these studies, an evaluation was carried out considering ten most visited websites in the world according to Alexa³. In this tool users can find lists with the visitation ranking of web pages with estimates based on global data. We chose 10 of the 50 most visited websites in the world and we used as inclusion criteria (In) and exclusion criteria (Ex): [In1] the website has relevant content for further analysis; [In2] the website is among the top 50 in the ranking; [Ex1] the website was developed in a non-English language and does not maintain its structure when translated automatically; [Ex2] the website has inappropriate content, or even improper content for further review, such as pornography content; [Ex3] more than three websites in the same functionality category have been selected; and [Ex4] the website is only a version in another language of any website already selected.

Table 1 summarizes the selected websites (URL), their respective positions in the ranking of visitation of Alexa, categories of functionality and ID for later reference in the analysis of the data.

Therefore, we investigated the 10 mobile apps to find out which UIDP the developers used, and the 12 resulting UIDP were:

1. *List of Links* - it is usually found on the first page, serving as a display center for a variety of links and icons for primary functions or usual pages that can be obtained from the app;
2. *Voice Search* - it usually appears as a clickable microphone icon, which when touched causes the device to go into listening mode;
3. *Auto-Complete* - it uses suggestions given by the system, which automatically completes the text typed by the user;
4. *Auto-Suggest* - the user types one or more characters in the search field, the system displays a series of additional suggestions containing one or more combinations corresponding to the keywords entered;
5. *Search From Action Bar* - the user can access the search through a dedicated clickable element in the application action bar that, when pressed, displays saved searches, search refinement options, usual searches, nearby locations, among others;

³<http://www.alexa.com/topsites/global>

6. *Tabs* - it is characterized when the tabs at the top of the page allow users to change pages or apply the usual sorting and filtering options;
7. *Multiple Select* - the user touches one or more items (from a gallery), the page is switched to multiple selection mode enabling mass changes, or also when the user touches one or more values (from a displayed list) and can apply/discard selections;
8. *Textbox with input Mask* - when a field accepts only specific data, such as email or phone number, the system can provide the right type of keyboard to facilitate data entry. Also, depending on the location of the label, the field can optionally show the input mask on the inside;
9. *Inline Error Message* - when an entry error occurs, the system notifies the user which fields need to be corrected. Generally, you can recognize two components of this UIDP: an error indicator in or around the field and a general error message (usually at the top of the page);
10. *Callback Validation* - when a portion of the user's input data needs to be validated on the server, the system detects when the input is complete and issues an asynchronous server call to validate the data, returning with one of two states, OK or Failed;
11. *Cancel/Ok* - in the form's design, the action buttons display Cancel / OK and usually are at the top or bottom of the form. The main action button is on the right and is sometimes longer or implemented with a more saturated color icon. Often, the main action button is disabled at the beginning of the process before a valid value is entered in the text field;
12. *Carousel* - the user sees several photos of the product along a horizontal line. Then the user can slide the line to scroll horizontally to the next set of products. An arrow indicating the direction of movement of the *Carousel* is usually provided as a tip to interact in the desired direction.

Considering our study of good practices for the development of applications for mobile devices, we defined the verification questions (evaluation verification items) considering, as parameters, the problems related to accessibility and usability. Thus, the first version of Acc-MobileCheck was composed of 241 verification items. The checklist has gone through a systematic development process and, currently with 47 verification items, is organized into four types of problems (**TP**), derived from the WCAG 2.0 principles [27].

