UNIVERSIDADE FEDERAL DE LAVRAS

DELVANI ANTÔNIO MATEUS

ACCESSIBILITY OF MOBILE APPS FOR VISUALLY IMPAIRED USERS:PROBLEMS ENCOUNTERED BY USER EVALUATION, INSPECTIONS, AND AUTOMATED TOOLS

LAVRAS – MG

2022

DELVANI ANTÔNIO MATEUS

ACCESSIBILITY OF MOBILE APPS FOR VISUALLY IMPAIRED USERS :
PROBLEMS ENCOUNTERED BY USER EVALUATION, INSPECTIONS, AND AUTOMATED TOOLS

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, linha de pesquisa de Engenharia de Software e Sistemas de Informação, para a obtenção do título de Mestre

DSc. Mauricio Ronny de Almeida Souza
Orientador

DSc. André Pimenta Freire
Coorientador

LAVRAS – MG
2022

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Mateus, Delvani Antônio.
Accessibility of Mobile Apps for Visually Impaired Users : Problems Encountered by User Evaluation, Inspections, and Automated Tools / Delvani Antônio Mateus. – Lavras : UFLA, 2022.
105 p. : il.

Dissertação (mestrado acadêmico)–Universidade Federal de Lavras, 2022.
Orientador: DSc. Mauricio Ronny de Almeida Souza.
Coorientador: DSc. André Pimenta Freire.
Bibliografia.

1. Problemas de Acessibilidade. 2. Acessibilidade Móvel. 3. Análise Qualitativa. I. Souza, Mauricio Ronny de Almeida. II. Freire, André Pimenta. III. Título.

DELVANI ANTÔNIO MATEUS

ACCESSIBILITY OF MOBILE APPS FOR VISUALLY IMPAIRED USERS: PROBLEMS ENCOUNTERED BY USER EVALUATION, INSPECTIONS, AND AUTOMATED TOOLS
ACESSIBILIDADE DE APLICATIVOS MÓVEIS PARA USUÁRIOS COM DEFICIÊNCIA VISUAL: PROBLEMAS ENCONTRADOS POR AVALIAÇÕES DO USUÁRIOS, INSPEÇÕES E FERRAMENTAS AUTOMATIZADAS

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, linha de pesquisa de Engenharia de Software e Sistemas de Informação, para a obtenção do título de Mestre

APROVADA em 16 de Dezembro de 2022.

DSc. Paulo Afonso Parreira Junior UFLA
DSc. Simone Bacellar Leal Ferreira UNIRIO

DSc. Mauricio Ronny de Almeida Souza
Orientador

DSc. André Pimenta Freire
Co-Orientador

LAVRAS – MG
2022

AGRADECIMENTOS

Primeiramente, expresso minha profunda gratidão a Deus e aos meus pais, bem como à minha esposa, que sempre esteve presente, apoiando e auxiliando em todos os momentos do meu processo de mestrado. Aos meus familiares, que também me apoiaram e incentivaram, sou imensamente grato. Agradeço de maneira especial aos meus amigos, que sempre me deram apoio e estiveram torcendo pelo meu sucesso. Não poderia deixar de mencionar e expressar minha profunda admiração e agradecimento aos meus orientadores, Prof. Mauricio Ronny de Almeida Souza e Prof. André Pimenta Freire, por sua orientação, apoio e incentivo em todo o processo. Por fim, dedico esta conquista em memória de Jéssica Oliveira.

Eu também gostaria de expressar minha gratidão à FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) e à UFLA (Universidade Federal de Lavras) pelos recursos disponibilizados durante todo o processo. Sua contribuição foi fundamental para o sucesso deste trabalho.

"Tudo tem o seu tempo e há tempo para todo propósito debaixo do céu."
(Eclesiastes 3:1)

ABSTRACT

Mobile apps have grown considerably, supporting many everyday activities such as news reading, social media, and online banking. Mobile apps should be usable and accessible to everyone, including visually impaired people. Accessibility evaluation techniques are essential to identify accessibility problems and help to avoid problems that hinder access by people with disabilities. Different accessibility evaluation methods may be used in different stages of software development. These methods include approaches such as employing automated tools to detect specific issues in the source code and more comprehensive methods, such as user evaluation and expert inspections. Expert inspections and automated tools can help detect problems earlier in the development process, though they cannot cover all the problems visually impaired users encounter. However, knowing the types of accessibility problems encountered by each method is vital for developers, testers, and designers to allocate the different methods appropriately in different stages of the development cycle. This project aimed to compare the outcomes of automated evaluations, manual inspection and user evaluation methods applied to mobile apps focusing on visually impaired users. The study compared the results from different methods by comparing results from a systematic mapping of the literature and the analysis of user evaluations performed in a previous study with two mobile apps. The study compared the results from the usability evaluations of Saraiva and Receita Federal by nine users with visual impairment and 189 instances of problems in two applications. These results were compared to the automated evaluation of the two apps by the tools MATE (Mobile Accessibility Testing) and Accessibility Scanner. Next, the apps were inspected by two groups of professionals: 17 experts in different areas of software development (full-stack developers, testers and front-end developers) and ten specialists in Human-Computer Interaction (HCI) with previous experience with accessibility. The results showed a difference between the different accessibility evaluation methods. Inspections by HCI specialists and tests with users with disabilities found a greater diversity of types of problems. Automated tools showed limited performance in the detection of types of problems. In accessibility inspections, There is a difference between specialists in software development and specialists in HCI. The results showed a difference in the number of violations that specialists in software development find compared to HCI specialists. This study contributes to the understanding of how different methods can contribute to the accessibility evaluation of mobile apps, helping different stakeholders make decisions about when and how to apply different methods.

Keywords: Mobile Accessibility. Users. Tools. Inspections.

