



Evaluation of the Accessibility of Mobile Applications: Current Approaches and Challenges

Maria Matos , Letícia Seixas Pereira ^(✉) , and Carlos Duarte 

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
lspereira@ciencias.ulisboa.pt

Abstract. The World Health Organization estimates that around 1.3 billion people worldwide are affected by some form of impairment. Ensuring accessibility of mobile applications has become imperative to meet the requirements of this group. Mobile accessibility involves the process of making mobile applications more inclusive for individuals with disabilities using mobile devices. This paper investigates the strong need for effective methodologies to evaluate the accessibility of mobile applications, particularly with the recent implementation of the Web Accessibility Directive in Europe. By reviewing existing research, analyzing monitoring reports from European Union Member States, and conducting interviews with mobile accessibility evaluators, this study identifies limitations in current evaluation methodologies and provides some insights into the challenges faced by evaluators. The findings reveal that there is no universally adopted methodology for evaluating the accessibility of mobile applications. Evaluators either adapt existing methodologies designed for web accessibility or develop their own approaches. The absence of specific guidelines for mobile accessibility poses a challenge, with evaluators often having to interpret web-focused documents. The lack of comprehensive automated tools for mobile accessibility evaluation further complicates the process. Through this research, it is anticipated that improvements in mobile application accessibility evaluations will contribute to creating more inclusive mobile experiences for all individuals.

Keywords: Accessibility · Mobile Accessibility Evaluation · Mobile Accessibility Monitoring · Accessibility Evaluation Methodologies · Mobile Applications

1 Introduction

According to the World Health Organization, approximately 1.3 billion people worldwide are affected by some sort of impairment [24]. Developers must consider these individuals' particular requirements and build accessible solutions. Accessibility refers to everyone's right to experience the environments, products, services, and information that come with living in society, regardless of

their sensorial, motor or cognitive skills, or social and cultural backgrounds. To be considered accessible, an information system should not include barriers that prohibit users, regardless of disability, from using it [27].

The term “mobile accessibility” refers to the process of making websites and applications more accessible to individuals with disabilities who use mobile phones and other mobile devices [14]. Smartphones, as a technological resource, are extremely beneficial to people with disabilities [27], affording them some independence. However, these devices bring with them new challenges [27]. Mobile devices are smaller than traditional desktops or laptops devices, which makes them challenging in a variety of ways, like tapping the right target, as discussed in Gonçalves et al. [23]. The vast number of applications and tasks, along with the small display size, makes them cognitively demanding [13]. Mobile devices and their applications have been evolving at a fast pace, driven by huge corporations. The lack of convergence and standards also marks this industry driven progress. Thus, impacting the treatment of accessibility [13].

The need to ensure accessible mobile applications is a pertinent, current issue, especially in Europe, where the Web Accessibility Directive (WAD) [6] has been recently implemented. This directive requires regular monitoring of public sector websites and mobile applications, demonstrated with periodical reports of accessibility evaluations. These reports [9] have revealed different methodologies and tools used by Member States to evaluate the accessibility of mobile applications, as well as challenges faced by monitoring bodies.

Although researchers have proposed evaluation methodologies, these have certain limitations. For instance, most of these methodologies do not provide complete information to conduct a thorough accessibility evaluation of mobile applications [3,27], and they are not able to identify all types of accessibility barriers [1,22]. To address these issues, an effective accessibility evaluation methodology for mobile applications is necessary. To obtain a more comprehensive understanding of the challenges faced by mobile application evaluators and address some of the issues of existing evaluation methodologies, we combined a review of the research on the subject with a thorough examination of the reports of the first period of monitoring from European Union (EU) Member States. Additionally, we conducted interviews with five mobile accessibility evaluators to understand the methods, tools, and challenges faced during evaluations.

This article is organized as follows: in Sect. 2, we discuss accessibility guidelines applicable to mobile applications, how accessibility is evaluated, methodologies to evaluate the accessibility of mobile applications, and existing tools to help in the evaluation; in Sect. 3, we analyze the monitoring reports provided by the EU Member States, the procedures followed in the monitoring activities, and the results from the accessibility evaluations; in Sect. 4, we report on the findings from a set of semi-structured interviews with mobile accessibility evaluators; in Sect. 5, we discuss all the findings in the context of existing guidelines, past evaluation studies and available evaluation methodologies; and finally in Sect. 6 we conclude the article.

2 Related Work

This research is related to prior work on accessibility guidelines, accessibility evaluation, methodologies and tools for accessibility evaluation.

2.1 Accessibility Guidelines

The W3C Web Accessibility Initiative (WAI) develops guidelines and resources to help make the Web accessible to all types of disabilities [19]. Among them, the Web Content Accessibility Guidelines (WCAG) [30] stand out, with versions that have evolved over the years. The WCAG 2.0 provides little guidance on creating accessible content for mobile devices. To fill this void, the Mobile Web Best Practices (MWBP) [26] were released. MWBP is a document that defines rules to make websites more usable in a mobile device. Similar to the WCAG, the MWBP defines a set of checkpoints that should be considered to guarantee that online content is suitable for access from mobile devices.

Another effort is the Guidance on Applying WCAG 2.0 to Non-Web Information and Communications Technologies (WCAG2ICT) document [16]. This document provides guidance on how to apply the concepts, guidelines, and success criteria of WCAG 2.0 to non-web documents and software. It was developed to give interpretations of how to use WCAG 2.0 in various contexts. WCAG2ICT addresses a broad variety of issues, however, it also acknowledges that it is unable to meet the needs of all individuals with disabilities. Because WCAG 2.0 was created for the web, addressing accessibility for non-web documents and software may need additional steps beyond those specified in the document [16].