Based on the initial set of checklist items, aiming to reduce access difficulties for people with some of the four types of disabilities (visual, hearing, motor, and intellectual), we relate each type of problem to the types of disabilities. It is worth noting that the types of problems could be related to more than one type of disability, and vice versa, because we had as a scenario, the context of the use of the 12 UIDP studied. The problems addressed at Acc-MobileCheck were:

- (i) **Comprehension (C)** - related to the user's cognitive difficulties to perform tasks when interacting with the application, such as: to understand what a clickable item or any non-textual element is for; to know where it is currently in the application, among others.
- (ii) **Operation (O)** - related to the user's difficulties to perform operations on tasks, such as: reaching an application resource efficiently or facilitated, pointing at the clickable element; using the application keyboard to enter data; obtaining from the visual, textual, sound or tactile context the current status of a task in progress, among others.
- (iii) **Perception (P)** - related to the user's difficulties to perceive information while performing tasks in the application, such as: identifying in the visual, textual, audible or tactile context notifications/suggestions given by the application, including during assistive technology usage; to identify the visual, textual, sound or tactile result of an interaction selection over an application element, among others.
- (iv) **Adaptation (A)** - related to the user's difficulties to face situations, such as: needs personalized/customized assistance to use the application, has difficulty using the application in only one specific orientation (vertical/horizontal), or different screen sizes.

The 47 Acc-MobileCheck verification items are related to each UIDP whose elements can be checked/questioned regarding accessibility/usability issues, and by type of problem. The total of questions per TP are 16 questions on (**C**) problems (the largest set of Acc-MobileCheck), 15 on (**O**) problems, 12 on (**P**) problems, and 4 on (**A**) problems⁴. There is an example of an Acc-MobileCheck verification item for a Comprehension problem, named as **C5**, in Table 2.

⁴From now on, we refer to these 4 types of problems through their initial letters: COPA.

In the definition of each verification item there is the title/statement of the question, followed by: (i) **Context** in which the question applies; (ii) **Example** that is necessary to check the interface under evaluation; (iii) **Motivation** in terms of accessibility gains in the interface, raising for which type(s) of disability(ies) is appropriate addressing the issue; (iv) **Recommendation** providing more comprehensive information on the scope of the issue; (v) **References** which indicate the documents used to substantiate the question; and (vi) **Answers** with the possible response options, in addition to observations by the evaluator.

<p>[C5] Is the language used in textual elements legible and understandable?</p> <p>Context: The mobile application contains textual content.</p> <p>Example: A text content does not have a clear language or it is not in the same language, or even uses abbreviations, technical terms, jargon.</p> <p>Motivation: To assist users with visual disabilities who use screen readers to get the context of the application and to understand the message of the content. To assist users with difficulties in understanding the meaning of the application language to meet the needs of intellectually disabled users.</p> <p>Recommendation: According to Nielsen’s second Heuristic [3], “the system should speak the user’s language instead of using system-oriented terms”. In the WCAG 2.0 [2] recommendations, the user must be able to read and understand what he/she is reading, regardless of whether he/she is using a screen reader or not. Textual parts with technical terms or distant from the common language of the user impair the understanding of the content of the screen for navigation, understanding for the use of tools, as well as the correction of possible errors reported by the software. Choose common words that are clear and easily understood by beginners and advanced readers. The text must be understandable by anyone, anywhere, regardless of their culture or language [1].</p> <p>References: [1] <i>Writing. Material Design</i>, Google Inc., 2014. [2] <i>Guideline 3.1 Readable</i>. WCAG 2.0, W3C, 2008. [3] Nielsen, J., (1994). <i>Heuristic evaluation. Usability Inspection Methods</i>. New York: John Wiley & Sons, Inc.</p> <p>Answers: [] Yes [] Partially [] No [] Not applicable</p>

Table 2: The C5 - a verification item of Acc-MobileCheck

The use of the checklist Acc-MobileCheck⁵, with its 47 verification items, has been validated by experts and developers (Section 4).

4 Acc-MobileCheck Evaluation

To evaluate the Acc-MobileCheck, experts and developers were invited to experiment with the developed checklist. Five experts (E1, E2, E3, E4, and E5) in accessibility and usability participated in the evaluation, in addition to two mobile application developers with more experience and a developer with little experience (D1, D2, and D3). All 8 evaluators are graduate students of University of São Paulo. Each evaluator performed two different tasks, each one analyzing a different type of problem, to perform the evaluation using the Acc-MobileCheck. Each Acc-MobileCheck type of problem is analyzed for unique verification items. Thus, each evaluator applied only one set of verification items of Acc-MobileCheck.