RESUMO

Os aplicativos móveis cresceram consideravelmente, suportando muitas atividades cotidianas, como leitura de notícias, mídia social e serviços bancários online. Os aplicativos móveis devem ser utilizáveis e acessíveis a todos, incluindo pessoas com deficiência visual. Técnicas de avaliação de acessibilidade são essenciais para identificar problemas de acessibilidade e ajudar a evitar problemas que dificultem o acesso de pessoas com deficiência. Diferentes métodos de avaliação de acessibilidade podem ser usados em diferentes estágios de desenvolvimento de software. Esses métodos incluem abordagens como o emprego de ferramentas automatizadas para detectar problemas específicos no código-fonte e métodos mais abrangentes, como avaliação do usuário e inspeções de especialistas. Inspeções especializadas e ferramentas automatizadas podem ajudar a detectar problemas no início do processo de desenvolvimento, embora não possam cobrir todos os problemas que os usuários com deficiência visual encontram. No entanto, conhecer os tipos de problemas de acessibilidade encontrados por cada método é vital para que desenvolvedores, testadores e designers aloquem os diferentes métodos adequadamente em diferentes estágios do ciclo de desenvolvimento. Este projeto teve como objetivo comparar os resultados de avaliações automatizadas, inspeção manual e métodos de avaliação de usuários aplicados a aplicativos móveis com foco em usuários com deficiência visual. O estudo comparou os resultados de diferentes métodos comparando os resultados de um mapeamento sistemático da literatura e a análise de avaliações de usuários realizadas em um estudo anterior com dois aplicativos móveis. O estudo comparou os resultados das avaliações de usabilidade do Saraiva e da Receita Federal por nove usuários com deficiência visual e 189 ocorrências de problemas em dois aplicativos. Esses resultados foram comparados com a avaliação automatizada dos dois aplicativos pelas ferramentas MATE (Mobile Accessibility Testing) e Accessibility Scanner. Em seguida, os aplicativos foram inspecionados por dois grupos de profissionais: 17 especialistas em diferentes áreas de desenvolvimento de software (desenvolvedores full-stack, testadores e desenvolvedores front-end) e dez especialistas em Interação Humano-Computador (IHC) com experiência anterior em acessibilidade. Os resultados mostraram uma diferença entre os diferentes métodos de avaliação de acessibilidade. Inspeções por especialistas em IHC e testes com usuários com deficiência constataram maior diversidade de tipos de problemas. As ferramentas automatizadas apresentaram desempenho limitado na detecção de tipos de problemas. Nas inspeções de acessibilidade, existe uma diferença entre especialistas em desenvolvimento de software e especialistas em IHC. Os resultados mostraram uma diferença no número de violações que os especialistas em desenvolvimento de software encontram em comparação com os especialistas em IHC. Este estudo contribui para a compreensão de como diferentes métodos podem contribuir para a avaliação de acessibilidade de aplicativos móveis, ajudando diferentes stakeholders a tomar decisões sobre quando e como aplicar diferentes métodos.

Palavras-chave: Acessibilidade Móvel. Usuários. Ferramentas. Inspeções.

LIST OF FIGURES

LIST OF TABLES

LIST OF CHARTS

CONTENTS

1 INTRODUCTION

During the COVID-19 pandemic, social distancing measures (OMS, 2021) boosted an increase in the use of digital services such as banking services, online medical appointments, social service apps, and others. However, these services are not usable by everyone. For example, visually impaired people have encountered significant problems in mobile applications due to the lack of accessibility to digital services (MARTINS et al., 2022; FOESCH et al., 2022).

In 2015, the Brazilian Law of Inclusion was approved (Governo Brasileiro, 2015). This law is fundamental as it makes accessibility to digital sites in the national territory mandatory, despite still being reticent regarding the accessibility of mobile applications. In the case of web applications, there was extensive research on how to improve the accessibility of web content to people with visual impairments and methods for evaluating websites and applications to identify problems (VIGO; BROWN; CONWAY, 2013; RØMEN; SVANÆS, 2012; POWER et al., 2012; JAEGER, 2006). Many studies focused on assessing web accessibility are focused on the use of international guidelines, such as the Web Content Accessibility Guidelines (WAG) (W3C, 2008), developed by the World Wide Web Web's Accessibility Initiative (WAI) Consortium (W3C).

Understanding the nature of the problems encountered by visually impaired users and the most effective methods to encounter them is a very relevant issue. The most used accessibility assessment methods include approaches based on tests performed by users, manual inspections carried out by specialists, and automated assessments carried out by tools that verify the compliance of applications with a set of requirements (JAEGER, 2006). Automated assessment tools are important assets in the assessment process, as they allow the automation of repetitive tasks performed by expert assessors. However, some studies have identified that they cannot identify all problems (VIGO; BROWN; CONWAY, 2013; JAEGER, 2006) in web applications.

1.1 Motivation and Justification

On my journey in the computing area in a certain service call, I assisted a user with a deaf and communication was very difficult because I had no domain in sign languages. At the same time, I started taking courses in LIBRAS (Brazilian Sign Language), which helped me to understand the difficulty that disabled people encounter during their lives. During this process, I started working with my advisor in the area of accessibility and with each violation of the right to use digital applications, I feel the need to contribute to society, trying to make digital applications accessible.

Despite having a relatively more recent uptake than on the web, mobile accessibility has increased interest in research (QUISPE; ELER, 2018; PARK; SO; CHA, 2019; DIAS, 2018; LOPES; FAÇANHA; VIANA, 2022; ALAJARMEH, 2021; FOESCH et al., 2022). Specific guidelines for mobile accessibility have been created, as the BBC Mobile Accessibility Standards and Guidelines v1.0 (BBC, 2014b), the Guide to the development of accessible mobile applications, by Samsung (SIEBRA et al., 2017), and Brazilian Standard (NBR) 17060 Mobile Applications Accessibility Standard (Governo Brasileiro, 2022).

Evaluation methods can be used at different stages of the project. Developers, testers and designers need to understand what problems are detected by expert inspection methods, automated testing, and user testing to appropriately employ them in different stages of the software development process.

Knowing that each method has a different cost-benefit ratio, it is crucial to analyse which test method should be performed throughout the development process, making it necessary to compare each method's coverage. For example, tests with design allow us to identify the level of knowledge inaccessibility and what problems can be avoided in development time (MANKOFF; FAIT; TRAN, 2005). The problems found by all methods are related to each other, and in some methods, practitioners may find more serious problems than in other methods (HARRISON; PETRIE, 2007; CALVO; SEYEDARABI; SAVVA, 2016). The existing guidelines cannot cover all the problems encountered (POWER et al., 2012; ALAJARMEH, 2021).