Shortly after the MWBP was released, an update to the WCAG was also released. WCAG 2.1 has one new guideline and 17 new success criteria to address mobile accessibility, people with low vision, and people with cognitive and learning disabilities [17].

There is now a public draft of the WCAG 2.2 with 9 new success criteria. These new success criteria enhance users' experiences by ensuring easy access to support resources and allowing for a wider range of assistive technologies. Related to mobile, there are two new success criteria: one about movements to facilitate drag and drop actions, and another about target size to prevent people from clicking on the wrong button due to lack of space between buttons [15].

Besides WCAG, there are other guidelines that can be applied to the mobile context. The User Agent Accessibility Guidelines (UAAG) 2.0 [29] are intended for user agent developers working on desktop or mobile operating systems. A user agent that follows UAAG 2.0 improves accessibility through its own user interface, options for rendering and interacting with content, and the ability to communicate with other technologies, including assistive technologies. Its support documentation contains numerous mobile examples. The Authoring Tool Accessibility Guidelines (ATAG) 2.0 [28] provide guidelines for developers of authoring tools for desktop or mobile operating systems. An authoring tool that adheres to ATAG 2.0 will be more accessible to authors with disabilities while also enabling, supporting, and promoting the creation of more accessible online

content by all authors. The support document to implement ATAG 2.0 contains multiple examples of mobile authoring tools [31].

2.2 Accessibility Evaluation

An accessibility evaluation aims to verify how well websites or mobile applications can be used by as many people as possible [25]. There are different methods for conducting an accessibility evaluation. One of them is automated testing, a starting point to evaluate the accessibility of websites or mobile applications with an automatic tool. This tool detects accessibility problems in the code of each page or screen. Usually, it follows a checklist of points that should be verified to ensure compliance [2, 22]. This testing technique is not capable of finding every problem in the website or mobile application being tested, so these need further testing. On the other hand, it is a fast method, saving time and effort [25].

There is also manual testing. In this, an inspection of the website or mobile application is done by an expert to check for issues that may cause problems for users with disabilities. Experts have to interpret the guidelines being used and determine if they are fulfilled by the mobile application [2, 22]. This is the method that allows evaluators to find the largest amount of errors (ideally all, depending on the expertise of the evaluator), especially the ones that cannot be programmatically detected, although it's a more time consuming method and prone to human error [25].

Another testing method that can be used is user testing. User testing should involve users with disabilities. It allows us to have a better understanding of the users' perspectives, including any issues they may encounter using or not using assistive technologies while navigating the website or mobile application. It also helps us understand if the accessibility criteria address all the issues that people with disabilities may face [25].

Combining these different methods, we can obtain different methodologies for evaluation.

2.3 Methodologies Used for Accessibility Evaluation

A review of the most relevant methodologies for evaluating web and mobile applications was conducted.

European Standard (EN) Methodology. The Web Accessibility Directive [6] is a work of the European Commission to ensure an inclusive Europe, accessible to all. The goal is to enable people with disabilities to better use websites and mobile applications of public services, by ensuring these meet all the requirements of the European Standard EN 301 549 [5]. In the monitoring process, member states must use the methodology chosen by the Commission to monitor compliance [6]. The methodology defines:

- The frequency with which the monitoring of website and mobile application is carried out;

- The sampling of web pages, their content, and the content of mobile applications;
- Provisions for automated, manual and usability testing;
- The guidelines used for determining compliance;
- A mechanism to assist public sector bodies in correcting any deficiencies discovered.

The WAD also defines two evaluation processes, simplified testing and detailed testing, to perform audits using some of the methods already mentioned [6]. Simplified testing uses automated tools and manual checks to examine a small section of the website or mobile application's pages. Detailed testing takes a more in-depth look at the website or mobile application. This monitoring method tests the platform against the WCAG success criteria, using assistive technology along with manual testing combined or not with automated testing to ensure compliance. To monitor the accessibility of mobile applications, WAD only requires the detailed testing method. It is important to note that, in the mobile context, different operating systems must be taken into account [7].

WCAG-EM. The Website Accessibility Conformance Evaluation Methodology (WCAG-EM) [18] is a methodology to determine how well a website, including web applications and mobile websites, complies with the WCAG. The conformance evaluation procedure is divided into five major steps, each broken out with recommendations, best practices, and guidance for evaluators:

1. Define the scope of the evaluation – in this step, the scope of the evaluation, the evaluation's goal, and the WCAG conformance level are defined.
2. Explore the website – the key web pages, key functionality, types of web content, designs, functionality, and required web technologies are all identified here.
3. Select a representative sample – when it is impossible to evaluate every web page on a website, guidance is given to choose structured and random samples of web pages.
4. Evaluate the selected sample – WCAG success and failure are determined in this phase, website features are checked for accessibility, and evaluation steps are documented.
5. Report the evaluation findings – here are the aggregation and reporting of evaluation findings, the formulation of evaluation statements, and the calculation of overall scores.

Appt-EM. The Appt-EM methodology [11] is based on the WCAG-EM, but instead of being developed for websites, the Appt-EM was developed for mobile applications. Out of the 50 A and AA success criteria from the WCAG 2.1, 6 are not applied in this methodology, 13 have undergone minor adjustments to the notes or definitions, and the other 31 are applied without changes.