The experts’ tasks (TE) were pre-defined and distributed among them with the types of problems (TP: (C), (O), (P), (A)). In this way, it was possible to analyze the same task by two evaluators and with different verification items. The distribution of their 4 tasks (TE1 to TE4), with the respective potential types of problems (TP) analyzed by the 5 experts are indicated in Table 3.

The pre-defined developers’ tasks (TD) should be performed on mobile application prototypes, and according to the types of problems (COPA), they were assigned to each evaluator. Similar to the distribution of the tasks to the experts, each developer evaluated a different type of problem, however, each developer task was performed only once. The distribution of tasks (TD1 to TD6) with the respective potential types of problems (TP) analyzed by the developers are shown in Table 4.

We can observe that there was a more proportional distribution in the allocation of TP (C, O, P, A) to experts (2, 3, 3, 2), than in the distribution to developers (3, 1, 1, 1), in which there was a concentration on the TP “Comprehension (C)”. However, adding the numbers of experts and developers who analyzed each TP, the following distribution was obtained: 5 participants evaluated (C), with 16 verification items;

⁵The complete description of Acc-MobileCheck is available at <http://tiny.cc/q2arez> (in Portuguese). Acc-MobileCheck’ online version is available at <http://limesurvey.labes.icmc.usp.br/index.php/783492> (in Portuguese).

Tasks x TP	C	O	P	A
TE1	-	E4	E5	E3
TE2	-	E2	E3	-
TE3	E1	-	-	E4
TE4	E5	E1	E2	-
Number of Experts / TP	2	3	3	2

Table 3: Distribution of 4 tasks and types of problems (TP) to the five experts (E)

Tasks x TP	C	O	P	A
TD1	D1	-	-	-
TD2	-	-	D1	-
TD3	D2	-	-	-
TD4	-	D2	-	-
TD5	D3	-	-	-
TD6	-	-	-	D3
Number of Developers / TP	3	1	1	1

Table 4: Distribution of 6 tasks and types of problems (TP) to the three developers (D)

4 participants evaluated (O), with 15 items; 4 participants evaluated (P), with 12 items, and 3 participants evaluated (A), with 4 items. The evaluations were carried out using all the verification items belonging to the Acc-MobileCheck TPs.

The evaluators who participated in the validation of the Acc-MobileCheck were asked to answer 10 questions regarding the *System Usability Scale* (SUS) questionnaire. SUS is a quantitative analysis method created by Brooke [28] to assess the usability of products, services, hardware, software, websites, and any other type of interface. The analysis considers the usability criteria suggested by ISO 9241-11 [29]: *Effectiveness*: “Can users complete their goals?”; *Efficiency*: “How much effort and resources are needed?”; *Satisfaction*: “Was the experience satisfactory?”

The objective of SUS is to provide a general indication of the subjective usability level of a system, so that it can be used for comparison with its competitors or its other versions [28]. This indication is obtained through a questionnaire with 10 statements, for which the evaluator can choose between 5 levels of agreement, from 1 (“strongly disagree”) to 5 (“fully agree”). Finally, it requires the following calculations:

1. In the answers to odd questions, the rating given by the evaluator will be subtracted from 1;
2. In the answers to even questions, the score will be 5 subtracted from the rating given by the evaluator;
3. All notes converted in the previous steps must be added;
4. The sum value must be multiplied by 2.5, resulting in the final SUS score.

The total validation period, including the application of Acc-MobileCheck and the response to SUS questionnaires lasted 12 consecutive days.

5 Discussion about Evaluation

One of the most common way to use questionnaire data is to report them, describing the distribution of the answers to each question or the most relevant ones [30]. Descriptive statistics must be appropriate. In the case of categorical measures, it is necessary to display the frequency distribution of the responses and, in order to summarize the result, the most frequent response is reported. In ordinal measurement questionnaires, as in the case of measuring usability in systems, the average of the answers is reported to summarize the result. An example is the SUS questionnaire, used to evaluate the Acc-MobileCheck. The data obtained indicate that the lowest SUS average value was assigned by an expert (E1, with 52.5) and the highest value was assigned by a developer (D3, with 100) as shown in Figure 2.