Several studies (DIAS, 2018; CARVALHO et al., 2018; POWER et al., 2012; FREIRE, 2012; RØMEN; SVANÆS, 2012; BRAJNIK, 2008; ALAJARMEH, 2021) identified that the guidelines are not able to cover all the problems found by users with disabilities. Inspection methods and user testing generally encounter more types of problems than automated tools (SILVA et al., 2019). Automatic tools can find more instances of problems due to how they are implemented. However, they have limitations in terms of the limitations to the types of problems (SILVA; ELER; FRASER, 2018a).

1.1.1 Objectives

The aim of this study is to analyze the types of problems encountered by automated tools, specialized inspection (containing developers and experts in human-computer interaction (HCI)), and user assessment of mobile app accessibility, focusing on visually impaired users.

1.2 Research gap identified

The knowledge about the coverage of problems encountered by testing methods by experts and automated tests and the coverage of accessibility problems that users would encounter in mobile applications is still limited. Many studies have shown the different accessibility problems encountered by different methods (VIGO; BROWN; CONWAY, 2013; JAEGER, 2006; SOUZA; CARDOSO; PERRY, 2019) including comparisons of assessments with users and automated tools in the particular context of Web applications. Understanding which problems the automated tests and tests with experts and designs can effectively find is e ssential. Being able to do so delimits the scope that developers and designers can use them and what problems need to be found by user evaluation in the context of mobile applications.

With the knowledge of the most appropriate method to be employed in each stage, more accessible mobile apps may be developed, improving the development process's efficiency.

Thus, this research aims to investigate the types of problems found in mobile applications through tests with visually impaired users, inspections by professionals, and automated tools. The study used a data set of problems encountered by visually impaired users Dias (2018), Carvalho et al. (2018) and compared them with the results of assessments made by different automated tools and with the results of inspections made by professionals involving people with different levels of experience in the evaluation using guidelines. The applications evaluated were identical to those evaluated by visually-impaired users in previous studies. All versions are available in the repository Apkpure [1]. Furthermore, the release period for the versions was after the publication of the WCAG 2.1 directive. Thus, the results were compared to categorize the types of problems encountered uniquely by each type of method, the coincident problems, and considerations on the applicability of the different types of methods at different stages of the development of mobile applications.

[1] https://m.apkpure.com/br/

1.3 Research problem

This research aimed to identify if there is a difference between assessment methods, namely: automated tools, user testing, and inspections by professionals. It also aimed to understand if there is a difference between expert inspections by specialists in software development and expert inspections with HCI specialists.

We defined the following specific questions to answer the research question:

RQ1: Is there a difference between accessibility inspections of mobile applications performed by software development specialists and HCI specialists?

1.4 Organization

The remainder of this document is organized as follows. Chapter 2 presents the theoretical foundations and related works. Chapter 3 presents the methodological framework and the general study design. Chapter 4 presents the systematic literature review of studies that conducted accessibility evaluations of mobile applications with automated tools, inspections and user evaluation. Chapter 5 presents an empirical study comparing the results from user evaluation of mobile apps and the results from automated evaluations. Chapter 6 presents the results from a comparison between user evaluation and manual inspections of mobile applications. Finally, Chapter 7 presents conclusions and future work.

2 THEORETICAL BACKGROUND

This section presents basic concepts and related studies that have approached accessibility inspections, tests with users, automated tests, and comparisons between evaluation methods in mobile applications.

2.1 Accessibility in Mobile Applications

Accessibility is defined by ISO 9241-11 as "Usability of a product, service, environment or installation by people with the widest range of resources" (ISO, 2018). ISO 9241-11 defines what usability is: "The extent to which a system, product or service can be used by specific users to achieve specific objectives with effectiveness, efficiency and satisfaction in a specific context of use" (ISO, 2018).

The ISO 9241-11 definition (ISO, 2018) implies that accessibility can be understood as enabling the largest number of users to use a system including people with visual impairments. Accessibility guidelines commonly used to guide the development of applications were intended to promote best practices. However, these guidelines alone are not able to cover all accessibility requirements by users (FREIRE, 2012).

Accessibility guidelines were created to promote accessibility on websites, applications and digital services, such as the Web Content Accessibility Guidelines (WCAG 2.1) (W3C, 2018), the British Broadcasting Corporation's (BBC) BBC HTML Accessibility Standards v 2.0 (BBC, 2014a) and BBC Mobile Accessibility Standards and Guidelines v1 .0 (BBC, 2014b). The Brazilian government also created its e-MAG (Governo Brasileiro, 2014) accessibility guidelines.

In October 2022, the NBR 17060 (Brazilian standard) was released. It deals with the accessibility standard in mobile applications (Governo Brasileiro, 2022). This standard is based on the WCAG (Web Content Accessibility Guidelines), whose principles are; 1 Perception and understanding; 2 Control and interaction; 3 Media; and 4 Codification. Its main objective is to support article 63 of the Brazilian Inclusion Law (LBI 13.146/2015) (Governo Brasileiro, 2015).

The set of guidelines developed by the World Web Consortium (W3C) as the Web Content Accessibility Guidelines (WCAG) has its version 3.0 under development (W3C, 2021). Version 2.2 was published as a working draft in 2021 (W3C, 2020), which extends from version 2.1, published in 2018 (W3C, 2018). Those are extensions to version 2.0, published in

2008 (W3C, 2008), which originated from the first publication published in 1999 (W3C, 1999). This set of guidelines is the most used worldwide.

These guidelines are aimed at the accessibility of web content to people with disabilities. The guidelines have three levels: the lowest (A), the intermediate (AA), and the highest (AAA). The success criteria are divided into five categories:

a) Perceivable: The components and information of the interface must be perceivable to users;
b) Operable: Navigation through the interface components must be operational;
c) Understandable: All the contents of the interface must be understandable;
d) Robust: The content must be able to be interpreted by assistive technologies, and can be used by all users;
e) Compliance: The content must support assistive technologies and must conform to the concept of compatible accessibility for each technology.