The evaluation procedure is also divided into five steps, each broken out into activities to guide the evaluators:

1. Define the scope of the evaluation – in this step, you start by determining the scope of the application by specifying the URL for downloading the application and the version to be tested, and knowing which screens are available to be evaluated; then you determine which compliance level to use; and finally you determine the scope of the hardware and software by defining screen sizes, system settings, operating system version, among other things.
2. Research mobile application – the most frequently used screens, the most critical flows, the most used elements, the techniques used, and other relevant screens are all identified here.
3. Select the sample to be evaluated – determine the screens to be assessed, both the structured sample and the random sample, and since the applications consist mainly of flows, add the flows to the scope.
4. Evaluate the selected sample – classify the screens, evaluate the flows, and compare the structured sample with the random sample.
5. Report the evaluation results – report must meet specific formal requirements of WCAG-EM.

This methodology also provides additional recommendations [11], such as providing sufficient context about the evaluation; for example, by saving screenshots of the application screen when some error occurs – since the application can be updated and the error can be gone. Another recommendation is that reports should be organised by success criteria and screen, making it easier for developers to understand how to correct the errors identified.

The authors also provide a list of the WCAG success criteria [12] with additional content tailored to the mobile context. This additional content includes a description of each success criteria, the impact it may have, a way to check if the requirements are being met, and some ways for developers to solve the problems, considering different operating systems.

Evaluation Methodologies in Academic Literature. In addition to the previous methodologies, there are also some that have been developed and used in studies to evaluate mobile applications. Acosta-Vargas *et al.* [1] explored a combination of automatic testing using an automatic tool with manual testing based on the WCAG 2.1 guidelines. After choosing a mobile application, the methodology goes through six phases:

1. Explore, interact, and navigate through each application.
2. Define activities and tasks for users to do in each application.
3. Define a list of accessibility barriers based on the four principles of WCAG 2.1.
4. Select users based on the barriers established in the previous phase.
5. Run the automatic test and review each guideline manually. The parameters selected to be evaluated by the authors in the automatic test were the number of elements, touch target, text contrast, item label, and image contrast.
6. Record the automatic and manual evaluation data and analyze it.

Using this methodology, the authors were able to determine which accessibility principles were being infringed the most in the applications, such as small touch targets, low text contrast, missing item labels, and low image contrast. The authors highlight the reliance on the expertise of the evaluator as a limitation of the method, but it can be mitigated by replicating the manual process with different people. This paper shows that automatic testing still needs some adaptations and tools to be effective on mobile devices, but there is the option to combine it with manual testing.

Ferreira da Silva *et al.* [27] used an observation method in the context of usage of mobile applications to assess their accessibility for people with visual impairment, based on a protocol tested in a previous study [10]. After choosing one mobile application, the methodology included the following stages: participant selection, equipment definition, task list definition, case study implementation, and results analysis. Difficulties and problems were analyzed and compared with the WCAG 2.0 guidelines. Overall, it was possible to observe that most of the users' difficulties are related to the way information and interface components are presented. Besides revealing some of the problems users with visual impairments face using mobile applications, this study supports the idea that designers and developers have difficulties understanding accessibility guidelines, which does not contribute to better accessibility in mobile applications.

Joshi *et al.* [21] proposes a methodology with four different testing approaches that can be applied throughout the mobile app development lifecycle. Authors argue that simple approaches can be carried out by any member of the development team because they do not require advanced knowledge of accessibility testing. The four testing approaches are: automated test, screen reader test, magnification/zooming test, and switch access and keyboard test. The authors report that adopting this methodology during the development phase can produce the most accessible application without incurring additional costs or time. It is simple to plan and implement without the use of expensive tools, and it can help people with visual, motor, and cognitive challenges avoid accessibility concerns. According to their findings, major issues can be identified and resolved if the appropriate testing procedures are employed at the appropriate development phase. Although these are promising developments, the fact is that mobile applications continue being inaccessible. This is, probably, a consequence of these suggestions not being applied by developers. They are, nonetheless, a set of recommendations that are useful for mobile accessibility evaluation even outside the scope of development.

Acosta-Vargas *et al.* [3] recognize that there is a lack of adequate methods to test mobile applications' accessibility and that has become a challenge for accessibility experts. To address this problem, they proposed a method of manual testing using WCAG 2.1 combined with automatic testing using the Accessibility Scanner [4]. The authors only worked with Android applications in this study. After choosing and installing the mobile applications that are going to be tested, the proposed method follows these steps:

- Install the Accessibility Scanner validator and activate the accessibility test;

- Interact with the mobile applications being tested running the accessibility test;
- Record the results in a Microsoft Excel spreadsheet;
- Analyze the problems identified by the tool;
- Review the barriers identified and propose some improvements.

The authors conclude that the information found can be helpful for designers, developers, and evaluators, giving some recommendations to overcome the most common problems encountered in their evaluation, such as lack of item labels, poor text contrast, or even issues with touch targets. They also suggest that to evaluate mobile applications, the process should use a combination of multiple methods, consider users with different disabilities, the different barriers raised by different disabilities, and the diverse usage scenarios of the mobile application. This study sustains the idea that combining methods is the best option to obtain a more accurate assessment and for providing better information to all parties involved in the creation and development of mobile applications.