An analysis on SUS was proposed by Bangor et al. [31] in which, according to the evaluated criteria, a system has a medium usability when it has a SUS score greater than 50.9 and a good usability, when

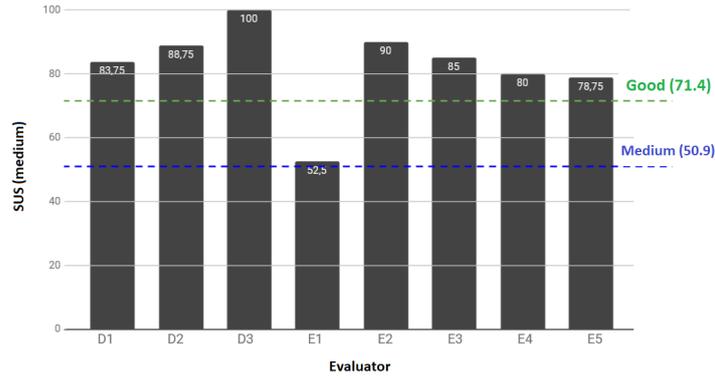


Figure 2: The SUS scores assigned by the 3 developers and 5 experts, and their averages

Evaluator Type x TP	C	O	P	A
Developer	89.2	90	87.5	100
Expert	62.5	75	86.7	81.3
Average (TP)	75.85	82.5	87.1	90.65

Table 5: Averages of problem types (TP) by developers and experts

the SUS score is greater than 71.4. These values, here referred to as cut score, are highlighted in Figure 2. Table 5 summarizes the comparison of the averages of evaluations.

The types of problems (TP) with the worst SUS rating by the evaluator, which were evaluated by the same expert, both scored the same (E1, with 52.5 in (C) and (O)). Thus, this score reduced the average for the evaluations made by experts for (C) and (O), when compared to the evaluations for (P) and (A).

Considering the cut score proposed by Bangor et al. [31], in Figure 3 we can observe that all types of problems scored within the medium usability (above 50.9) and three of them are classified above the score considered good usability (71.4) by both groups of evaluators.

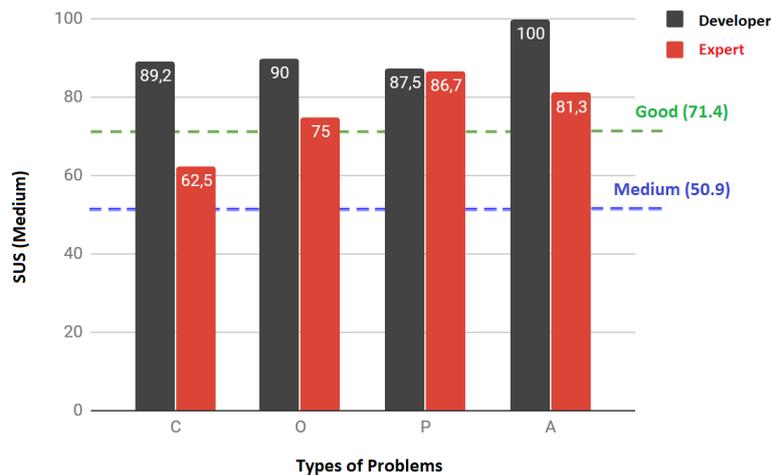


Figure 3: Average SUS score of problem types by type of evaluator

Regarding the scores, as can be seen in Figure 3, the biggest difference between the developer and expert scores was given to the items of Acc-MobileCheck related to the (C) problem, differing in 26.7. The smallest difference was given to the (P) problem, differing by only 0.8. Overall, the average SUS score (considering all the types of problems) by the experts was lower than that of the developers.