The Brazilian federal government created its e-government accessibility model (e-MAG). This model has the principle of guiding developers to good accessibility practices so that each person, regardless of disability, can have unrestricted access to information (Governo Brasileiro, 2014). e-MAG is a version adapted from WCAG for use by the Brazilian government.

The Brazilian Law of Inclusion of Persons with Disabilities (Statute of Persons with Disabilities), Law No. 13,146, was created in 2015. It states that: accessibility is the possibility, condition, and use of communication, information, multimedia systems, and other technologies. Chapter II, which deals with access to information and communication, in its Art. 63, states that “Accessibility to websites maintained by companies with headquarters or commercial representation in the country or by government agencies is mandatory, for the use of persons with disabilities, ensuring that access to available information, following the best practices and accessibility guidelines adopted internationally”. Web sites must contain highlighted accessibility symbol (Governo Brasileiro, 2015). Law number 14.129, of March 29, 2021, brings the “principles, rules and instruments for increasing the efficiency of public administration, especially through reducing bureaucracy, innovation, digital transformation, and citizen participation” (Governo Brasileiro, 2021).

Other local initiatives have also emerged in Brazil. The city of São Paulo created the CPA (Permanent Accessibility Commission) to assess and provide accessibility seals in sites with 95% accessibility as evaluated by automated tools. This commission offers technical reports pointing out accessibility violations; companies have 60 days to make the correction and to go through the evaluation again. In 2022, the CPA had already issued 70 Digital Accessibility Seals delivered to websites since its launch in 2015 (Prefeitura Paulista, 2018). Two studies carried out by the *Movimento Web Para Todos* (WPT, 2019), and *BigDataCorp* (BigDataCorp, 2019) show that less than 1% of Brazilian websites are accessible and 100% of the sites accessed from Brazilian e-commerce present barriers to the navigation of this audience. For this reason, this concern needs to be constant.

2.2 Accessibility Evaluation

In the literature, there are different methods to evaluate accessibility and usability. Different methods may involve tests with users, inspections conducted by specialists and automated tests. Manual inspections involve the analysis by specialists of the interface and source code (BRAJNIK, 2008).

2.2.1 Automated tests

Automated evaluations usually are part of a broader accessibility inspection procedure in which an expert uses a tool to inspect the application or its source code according to guidelines and recommendations (BRAJNIK; YESILADA; HARPER, 2011). Automated evaluation tools can provide faster feedback on conformance to guidelines. However, they have many limitations, as they cannot identify all problems listed in the guidelines, as not all tests are automatable (BRAJNIK, 2008). Many studies conduct evaluations of large amounts of websites using only automated evaluation tools. However, those studies cover only a tiny portion of accessibility recommendations (ELER et al., 2018).

Tools can help with repetitive tasks such as identifying colour contrast, target size, and lack of content description (BRAJNIK; YESILADA; HARPER, 2011; SILVA; ELER, 2018). This method is limited and cannot identify all accessibility issues. The tool checks for presence and absence but cannot verify that the content description is correct. (BRAJNIK; YESILADA; HARPER, 2011; ELER et al., 2018; MATEUS et al., 2020).

2.2.2 User Tests

User testing is considered the most comprehensive method of accessibility evaluation (BRAJNIK; YESILADA; HARPER, 2011; PETRIE; BEVAN, 2009). They consist of performing evaluations involving users with disabilities, who attempt to carry out tasks in interactive systems to identify possible problems (BRAJNIK, 2008). This method is the most effective in the sense that it addresses the issue of whether users with disabilities can or cannot use a system.

Involving users with disabilities in the accessibility verification process is very important to reveal accessibility problems they might have with a system. User evaluation usually employs think-aloud techniques and the observation of users' facial expressions from an expert's analysis. Major accessibility problems can be found, and the severity (SILVA et al., 2019; BRAJNIK; YESILADA; HARPER, 2011) may also be verified by observing the actual impact of possible problems on the tasks. However, recruiting users with different types of disabilities is not an easy task (PETRIE; POWER, 2012).

2.2.3 Expert inspections

Manual inspection methods involve the conduction of verifications performed by accessibility specialists, in which they perform manual tests on the application and its source code to verify whether it complies with recommendations, such as WCAG 2.1 (W3C, 2008), the BBC accessibility recommendations (BBC, 2014b), the Brazilian Accessibility Model for Electronic Government (e-MAG), among others. Such verifications aim to establish whether applications conform to standards and anticipate problems that users may encounter.

This method is vital for finding issues on mobile platforms that cannot be automatically verified and that users do not have the technical knowledge to find (QUISPE; ELER, 2018). Inspections by specialists may also help find problems that users might not perceive during accessibility evaluations. For example, a blind person might not even notice that an image with an empty textual description is present.

Chart 2.1 presents some advantages and disadvantages of each accessibility evaluation method.

Chart 2.1 – Applications evaluated by blind and partially-sighted users and problems

Method	Advantage	Disadvantage
Automated tests	It has a relatively low cost, can find larger amounts of problem instances, is great for repetitive tasks, and has faster results.	It has a lower detection power in relation to problem types, it can only find violation items that can be found by presence and absence.
User tests	It has the power of detecting types of problems greater than any other method. can find issues that are not covered by the guidelines.	It has a high cost and the tests take a longer time to be carried out.
Expert test	It performs better than automated tests, it finds issues that are not covered by the guidelines.	It has a lower performance than user tests, it has a relatively low power to detect instances of problems compared to automated tools.

Source: Own author

2.3 Related work

This section discusses related studies that performed different accessibility evaluations and comparisons between evaluation methods.

The study by Power et al. (2012) found web accessibility problems in sixteen sites from a study with thirty-two users with visual impairment. Using the WCAG 2.0 guidelines, they observed that only 50.4% of the problems encountered by users were covered. This fact shows that guidelines alone cannot cover all accessibility problems.

To verify the coverage of the WCAG 1.0 and WCAG 2.0 guidelines at two sites in Norway, Rømen e Svanæs (2012) conducted tests with a total of thirteen users (three people with visual impairment, two users with dyslexia, two members with motor disabilities and six people without disabilities). As a result, it showed that WCAG 2.0 coverage corresponds to approximately 49% of the problems and WCAG 1.0 to 42%. Even if developers followed the guidelines, they would not cover all barriers.