Mateus *et al.* [22] presents a study that compares different methodologies used to evaluate mobile applications, using problems encountered only by users and problems encountered by both tools and users. It reveals that automated tools can only encounter a small portion of problems while users find a more significant number of problems. The authors came to the conclusion that, despite the fact that automated evaluation techniques might encounter a small but significant portion of difficulties, they cannot replace user testing, which is still a safe way of detecting accessibility barriers. Tools do not identify all of the issues that users describe, but they can be valuable for developers and evaluators in the early stages. They are an evaluation method that identifies a great number of problems at a lower cost and with greater agility.

2.4 Existing Tools for Accessibility Evaluation

Using an automated tool can help verify criteria in a faster way, but they can have disadvantages, depending on the tool, operating system, and even the task performed. Examples of automatic tools used for testing mobile applications are:

Accessibility Scanner: This tool checks Android apps against a set of rules that identify possible problems for people with disabilities. When opened, it scans the screen of the application and gives suggestions for improving its accessibility. It is a tool that can be used directly in a mobile application by anyone since all functions are user-friendly; it can evaluate any screen element; it allows people to leave suggestions about possible alterations for the developers; and it shows the code that is generating accessibility problems. It tests content labels, touch target size, clickable items, text and image contrast. On the downside, this tool only works for Android, and it does not replace manual testing; it only complements it [4].

Accessibility Inspector: This tool inspects an iOS application, looking for potential optimizations that allow people with disabilities to use the app.

Accessibility Inspector makes it possible to conduct automatic audits on any page; to see all element properties at once; to explore switch control, braille boards, and alternative input devices; to help develop apps more accessible to VoiceOver (Apple’s voice assistant) and other accessibility technologies; and to run the application as if it were running with VoiceOver but using a mouse to see the elements’ labels. On the other hand, this tool is only for iOS, only tests on the source code, and its automatic audits must be manually reviewed due to false positives [20].

Apart from automatic tools, there are other types of tools, called support tools, used to evaluate some specific aspects of mobile applications. These support tools are used to help in the manual testing of applications. The most common support tool used to help evaluate the accessibility of mobile applications is color contrast analyzers, which allow you to determine the contrast between two colors.

For websites, there are many more support tools for other functionalities, for example, a tool that helps to check if there are any issues with the content or functionality of your website after increasing the line, paragraph, letter, and word spacing according to the “text spacing” success criterion of the WCAG or even another tool to examine the HTML markup on a web page for syntax and lexical mistakes. These tools do not exist for mobile.

3 Monitoring Reports

In the monitoring process defined by the European Commission’s methodology [6], member states describe the results in reports. These monitoring reports are reports of accessibility evaluations of websites and mobile applications done by each member state, showing some of the shortcomings and difficulties faced when evaluating the applications [9]. From these reports, it is possible to draw some conclusions about what is missing in order to make evaluations more accurate and complete toward achieving more accessible mobile applications.

3.1 Review of Monitoring Reports

To understand what methods were used by member states to assess the accessibility of their platforms, the problems they have faced, and the conclusions they have reached, the monitoring reports [9] were analyzed.

Table 1 shows the methods used by each country to evaluate websites and mobile applications. This table does not contain the reports from Hungary, whose report was difficult to analyze due to a lack of information and structure, and France and Cyprus, who had not published a report by the time of this analysis.

To evaluate websites, almost every country employed a methodology combining both automated and manual testing for a more efficient and accurate evaluation. To evaluate mobile applications, the most common method is manual testing, with some countries also using automatic testing. In general, user

Table 1. Evaluation Methods used by each country

Country	Evaluation Method for					
	Website			Mobile Applications		
	Automated Testing	Manual Testing	User Testing	Automated Testing	Manual Testing	User Testing
Belgium	✓	✓	—	—	—	—
Bulgaria	✓	✓	✓	✓	✓	—
Czechia	✓	✓	—	✓	✓	—
Denmark	✓	✓	—	—	✓	—
Germany	✓	✓	—	✓	✓	—
Estonia	✓	✓	—	—	✓	—
Ireland	✓	—	—	✓	✓	—
Greece	✓	✓	—	✓	✓	—
Spain	✓	✓	—	—	✓	—
Croatia	✓	✓	—	—	✓	—
Italy	✓	✓	—	✓	✓	—
Latvia	✓	✓	✓	—	✓	✓
Lithuania	✓	✓	✓	—	✓	✓
Luxembourg	✓	✓	—	—	✓	—
Malta	✓	✓	—	✓	✓	—
Netherlands	✓	✓	—	—	✓	—
Austria	✓	✓	—	—	✓	—
Poland	✓	✓	✓	✓	✓	—
Portugal	✓	✓	✓	—	✓	✓
Romania	✓	✓	—	—	✓	—
Slovenia	✓	✓	—	—	✓	—
Slovakia	✓	✓	—	—	✓	—
Finland	✓	✓	—	—	✓	—
Sweden	✓	✓	—	—	✓	—
United Kingdom	✓	✓	—	✓	✓	—

testing is not widely used to assess websites or mobile applications. Still, we can see that there are some countries, although few, that make use of this technique combined with others to ensure that the users' point of view is taken into consideration when evaluating mobile platforms.

Overall, the two evaluation processes defined by the WAD were used to perform the audits. For websites, the evaluation processes mostly used were simplified and detailed testing, while for mobile applications, only detailed testing was used.