As highlighted by Brooke et al. [28], the SUS score is a subjective measure of a system’s usability, used to obtain an overview of its quality in this context. How to obtain each part of this measure depends on the personal perceptions of each evaluator involved, and different results may occur. When considering the average score of a large sample, possibly a very different result would not be decisive. Regardless of the

number of evaluators, it is interesting to analyze the reason for a bad experience with the evaluated interface.

Following, we present comments provided by reviewers about their experience with Acc-MobileCheck.

E1 - "... excessive information. I think it is better to separate the content description of its set of guidelines from the evaluation tool. The interaction options (checklist and text box) were "squeezed" by texts above and below, which are not always read. I would prefer separate screens, one for describing the guidelines and another for checking them.

E2 - "I believe it would be interesting to be able to specify the task that was evaluated in Acc-MobileCheck."

D3 - "The questionnaire in general is simple to answer, with direct questions about the use of the system as a whole. However, the requested tasks do not cover all issues."

Analyzing E1's comment, the reason is that, in fact, for experts in usability and accessibility, the explanations are too much in each verification item can be exhaustive. D2 and E1 evaluators reinforced this argument. D2 is the only non-expert on usability and accessibility among developers and has assessed the same types of problems as E1, assigning different scores. Only D3 highlighted as a problem the existence of issues not used in the task. Therefore, the importance of help texts for evaluators who do not have specialized knowledge on the subject addressed in the verification items is justified.

Among the advantages of a checklist is the ease and low complexity of interpretation of the questions, allowing their application by evaluators with prior knowledge about the recommendations [32]. Acc-MobileCheck is a checklist that aims to assist in assessing usability and accessibility in the process of developing mobile applications. It has simple questions and aid's texts for possible doubts about each verification item so that inexperienced evaluators in the subject can also use it.

In this sense, the observation about the experience with Acc-MobileCheck by evaluators with different levels of knowledge about usability and accessibility shows progress in meeting the objectives of its purpose.

In summary, according to the Bangor et al. criteria [31], 87.5% of the evaluators (Figure 3) mentioned that the checklist has good usability; the checks on all types of problems, on average between evaluators, were also indicated with good usability (Table 5), as well as the general average. Although the sample space is small, composed of 8 evaluators, the evaluations show signs of a satisfactory and positive result. Finally, the total SUS average obtained among all evaluators was 84, considered as **good usability**, with a standard deviation of approximately 16.7.

6 Conclusion

Methods for accessibility and usability evaluation of mobile applications have been proposed to meet the current demand for quality development for a diverse and increasing population. However, the need for a more objective method that addresses both usability and accessibility was the main motivation for the Acc-MobileCheck. This checklist has 47 items of verification that consider problems about **comprehension (C)**, **operation (O)**, **perception (P)**, and **adaptation (A)** in interacting of mobile interfaces, in addition to checking issues involving four types of disabilities: visual, hearing, intellectual, and motor. To validate the Acc-MobileCheck, we have evaluated it with five experts and three developers. Through the development of tasks and the answer to the SUS questionnaire, it was possible to obtain a positive result regarding the use of the proposed checklist.

Considering the results and comments obtained from validation with experts and developers, we performed a detailed review of all verification items of Acc-MobileCheck, to elaborate a new version of the checklist. This new version included captions in abbreviations, explanations for technical and specific terms (for example, assistive technology), definitions for the types of disabilities more related to barriers of the problem types (C, O, P, A), etc.

Future work mainly keeps on the evaluation (conducting new case studies) of the new version of the Acc-MobileCheck to observe if the changes made were sufficient, as well as to identify new points for improvement. Besides, more mobile UIDPs should be selected to increase the scope of analyzes. Indeed, future versions of Acc-MobileCheck have potential improvements to reflect it.

Acknowledgment

The authors thank the financial support grant 2015/24525-0, São Paulo Research Foundation (FAPESP).