A study by the ALCANCE research group (DIAS, 2018; CARVALHO et al., 2018) aimed to understand the types of problems encountered by visually impaired users in mobile applications. This study provided input from the user evaluations used in the present dissertation. The study involved user evaluations of four applications: Caixa Econômica Federal (government-owned bank), Receita Federal (National Treasury), Decolar (tourism agency), and

Saraiva (bookseller) on Android and iOS platforms. The study recruited eleven users, six blind and five with low vision. The results yielded 39 types of problems and a total of 415 problem instances. Table 2.1 presents the problems encountered on each application in that study. Table 2.2 and Table 2.3 present the problems with higher instances encountered by blind users and users with low vision, respectively. In addition to the accessibility problems identified, this study (DIAS, 2018; CARVALHO et al., 2018) showed that users with visual impairments had more significant difficulties in completing tasks than users without disabilities and that the severity of the problems encountered is also greater.

Table 2.1 – Applications evaluated by blind and partially-sighted users and problems

-		Blind users		Partially-sighted users	
Apps	**Total**	**# Users**	**Problem Instances**	**# Users**	**Problem instances**
Caixa	101	4	70	5	31
Decolar	124	4	61	5	63
Saraiva	76	3	37	4	39
Receita	114	5	94	4	20

Source: Carvalho et al. (2018), Dias (2018)

The problems encountered in that study (DIAS, 2018; CARVALHO et al., 2018) were categorized using a categorization scheme adapted from (POWER et al., 2012). The most frequently encountered problems by blind users and their frequencies are shown in Table 2.2.

Table 2.3 presents the list of the most frequently encountered problems by partially-sighted users by the same study (DIAS, 2018; CARVALHO et al., 2018).

Other related studies have also investigated the interplay between different accessibility evaluation methods, including inspections by specialists. Accessibility studies using different assessment methods can target whether websites follow accessibility guidelines. In the study by Jaeger (2006) that sought to identify the accessibility of ten US government websites to find out if they are following the Section 508 accessibility guide, the following evaluation methods were used: tests by specialists, tests with users being carried out by people visually and motor impaired; automated tests and questionnaire for webmasters, involving professionals from the selected sites. This study showed that if the 508 guide recommendations are implemented, the sites will be accessible to people with disabilities. Furthermore, it showed that testing with users offered greater detail about the problems encountered.

In tests carried out by Mankoff, Fait e Tran (2005), they evaluated the accessibility of 50 websites, comparing them with inspection methods by experts, automated tests, and user tests.

Table 2.2 – Most frequently encountered problems by blind users and their frequencies

Problem category	Total (N)
3.4.1 Inappropriate feedback (Controls, forms and functionality)	34
3.4.5 Unclear or confusing functionality (Controls, forms and functionality)	25
3.5.1 Lack of identification (Buttons)	22
3.4.9 Users inferred that there was functionality where there wasn't (Controls, forms and functionality)	18
1.1.3 Users cannot make sense of content (Content - meaning)	15
3.4.8 Sequence of interaction is unclear or confusing (Controls, forms and functionality)	15
3.5.2 Button functionality is nuclear or confusing (Buttons)	15
2.2.1 No textual alternative (Image)	14
1.1.4 Inconsistent Content organization (Content - meaning)	12
3.4.10 Default presentation of control or form element is not adequate (Controls, forms and functionality)	11
3.4.4 Expected functionality not present (Controls, forms and functionality)	10
6.1.3 System problems with assistive technology (System characteristic)	8
1.1.7 Meaning in content is lost (Content - meaning)	6
2.3.1 No textual alternative (Audio, video and multimedia)	6
3.4.2 Functionality does not work (as expected) Controls, forms and functionality	6

Source: Carvalho et al. (2018), Dias (2018)

Specialists were developers with an average of between 2 and 8 years of experience in a total of seventeen participants. Some of the tests performed by programmers would have had to turn off monitors and use screen readers. In user tests, five blind users participated, one with less than two years of experience, two with two to six years of experience, and one with more than six years of experience. In comparison, the researchers found a great variety of problems found individually by the evaluated ones. In this way, they decided to analyze the total of problems by a group of evaluators. User tests were divided into two stages, laboratory tests, and remote tests. They observed that 50% of the problems encountered were due to the combination of screen readers and monitors. However, other problems were not detected.

A study with 12 Brazilian government applications Quispe e Eler (2018) suggests the adaptation of the Brazilian government's e-MAG Web accessibility guidelines, which can be applied in the development of mobile applications. The Accessibility Scanner tool was used to evaluate the applications. As a result of the evaluations, the authors made 35 recommendations for adapting e-MAG. The results show that it is necessary to have an accessibility guide for developing mobile applications by the Brazilian government. In addition, the authors conclude

Table 2.3 – Most frequent problems encountered by partially-sighted users

Problem category	Total (N)
3.4.1 Inadequate feedback (Controls, forms and functionality)	15
3.4.10 Default presentation of control or form element is not adequate (Controls, forms and functionality)	12
2.1.5 Default presentation is not adequate (Text)	11
6.1.4 System too slow (System characteristic)	11
3.4.2 Functionality does not work (as expected) Controls, forms and functionality	10
3.4.9 Users inferred that there was functionality where there wasn't (Controls, forms and functionality)	9
3.5.1 Lack of identification (Buttons)	8
6.1.1 Server not working appropriately (System characteristic)	8
1.1.4 Inconsistent content organization (Content - meaning)	6
3.4.8 Sequence of interaction is unclear or confusing (Controls, forms and functionality)	6
1.1.3 Users cannot make sense of content (Content - meaning)	5
3.4.4 Expected functionality not present (Controls, forms and functionality)	5
4.1.1 Inadequate navigation elements (Navigation)	5
1.1.7 Meaning in content is lost (Content - meaning)	4
2.2.5 Default presentation not adequate (Image)	4

Source: Carvalho et al. (2018), Dias (2018)

that the flaws found in all applications can prevent disabled users from accessing all application functionality.