When it comes to mobile applications, Table 2 shows that the samples differ slightly between countries. According to the methodology defined by the European Commission, the sample size should be proportional to the size of the country's population, consisting of one application per million inhabitants plus six applications. Therefore, there are larger and smaller samples, some of them too small to draw any significant conclusions. According to the Study support-

Table 2. Number of mobile applications per country (* Countries reporting only the total number of mobile applications).

Country	Android Apps	iOS Apps	Total
Belgium	0	0	0
Bulgaria	13	0	13
Czechia	17	0	17
Denmark	4	4	8
Germany	23	34	57
Estonia	4	3	7
Ireland	2	2	4
Greece	17	17	34
Spain	8	10	18
Croatia	8	0	8
Italy	11	11	22
Latvia	–	–	8*
Lithuania	10	8	18
Luxembourg	1	1	2
Hungary	12	12	24
Malta	2	2	4
Netherlands	4	4	8
Austria	3	2	5
Poland	8	7	15
Portugal	8	8	16
Romania	25	0	25
Slovenia	–	–	3*
Slovakia	3	3	6
Finland	2	2	4
Sweden	0	2	2
United Kingdom	2	2	4

ing the review of the WAD [8], the countries with small samples are Luxembourg, Ireland, and Sweden. There is one country (Belgium) that, at the time of this analysis, had yet to conduct mobile application evaluation. Latvia's and Slovenia's reports did not state how many mobile applications of each operating system were evaluated, only the total number. Based on the number of mobile applications evaluated in the countries that have monitored them, the average is 13 mobile applications evaluated per country.

The reports show the contrast between Android and iOS. Table 2 presents how many Android or iOS applications were evaluated in each country. There are

some that do not evaluate iOS because it is not the most used operating system by people with disabilities in the country, or because the country's language is not supported by the screen reader. There is only one country that does not evaluate Android mobile applications, Sweden, but they do not explain why.

Looking at the reports, it is possible to see that some mobile apps do not have any issues complying with all the criteria. Still, other problems were identified in the monitoring of mobile applications:

- The applications miss some criteria;
- There is a lack of information about the results or conclusions;
- Mobile applications are given little importance compared to websites;
- There is no fixed methodology that can be used by everyone in the same way to draw the best possible conclusions;
- There is a lack of accessibility statements.

On the bright side, some reports show detailed conclusions about the accessibility of the applications monitored, and the user testing allows the users to give their perspective on problems they face daily, including some that people without their disabilities wouldn't think of.

3.2 Problems Identified in the Reports

In the reports of the member states of the European Commission it is possible to see which methods were used in the evaluation process of mobile applications. Starting with automated testing, out of the 24 countries that evaluated mobile applications, only nine of them conducted this type of testing. We can deduce the reasons why this strategy is not widely used based on what was said in these reports:

- Not having coding skills to be able to use a certain tool;
- Not testing all the criteria desired to be tested, since these tools cannot identify all problems listed in the guidelines [22];
- Automatic tools continue to face challenges in accurately identifying false positives and handling dynamic content [8]. This further contributes to the existing limitations of automated tools, as discussed in previous sections. Additionally, the majority of these tools only support one of the two major mobile operating systems (Android or iOS), which adds another layer of complexity to automated testing.

Regarding manual testing, as reported by the countries that conducted this type of evaluation:

- There are success criteria that are difficult to evaluate in mobile applications for lack of support tools or an easier process to do it;
- Manual testing increases the time and cost of operations, which are worsened by the difficulties encountered when interpreting the directive's guidelines [8].

About user testing, we can see in Table 1 that only 3 countries conducted this test. This can be partially explained due to:

- User testing takes work to plan and execute and time to do it all;
- It can be hard to find people with disabilities to participate in these tests, either because of deadlines for submitting results, a lack of funds to be able to compensate the participants or a lack of people available to do so.

4 Interviews with Accessibility Evaluators

To better characterize the issues faced when monitoring and evaluating mobile applications, for the following stages, we established two research questions:

- What are the main problems evaluators face when evaluating the accessibility of mobile applications?
- Why they choose the methodologies and tools they use for evaluations?

To answer the research questions we conducted semi-structured interviews with accessibility evaluators to discuss their experience with accessibility evaluations. These interviews aimed to understand the problems they face when assessing mobile applications, the methodologies and tools they use and why, and what they would change to make this process easier and more reliable.

4.1 Participants

Five accessibility evaluators (Table 3) from four different European countries were interviewed remotely. The evaluators had between 1 to 9 years of experience, with an average of three years. All of them had experience evaluating Android, and only one didn't have with iOS. The only requirements for the choice of participants were that they should be of legal age and have previous experience evaluating mobile applications.

Table 3. Characterization of the evaluators

Evaluator	Country	Years of experience with mobile accessibility	Operating Systems used in accessibility evaluations
P1	Portugal	9 years	Android and iOS
P2	Denmark	3 years	Android and iOS
P3	Norway	1 year	Android and iOS
P4	Portugal	1 year	Android
P5	Netherlands	2 years	Android and iOS

4.2 Procedure

Participants were recruited through the research team's network. An email was sent to them with information about the study and an Informed Consent Form to be signed. A convenient time was arranged for both the researcher and the participant to conduct the interview. Sessions were conducted through Zoom and were recorded with the permission of the participants. One of them did not authorize recording the session, so the researcher took notes.