References

- [1] IBGE, *Pesquisa nacional por amostra de domicílios 2005-2014*. Centro de Documentação e Disseminação de Informações. Fundação Instituto Brasileiro de Geografia e Estatística, 2015. [Online]. Available: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv95753.pdf>
- [2] “Ict facts and figures - the world in 2015,” International Telecommunication Union, jun. 2015. [Online]. Available: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf>
- [3] C. Poore-Pariseau, “Online learning: Designing for all users,” *J. Usability Studies*, vol. 5, no. 4, pp. 15:147–15:156, Aug. 2010. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2019116.2019119>
- [4] G. A. García-Mireles, M. A. Moraga, F. García, and M. Piattini, “The Influence of Process Quality on Product Usability: A Systematic Review,” *CLEI Electronic Journal*, vol. 16, pp. 6 – 6, 08 2013. [Online]. Available: http://www.scielo.edu.uy/scielo.php?script=sci_arttext&pid=S0717-50002013000200006&nrm=iso
- [5] “ISO/IEC 25000 SQuaRE - System and Software Quality Requirements and Evaluation,” International Organization for Standardization, 2014. [Online]. Available: <http://iso25000.com/index.php/en/iso-25000-standards>
- [6] C. Power and H. Petrie, “Accessibility in non-professional web authoring tools: a missed web 2.0 opportunity?” in *Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A)*. ACM, 2007, pp. 116–119.
- [7] A. de Lima Salgado, L. A. Amaral, A. P. Freire, and R. P. M. Fortes, “Usability and UX Practices in Small Enterprises: Lessons from a Survey of the Brazilian Context,” in *Proceedings of the 34th ACM Intl. Conf. on the Design of Communication*, ser. SIGDOC '16. New York, NY, USA: ACM, 2016, pp. 1–9. [Online]. Available: <http://doi.acm.org/10.1145/2987592.2987616>
- [8] L. M. B. Oliveira, “Cartilha do censo 2010—pessoas com deficiência,” *Brasília: Secretaria de Direitos Humanos da Presidência da República (SDH/PR), Secretaria Nacional de Promoção dos Direitos da Pessoa com Deficiência (SNPD), Coordenação-Geral do Sistema de Informações sobre a Pessoa com Deficiência*, 2012.
- [9] WHO and W. B. Group, “World report on disability,” 2011, access: 13/09/2019. [Online]. Available: http://www.who.int/disabilities/world_report/2011/en/
- [10] S. Barbosa and B. Silva, *Interação humano-computador*. Elsevier Brasil, 2010.
- [11] S. Blomkvist, “Towards a model for bridging agile development and user-centered design,” in *Human-centered software engineering-integrating usability in the software development lifecycle*. Springer, 2005, pp. 219–244.
- [12] M. Y. Ivory and M. A. Hearst, “The state of the art in automating usability evaluation of user interfaces,” *ACM Computing Surveys (CSUR)*, vol. 33, no. 4, pp. 470–516, 2001. [Online]. Available: <http://webtango.berkeley.edu/papers/ue-survey/p470-ivory.pdf>
- [13] L. A. do Amaral, R. P. M. Fortes, and T. J. Bittar, “A4U - an Approach to Evaluation Considering Accessibility and Usability Guidelines,” in *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, ser. WebMedia '18. New York, NY, USA: ACM, 2018, pp. 295–298. [Online]. Available: <http://doi.acm.org/10.1145/3243082.3264666>
- [14] F. Paz and J. A. Pow-Sang, “A systematic mapping review of usability evaluation methods for software development process,” *Int. J. of Software Engineering and Its Applications*, vol. 10, no. 1, pp. 165–178, 2016. [Online]. Available: http://www.sersc.org/journals/IJSEIA/vol10_no1_2016/16.pdf
- [15] L. Rivero, R. Barreto, and T. Conte, “Characterizing Usability Inspection Methods through the Analysis of a Systematic Mapping Study Extension,” *CLEI Electronic Journal*, vol. 16, pp. 12 – 12, 04 2013. [Online]. Available: http://www.scielo.edu.uy/scielo.php?script=sci_arttext&pid=S0717-50002013000100012&nrm=iso
- [16] E. Folmer, *Interaction Design Patterns*. The Interaction Design Foundation, 2015, ch. 41. [Online]. Available: <https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/interaction-design-patterns>