The study by Silva, Ferreira e Sacramento (2018a) investigated accessibility barriers when observing the interaction of five visually impaired people, of which four had some residual vision. They pointed out gaps in the technical guidelines of WCAG 2.0 in the application aimed at e-commerce, Mercado Livre. Users used a screen reader to perform tasks and later answered a questionnaire to record their thoughts during this execution. This process characterizes the retrospective verbalization protocol (SILVA; FERREIRA; SACRAMENTO, 2018a). As a result, users encountered several difficulties in interacting with the application were recorded, in addition to violations related to the WCAG 2.0 guidelines, such as the presence of icons with alternative text that did not adequately describe the features, violating one of the WCAG 2.0 guidelines.

To investigate mobile accessibility issues, Carvalho e Freire (2017), Carvalho, Dias e Freire (2018) analyzed the suitability of interface components in developing mobile systems. They used three prototypes of mobile systems based on accessibility problems for people with visual impairments. They selected thirty Android interface components based on the investigation of documents on standard components related to HTML and the Android system. Based on WCAG 2.0 success criteria, an expert evaluator audited all sample interface components present in the three prototypes. These prototypes were implemented using three methods: generating a native application with standard Android Studio components, another generated a system with web resources developed in HTML components, and the last one a hybrid application.

With the same objective of identifying accessibility barriers, Silva, Ferreira e Ramos (2016a) conducted an empirical study based on mobile systems. They evaluated the WhatsApp application. Five blind users performed eleven tasks and verbalized their interaction experiences. They identified barriers such as no feedback and unlabeled buttons and identified and classified them according to the success criteria of WCAG 2.0. They also noted violations of the guidelines in almost all principles except the robust principle.

When evaluating the types of interaction facilitators used in mobile systems by people with visual impairments Ghidini et al. (2016) developed a prototype of an electronic diary and tests with the participants. For this, they interviewed six visually impaired people to identify the most used means of interaction with smartphones and the facilities and barriers. They developed a prototype to obtain other results, replacing the native calendar. Subsequently, they performed tests with four users using the native calendar and the prototype, who performed tasks with a screen reader. The authors report that users describe the native calendar as having reduced usability. When implementing observations pointed out by users, a new test was performed with a low-vision user, involving the native application and the researchers' prototype. Again, the prototype results demonstrated improved usability.

The study by Oncins (2021) qualitatively researched accessibility in seven videoconferencing platforms. The tests carried out with users happened synchronously. The results were categorized according to WCAG 2.1 guideline. Of the seven platforms, those with the best accessibility were Zoom, Google Meets, and Microsoft teams. The author concludes that, although accessibility on videoconferencing platforms is improving, users with disabilities still face difficulties using them.

Another survey involving 16 visually impaired users verified accessibility on mobile devices, totalling 34 problems found concerning WCAG 2.1. The researchers (ALAJARMEH, 2021) compared the problems found with the WCAG 2.0 and WCAG 2.1 guidelines to verify the level of coverage of the problems encountered by the users. They verified the existence of success criteria related to each problem and the sufficiency of the compliance and success criteria. The study examined whether there are relevant success criteria in WCAG 2.1 for a problem identified or not by users. The authors indicated that the analysis is key to suggesting appropriate compliance levels for WCAG 2.1 and incorporating more access barriers encountered by blind or visually impaired users. The authors conclude that the WCAG 2.1 version significantly improved over the previous version. However, the developers had not yet fully implemented these criteria. In addition, the researchers report that there are issues not covered by the guideline, with room for future improvements in its new version.

Along the same lines for verifying accessibility through automated tests, Silva, Eler e Fraser (2018a) sought to identify which accessibility properties can be automated, partially automated, and which cannot be automated. For this, they used seven automatic tools, such as:

Accessibility Scanner (GOOGLE, 2015) is a mobile application on the Android platform, which, when printing the screen of an application in use, generates recommendations for improving accessibility.

MATE (ELER et al., 2018) is an automated accessibility verification tool for Android applications. When using the tool, each screen accessed from the application under evaluation is checked for accessibility guidelines, at the end, a file is generated containing the problems found.

All tools are open source and available for use. With the results, the authors made a guide for the developers of accessibility tools to create robust tools to use the designs.

The studies analyzed in this section show that many advancements have been achieved in understanding the accessibility problems encountered by visually impaired users. However, there is little information about the types of problems each method can encounter, and how to provide orientation to developers, testers and designers to employ them in the development process.

3 METHOD

This chapter presents the methods used to answer the research questions. The study aimed to compare accessibility problems in mobile applications found by expert inspections, user tests, and automated assessment tools.

3.1 Study Design

To answer the research questions, this dissertation was divided into three studies. The first is a systematic mapping study to survey the literature for types of problems encountered by the different accessibility assessment methods. The second study aims to compare accessibility problems in mobile applications found by expert inspections and user tests. Similarly, the third study aims to compare the results of user evaluations and manual inspections performed by specialists. The three studies provide evidences for addressing RQ1 (“Is there a difference between the accessibility assessment methods in the mobile context?”). The third study specifically address RQ2 (“Is there a difference between software development specialists and HCI specialists in the types of problems encountered in the accessibility evaluation of mobile applications?”).

The following subsections provide further details of the methods used in each study.

3.2 Study 1: Systematic Mapping Study

Systematic mapping study is a secondary study method that systematically explores and categorizes studies in a given research field, and provides a structure of the type of research reports and results that have been published, hence it contributes to confirm research gaps (PETERSEN et al., 2008). This study investigated the types of problems encountered by the different accessibility assessment methods, present in the literature on Human-Computer Interaction, the investigation focuses on people with disabilities including the visually impaired. Accessibility assessments can be performed through user testing, manual inspections by experts, and automated testing. Effective accessibility assessments are very important for people with visual impairments, as many interfaces are developed with a focus on visual perception (STEPHANIDIS, 2009).

The problems found were categorized according to the method used for evaluation, with this data we identified which methods are most used and what are the limitations. Studies on a web platform have identified that the problems encountered by users are not all covered by accessibility guidelines (FREIRE, 2012; BRAJNIK; YESILADA; HARPER, 2011; BRAJNIK, 2008; RØMEN; SVANÆS, 2012) the (VIGO; BROWN; CONWAY, 2013) study showed how harmful it is to use only the automated method. The result confirmed the results of the studies cited above. The study is further detailed in Chapter 4.