The interview started with a brief introduction about the study. It was checked if there were any questions regarding the study and if the participant authorized the interview recording. It was explained that the participant would be asked a few questions about their experience evaluating mobile applications and to provide feedback on anything they felt relevant to the study. Next, questions were asked related to the problems faced during mobile applications accessibility evaluation, methodologies used in the evaluations along with tools, their thoughts on the ideal methodology, and what they would change in the evaluation process to make it easier. They were still asked if they would like to add anything. In the end, they were thanked for their participation and time. This procedure was approved by the Ethics Committee of the Faculty of Sciences of the University of Lisbon.

4.3 Data Analysis

To analyse the data from the interviews we conducted a thematic analysis. We started by transcribing and coding the recorded interviews. Notes from the non-recorded interview were also coded. Multiple iterations of the codes and themes were discussed among the researchers until a final consensus was reached. The thematic analysis was used to reach conclusions about the research questions established and to understand common problems, methodologies, and tools.

4.4 Findings

From the five interviews with the evaluators, six themes emerged: "Lack of a common methodology for mobile accessibility assessment", "Automation of methodology processes", "Methodologies are not adapted for mobile applications", "Main accessibility flaws found in mobile applications", "Reporting errors to developers in a perceivable way is not easy", "Lack of information and guides to assess mobile accessibility".

Through the coding it is possible to understand that there is no methodology for evaluating the accessibility of mobile application that is common to all evaluators. Some of them developed their own methodology based on an existing one (P5: "so we are testing 44 criteria and besides that we rewrote the WCAG-EM and did some pointers as well") or from scratch (P2: "we've actually built out our own methodology"), arguing that there should be a common methodology to all European Union countries (P3).

Given the problems of the existing methodologies, some ideas emerge that automating the processes of the methodology (P4: “Ideally it would be all automatic, that’s impossible, but automating as much as possible”) would be positive to decrease time and resources (P3), even if the evaluation process was only semi-automatic, and some specific criteria had to be manually evaluated (P3). Still, there are some who say that the evaluation process they use does not present major problems (P4: “I don’t think we had any major problems in the evaluation process”).

To automate methodology processes one can resort to the use of tools, both automatic and support tools since both help in the evaluation process. However, evaluators agree that there is a lack of tools to evaluate mobile application accessibility. Automatic tools are either non-existent (P3), or they only test some problems and the results may not even be right (P1: “sometimes you find things that turn out to be right”). Still, evaluators say that an automatic tool for everything would be ideal (P2: “I mean anyone would wish that there was an automatic tool that could do anything”), or that, at least, a tool that could detect errors that are not always detected manually (P1: “Detect errors that are not so noticeable in manual validation”). Support tools are either not used or limited in their scope, with the usual mention being the color contrast analyzer (P2: “we use color contrast checkers”). Another idea given by the evaluators is the possible use of Artificial Intelligence (AI) (P2: “I also believe that we can use the power of AI to some extent, but still it will require manual assessments”).

Evaluators reported that there is a lack of guidelines to evaluate mobile accessibility (P3) and that the documents that do exist are geared towards the web (P3) which leads to everyone having to interpret these documents differently (P3). The evaluators feel that there is lack of information (P1: “There is no specific thing for native apps”) and of examples regarding the evaluation of mobile apps (P3). As much as they may use a certain methodology, they don’t know if it is the right one (P1: “I can use my methodology all right, but for what it’s worth, there is no standard”) because there is no concrete document for mobile applications (P1: “what applies to mobile? There is no specific thing for native applications”). They also mention that there are mismatches between the European Standard and the WCAG (P3), that user needs differ in the two documents (P3), and that it would be important to make an accurate interpretation of the standard and the WCAG (P5: “I think you should first start with EN standard with a good interpretation of WCAG, because otherwise it doesn’t make sense”). Some evaluators argue that the user’s perspective can provide some important information in the evaluation process (P2: “what we have to do then is to put ourselves in the place of the user, which is in principle really, really good because what we are evaluating is not so much compliance but more the actual user experience”).

One of the biggest difficulties of the existing methodologies is the fact that they are not adapted for mobile devices (P4: “as it is not adapted, as it was not made on purpose for mobile applications, there are some things that do not make so much sense”), having differences between operating systems, both in terms

of the criteria that the operating system itself requires to be met (P1: “I think Apple does something better, which is that when they make an application they have to meet certain criteria”) and in terms of automatic tools available for each one (P2: “I mean they are only available for Android primarily”). Participants also mentioned it could be advantageous to have access to the code to inspect some problems that may go unnoticed when using only the app (P3).

Some evaluators also discussed the way the problems are reported to the developers, which may not be understandable by developers if they don’t have some knowledge of accessibility (P2: “typically the recipient of an evaluation report like this, they don’t have knowledge enough about, you know, what the reasons would be for that symptom”). They also argue that the methodology should be oriented to solving the problems, not just finding them (P5: “it’s all about WCAG and finding the issues. And I think that’s wrong because it’s about fixing the issues”).

Finally, some of the accessibility flaws still found in mobile applications were also mentioned, the most common being the lack of labels on elements (P1: “Labels, labels of the fields and the buttons”), contrast between elements (P5: “about contrast elements, contrast of text”), problems related to navigation in the application (P4: “the navigation, the fact that they are adapted from web pages makes the navigation a lot worse”) and with the use of screen readers (P5: “The focus order for the screen reader that goes wrong quite often”).

5 Discussion

From the problems identified in the reports and the findings of the interviews, the most pressing issues preventing effective and efficient mobile accessibility evaluations emerged.