- [17] M. Van Welie and G. C. Van der Veer, "Pattern languages in interaction design: Structure and organization," in *Proceedings of interact*, vol. 3, Set. 2003, pp. 1–5.
- [18] G. Nudelman, *Android design patterns: interaction design solutions for developers*. John Wiley & Sons, 2013.
- [19] L. S. G. Piccolo, E. M. De Menezes, and B. De Campos Buccolo, "Developing an accessible interaction model for touch screen mobile devices: preliminary results," in *Proceedings of the 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American Conference on HCI*. Brazilian Computer Society, 2011, pp. 222–226.
- [20] H. K. Kim, S. H. Han, J. Park, and J. Park, "The interaction experiences of visually impaired people with assistive technology: A case study of smartphones," *International Journal of Industrial Ergonomics*, vol. 55, pp. 22–33, 2016.
- [21] S. S. Nathan, A. Hussain, N. L. Hashim, and M. A. Omar, "Dimensions for hearing-impaired mobile application usability model," in *AIP Conference Proceedings*, vol. 1891, no. 1. AIP Publishing, 2017, p. 020108.
- [22] M. C. N. Carvalho, F. S. Dias, A. G. S. Reis, and A. P. Freire, "Accessibility and usability problems encountered on websites and applications in mobile devices by blind and normal-vision users," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. ACM, 2018, pp. 2022–2029.
- [23] M. M. Eler, J. M. Rojas, Y. Ge, and G. Fraser, "Automated Accessibility Testing of Mobile Apps," in *2018 IEEE 11th International Conf. on Software Testing, Verification and Validation (ICST)*. IEEE, 2018, pp. 116–126.
- [24] C. J. M. Moura, S. de Oliveira, K. A. C. Faria, and E. N. de Andrade Freitas, "A New API for Android Accessibility Testing," in *2017 International Conf. on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2017, pp. 594–598.
- [25] N. Patil, D. Bhole, and P. Shete, "Enhanced UI Automator Viewer with Improved Android Accessibility Evaluation Features," in *2016 International Conf. on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*. IEEE, 2016, pp. 977–983.
- [26] M. H. d. S. Batista, "Uma Abordagem para Verificação de Acessibilidade e Usabilidade em Aplicativos Móveis," Master's thesis, Universidade de São Paulo, São Carlos, SP, Brazil, 2018. [Online]. Available: <https://www.teses.usp.br/teses/disponiveis/55/55134/tde-07012019-091622/pt-br.php>
- [27] "Mobile web initiative (mwi)," World Wide Web Consortium Brasil, 2009. [Online]. Available: <https://www.w3.org/Mobile/About>
- [28] J. Brooke *et al.*, *SUS - A quick and dirty usability scale*. Londres: Taylor & Francis Ltd., 1996, ch. 21, pp. 189–194.
- [29] "ISO 9241-11: 2018 Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts," International Organization for Standardization, 2018.
- [30] C. Hestomo, E. K. Budiardjo, and A. Ferdinansyah, "Quality Function Deployment Analysis in Selecting Software Development Methodology: Case Study of ABC-CORP," in *Proceedings of the 2nd Intl Conf. on Software Engineering and Information Management*, ser. ICSIM 2019. New York, NY, USA: ACM, 2019, pp. 63–68. [Online]. Available: <https://doi.org/10.1145/3305160.3305202>
- [31] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009. [Online]. Available: http://uxpajournal.org/wp-content/uploads/pdf/JUS_Bangor_May2009.pdf
- [32] M. C. Camargo, R. M. Barros, and V. T. O. Barros, "Visual Design Checklist for Graphical User Interface (GUI) Evaluation," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, ser. SAC 18. New York, NY, USA: ACM, 2018, pp. 670–672. [Online]. Available: <https://doi.org/10.1145/3167132.3167391>