3.3 Comparison user data vs automated tools

This study aimed to compare accessibility problems in mobile applications found by user tests and automated assessment tools. We used data from a previous study involving blind users and users with low vision that evaluated four (DIAS, 2018; CARVALHO et al., 2018) apps: Caixa (Brazilian government app for banks, lotteries, housing simulator), Saraiva (the which allows the purchase of goods, such as books, Receita Federal (Brazilian government application where you can check CPF and income tax) and Decolar (application that sells travel packages, car rentals and hotel reservations). We compared the results of user tests with the results of two assessment tools (Google's Accessibility Scanner and MATE (ELER et al., 2018)) employed to perform the automated evaluation of the same four mobile apps.

This study is further detailed in Chapter 5. This study was approved by the Research Ethics Committee of the Federal University of Lavras, with CAAE code 49781115.9.0000.5148.

3.4 Comparison user data vs manual inspections

This study aimed to compare accessibility problems in mobile applications found by expert inspections and user tests. We considered the problems identified by blind users and users with low vision in the study of Dias (2018), Carvalho et al. (2018). We recruited software development and HCI professionals to perform manual inspections on two of the apps tested by users: Receita Federal and Caixa. We applied observation and interviews to collect data. We compared the types of violations found by each group of participants, and we investigated factors that impact the inspections.

This study is further detailed in Chapter 6. The study was approved by the Research Ethics Committee of the Federal University of Lavras, with CAAE code 41956121.7.0000.5148.

4 SYSTEMATIC MAPPING OF THE LITERATURE

This chapter presents a published systematic mapping study in title (A Systematic Mapping of Accessibility Problems Encountered on Websites and Mobile Apps: A Comparison Between Automated Tests, Manual Inspections and User Evaluations), which analyzed studies that assessed accessibility on the web platforms and mobile platforms (MATEUS et al., 2021). The results in dissertation focus on the data regarding the accessibility of mobile applications.

This chapter is organized as follows: Section 4.1 presents the methods used in this study, Section 4.2 presents all the results obtained, Section 4.3 presents the discussions, Section 4.4 presents the limitations of the study and Section 4.5 presents the final considerations.

4.1 Method

This section explains how we planned and carried out this study. We carried out a systematic mapping study, which aims to identify gaps in the literature, evaluate the findings and analyze the results for a new research question (PETERSEN et al., 2008). The goal of this study is to, we understand the behavior of accessibility assessment methods, and the types of issues that user testing, automated testing, and specialized inspections can cover.

During the mapping, we analyzed 38 articles related to the research question of the dissertation **RQ1**: Is there a difference between the accessibility assessment methods in the mobile context? which can be translated to the following research question.

When performing automated assessments, inspections by experts, and tests with users in web and mobile applications focusing on people with visual impairments, what are problems identified?

We defined the following specific questions to answer the research question:

Q1: Among the problems identified in accessibility evaluations, what are problems found by any combinations of methods?

Q2: What are the benefits and limitations of each method for evaluating accessibility on mobile platforms?

4.1.1 Search strategy

The search string was built to find studies where some accessibility assessment was applied to the web or mobile apps, using automated testing, user testing, and expert inspections:

TITLE-ABS-KEY ((accessibility OR accessible) AND (mobile OR android OR apps OR ios OR talkback OR "talk back" OR "voice over" OR voiceover OR web OR website OR "web site") AND ("visual impairment" OR blind OR blindness OR "visual disability" OR "low vision" OR "partially sighted") AND (evaluation OR assessment OR testing OR test OR inspection OR audit) AND (specialist OR expert OR appraiser OR estimator OR evaluator OR assayer OR manual OR automatic OR automated OR tool OR tools OR user OR users)) AND (LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014))

The search string was used in the Scopus scientific articles repository, which contains the most relevant publications in Computer Science and Human-Computer Interaction. 267 studies were found with this string, and the search was performed from November 14, 2020, to June 13, 2021.

4.1.2 Inclusion and Exclusion Criteria

For including the studies in the systematic literature mapping, the following inclusion criteria were defined:

a) Studies should report assessing the accessibility of Web sites or mobile applications using automated tests, inspections by experts or user tests;
b) Studies should focus or address evaluations targeted at visually impaired users;
c) Studies must explicitly report the types of accessibility problems encountered;
d) The studies' full text must be available through the Brazilian Capes Portal;
e) Studies should report in detail the methods used and the procedures for evaluation;
f) Studies must be published up to February 2021.

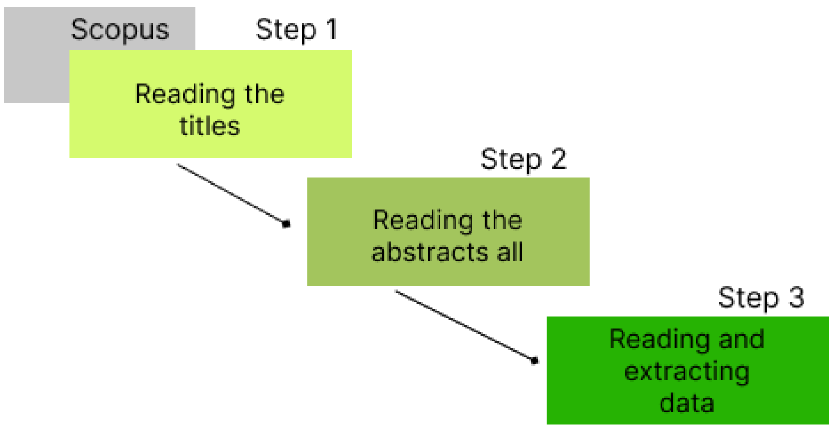
In addition, the following exclusion criteria were defined:

a) Short paper studies with non-detailed presentation of methods used;
b) Studies that only report the number of problems encountered, without qualifying the types of problems.
c) Articles not written in English or Portuguese.