- Both the reports and the interviews indicate that there is no agreed upon methodology to be used when evaluating mobile applications. This originates from a clear lack of accessibility guidelines for mobile applications, which could be seen from several of our interviewees referring to WCAG. Unlike the web, with the WCAG, there is no set of guidelines to orient the development of accessible mobile applications. The European Standard has adopted the WCAG for the mobile context, but what is expressed in the reports and by our interviewees is that is not the proper solution. Our findings show that even if mobile accessibility evaluation methodologies exist, they were originally designed for the web and require adaptation, or are being developed independently by the evaluators or monitoring bodies themselves driven by their necessity.
- Our findings highlight that there is a lack of documentation about how the existing mobile accessibility guidelines should be applied. This problem is compounded by the fact that these guidelines have been originally developed for the web and, consequently, most documentation available was created for the web platform. It should be noted that, while our focus is on the evaluation

and the impact that the lack of documentation has on accessibility evaluators, this issue also applies to designers and developers that lack an understanding of how to avoid accessibility barriers during the design and development process. The way guidelines are presented right now leaves space for different interpretations and lacks descriptions of how they should be applied in the mobile context, therefore preventing consistent and effective mobile accessibility evaluations.

- There is a discrepancy in the availability of tools to evaluate the accessibility of web and mobile applications. There are a lot of tools, automatic or support, available for the web, while there is a lack of tools for mobile. Many web accessibility evaluation tools rely on the inspection of the DOM. This is usually not an issue, given that many web user agents make the DOM accessible for inspection. However, mobile platforms operate differently and it is much harder to programmatically access the underlying representation of the interface elements. For that reason, having access to the source code of the mobile application makes the evaluation process easier. But it is not always possible to access the source code of mobile applications when you are not their developer, which is the situation of monitoring agencies. Therefore, this is still an added limitation for mobile accessibility evaluation. Ideally, a methodology to monitor and evaluate the accessibility of a mobile application would be fully automatic. While this does not happen, developing automated or support tools to help with the evaluation would minimize the resources used.
- Since evaluations of mobile applications are conducted to understand how accessible they are for people with disabilities, it is important to make sure they are as accurate and complete as possible. There is no better way to understand their point of view than to listen to their feedback while using the application. Automatic accessibility evaluations are known to be limited in the type of barriers they can find. Manual evaluations depend on the expertise of the evaluator, which means that they also do not guarantee the identification of every possible barrier. For that reason, complementing automated and manual evaluations with user testing with participants with disabilities increases the changes that as many barriers as possible are identified. Still, the EN methodology only enforces automated and manual evaluations, leaving user testing as optional. Additionally, of the methodologies reviewed, those that mention user testing are rare. Although a manual evaluation will try to identify every possible barrier, this might not be possible every time, so, complementing accessibility evaluations with user tests should be further encouraged and incorporated in future methodologies.

Even though these are not directly related to the evaluation methodologies or processes, our study also allowed us to find what are the most often identified mobile accessibility problems in the reports and by the evaluators interviewed. These problems are often related to basic accessibility principles, such as elements without labels, low contrast between elements, text size that cannot be changed, missing language for screen readers, wrong focus setting, or naviga-

tion that does not work properly, with and without an external keyboard. These problems impact elements which support basic tasks that users need to perform and, thus, are not able to. Most of these problems are easy to solve at the source if designers and developers are educated and aware of mobile accessibility.

6 Conclusion

Ensuring the accessibility of mobile applications is a critical concern in today's digital landscape, particularly for the 1.3 billion people worldwide affected by disabilities. Mobile accessibility involves addressing the unique needs of individuals with disabilities using mobile devices. While smartphones offer benefits in terms of independence, they also present challenges due to their smaller screens and evolving nature. The lack of convergence and standards in the mobile industry further complicates accessibility efforts. However, the implementation of the Web Accessibility Directive in Europe highlights the importance of monitoring and evaluating the accessibility of public sector websites and mobile applications. Through periodic reports, Member States have shed light on various evaluation methodologies, tools, and challenges faced by monitoring bodies.

To better characterise these methodologies and challenges, we performed a review of existing research on mobile accessibility, analyzed monitoring reports from EU Member States, and conducted interviews with mobile accessibility evaluators. This procedure provided valuable information about the limitations of the current evaluation practices and the challenges faced by evaluators. The results of the reports and interviews highlight the absence of a methodology designed specifically for mobile applications, which leads to inconsistent evaluations. To improve the evaluation process, it is crucial to develop a standard methodology that takes into account the unique characteristics of mobile platforms. In addition, the limited supply of automated and support evaluation tools for mobile applications complicates the process by requiring more resources. Efforts should be made to improve existing tools and create new ones to meet the specific needs of mobile accessibility evaluations. By addressing these challenges and promoting collaboration between evaluators, developers, and users, we can contribute to a world more accessible and inclusive for all.

References

1. Acosta-Vargas, P., Guña-Moya, J., Jadán-Guerrero, J., Alvites-Huamaní, C., Salvador-Ullauri, L.: Towards accessibility assessment with a combined approach for native mobile applications. In: *Advances in Human Factors and System Interactions*, vol. 265, pp. 234–241 (2021). https://doi.org/10.1007/978-3-030-79816-1_29
2. Acosta-Vargas, P., Salvador-Acosta, B., Salvador-Ullauri, L., Villegas-Ch, W., Gonzalez, M.: Accessibility in native mobile applications for users with disabilities: a scoping review. *Appl. Sci. (Switzerland)* **11**, 5707 (2021). <https://doi.org/10.3390/app11125707>

3. Acosta-Vargas, P., et al.: Accessibility Assessment in Mobile Applications for Android, pp. 279–288 (2019). https://doi.org/10.1007/978-3-030-20040-4_25
4. Android: Accessibility scanner (2022). <https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.auditor&hl=en&gl=US>
5. CEN, C., ETSI: En 301 549 v3.2.1 (2021). https://www.etsi.org/deliver/etsi_en/301500_301599/301549/03.02.01_60/en_301549v030201p.pdf
6. European Commission: Web accessibility directive (2016). <https://eur-lex.europa.eu/eli/dir/2016/2102/oj>
7. European Commission: Commission implementing decision (EU) 2018/1524 (2018). https://eur-lex.europa.eu/eli/dec_impl/2018/1524/oj
8. European Commission and Directorate-General for Communications Networks, Content and Technology, et al.: Study supporting the review of the application of the Web Accessibility Directive (WAD): VIGIE 2020–0656: executive summary in English. Publications Office of the European Union, Luxembourg, UE (2022). <https://doi.org/10.2759/25194>
9. Member States of EU: Web accessibility directive - monitoring reports (2022). <https://digital-strategy.ec.europa.eu/en/library/web-accessibility-directive-monitoring-reports>
10. Ferreira, S.B.L., da Silveira, D.S., Capra, E.P., Ferreira, A.O.: Protocols for evaluation of site accessibility with the participation of blind users. *Procedia Comput. Sci.* **14**, 47–55 (2012). <https://doi.org/10.1016/j.procs.2012.10.006>
11. Appt Foundation: Appt evaluation methodology (Appt-EM) (2022). <https://appt.nl/toolkit/appt-em>
12. Appt Foundation: Appt web content accessibility guidelines (WCAG) (2022). <https://appt.org/en/guidelines/wcag/principle-1>
13. Guerreiro, T., Carriço, L., Rodrigues, A.: *Mobile Web*. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7440-0_37
14. Henry, S.L.: Mobile accessibility at W3C (2021). <https://www.w3.org/WAI/standards-guidelines/mobile/>
15. Henry, S.L.: What's new in WCAG 2.2 draft (2022). <https://www.w3.org/WAI/standards-guidelines/wcag/new-in-22/#257-dragging-movements-aa>
16. Henry, S.L., Brewer, J.: WCAG2ICT overview (2022). <https://www.w3.org/WAI/standards-guidelines/wcag/non-web-ict/>
17. Henry, W.S.L.: What's new in WCAG 2.1 (2020). <https://www.w3.org/WAI/standards-guidelines/wcag/new-in-21/>
18. Henry, W.S.L., Abou-Zahra, S.: WCAG-EM overview: Website accessibility conformance evaluation methodology (2020). <https://www.w3.org/WAI/test-evaluate/conformance/wcag-em/>
19. Henry, W.S.L., McGee, L.: Web accessibility initiative (WAI) at W3C (2018). <https://www.w3.org/standards/webdesign/accessibility#wai>
20. Apple Inc: Testing for accessibility on OS X (2018). <https://developer.apple.com/library/archive/documentation/Accessibility/Conceptual/AccessibilityMacOSX/OSXAXTestingApps.html>
21. Joshi, P., Bhagat, S.: Effective accessibility testing methodologies and seamless accessibility integration in mobile applications. In: ICEGOV, pp. 449–455. ACM, New York (2022). <https://doi.org/10.1145/3560107.3560175>
22. Mateus, D.A., Silva, C.A., Eler, M.M., Freire, A.P.: Accessibility of mobile applications: evaluation by users with visual impairment and by automated tools. In: IHC. ACM, New York (2020). <https://doi.org/10.1145/3424953.3426633>

23. Nicolau, H., Guerreiro, T., Jorge, J., Gonçalves, D.: Mobile touchscreen user interfaces: bridging the gap between motor-impaired and able-bodied users. *Univ. Access Inf. Soc.* **13**, 303–313 (2014). <https://doi.org/10.1007/s10209-013-0320-5>
24. World Health Organization: Disability (2022). <https://www.who.int/news-room/fact-sheets/detail/disability-and-health>
25. ProfessionalQA: Accessibility testing: Complete guide (2020). <https://www.professionalqa.com/accessibility-testing>
26. Rabin, J., McCathieNevile, C.: Mobile web best practices 1.0 (2008). <https://www.w3.org/TR/mobile-bp/#d0e128>
27. da Silva, C.F., Ferreira, S.B.L., Sacramento, C.: Mobile application accessibility in the context of visually impaired users. In: *IHC. IHC 2018*, ACM, New York (2018). <https://doi.org/10.1145/3274192.3274224>
28. W3C: Authoring tool accessibility guidelines (ATAG) 2.0 (2015). <https://www.w3.org/TR/ATAG20/>
29. W3C: User agent accessibility guidelines (UAAG) 2.0 (2015). <https://www.w3.org/TR/UAAG20/>
30. W3C: WCAG 2 overview (2022). <https://www.w3.org/WAI/standards-guidelines/wcag/>
31. W3C Kim Patch, Jeanne Spellman, K.W.: Mobile accessibility: How WCAG 2.0 and other W3C/WAI guidelines apply to mobile (2016). <https://www.w3.org/TR/mobile-accessibility-mapping/>