4.1.3 Study Selection

Here are listed the main steps to carry out the systematic mapping as Figure 4.1. The first step is to connect to the Scopus database and execute the search string, which returned two hundred and sixty-seven potential studies. This step also consisted of reading the titles, excluding only those that presented disparity in the title description, moving from one hundred and eight to the next step. However, some assessed web or mobile accessibility used one or more different methods among the selected studies. The Chart 4.1 presents all studies selected in the mobile context. The Table 1 in appendix presents all selected studies in the full reading phase. The entire process of analyzing the studies and extracting the data took place manually, no tools were used.

Figure 4.1 – Step was the execution of the search string in the Scopus database



Source: Own author

The entire process of analyzing the studies and extracting the data happened manually, no tools were used.

a) **Step 1**: The selected database was searched using a previously defined search string. From this, 267 potential studies were found. All titles, 190 studies with potential for systematic mapping were selected for the next step (Step 2), according to the application of the inclusion and exclusion criteria;
b) **Step 2**: Of the 190 studies identified in Step 1, 108 studies were accepted to be analyzed in this stage. The abstracts of these studies were read, again using the inclusion and exclusion criteria;
c) **Step 3**: In this step, of the 108 consolidated studies of Step 2, 38 studies presented information relevant to the topic. From these studies, data were extracted to answer the

research questions. These data were gathered in a spreadsheet to be analyzed.

Chart 4.1 – List of twelve papers

ID	Title	Citation
A2	Accessibility and usability problems encountered on websites and applications in mobile devices by blind and normal-vision	(CARVALHO et al., 2018)
A22	WhatsApp accessibility from the perspective of visually impaired people	(SILVA; FERREIRA; RAMOS, 2016b)
A23	Accessibility of the smart home for users with visual disabilities: An evaluation of open source mobile applications for home	(OLIVEIRA; BETTIO; FREIRE, 2016)
A24	Accessible smart cities?: Inspecting the accessibility of Brazilian municipalities' mobile applications	(CARVALHO et al., 2016)
A26	Heuristic method of evaluating accessibility of mobile in selected applications for air quality monitoring	(ACOSTA-VARGAS et al., 2020b)
A27	Accessibility Assessment of Mobile Meteorological Applications for Users with Low Vision	(ACOSTA-VARGAS et al., 2020a)
A31	Digital equity and accessible MOOCs: Accessibility evaluations of mobile MOOCs for learners with visual impairments	(PARK; SO; CHA, 2019)
A33	Mobile Application Accessibility in the Context of Visually Impaired Users	(SILVA; FERREIRA; SACRAMENTO, 2018b)
A34	Evaluation of tablet PC application interfaces with low vision users: Focusing on usability	(KULPA; AMARAL, 2014)
A35	The interaction experiences of visually impaired people with assistive technology: A case study of smartphones	(KIM et al., 2016)
A37	Accessibility of mobile applications: Evaluation by users with visual impairment and by automated tools	(MATEUS et al., 2020)
A38	The Current Status of Accessibility in Mobile Apps	(YAN; RAMACHANDRAN, 2019)

Source: Own author

4.1.4 Data extraction

After selecting the studies, the data were extracted from analyzed and consolidated. The following data were extracted from each study:

a) Instances of accessibility problems encountered;

b) Type of method used to find each problem;

c) Automated tool used;

d) Number of pages and evaluated;

e) Number of applications evaluated;

f) Specialists' profile involved in manual inspections;

g) Browser used in evaluations;

h) Mobile operating system;

i) Screen readers used in the evaluations;

j) Users' profile who participated in the evaluations;

k) Used smartphones.

We carried out an analysis to consolidate the types of methods used and the types of problems encountered from the data extraction. An analysis of types of problems and a unique category is assigned to the types of problems to make it possible to compare the types of problems encountered by the different methods.

4.2 Results

This section presents the results obtained in the mapping study. We present the accessibility evaluation methods examined and accessibility problems identified. Data identified by the three main methods are shown separately. Similarly, we also summarized the methodological approach used and their characteristics. For example, we discussed the tools involved and the participants' profiles in user studies. Besides, accessibility problems are made explicit, relating them to the studies that identified them. Some accessibility problems were identified by more than one study. Even when it comes to a single evaluation method, there are also problems with accessibility identified by one study only, and there are accessibility issues found on the web and mobile platforms.

4.2.1 Automated Evaluation

Only two studies performed automated tests with a set of three tools, and only one study used more than one tool. The table 4.1 presents the tools used, the number of tools, and the number of applications evaluated. All were performed with the Android system.

Table 4.1 – Automated tools used in the mobile studies

# of tools	Tools used	# of apps evaluated	System Smartphone	Citation
1	IBM Mobile Accessibility Checker	479	Android	(YAN; RAMACHANDRAN, 2019)
2	Acessibility Scanner, MATE	4	Android	(MATEUS et al., 2020)

Source: Mateus et al. (2021)

The problems encountered by these studies had eleven barriers, with related violations: Inappropriate description in controls, target size, insufficient contrast, spacing and inappropriate title, among others. Chart 4.3 shows all the accessibility problems they found and the number of studies they found.

4.2.2 Inspection by Experts

All two studies that performed expert inspections used three or more evaluators. The two studies employed the screen reader Talkback. Table 4.2 shows characteristics related to the methodology used by these studies, explaining the data associated with the method used, including the number of applications, the profile of the specialist, the number of specialists, and the number of tools involved.

Studies involving specialized inspections of mobile applications presented sixteen barriers. The Chart 4.4 presents five problems found related to: form fields that have no labels on their purpose; the user cannot understand what the system expects him to do; components that have the same functionality in a set of web pages are consistently identified. The Table 2 in appendix lists all types of accessibility problems identified in the studies.

4.2.3 User Tests on Mobile Platforms

Of eight studies that performed evaluations with blind users, with low vision, and with normal vision, two of these studies recruited ten or more users, and only two carried out tests in Android and iOS. Table 4.3 presents the characteristics of the methodological approach used in

Chart 4.3 – Accessibility problems encountered by automated evaluation tools

Barrier code	Barrier	Description	Studies
1	Absence of labels	Form fields that have no labels on their purpose.	(MATEUS et al., 2020)
17	Inappropriate description in controls	Controls, such as a link or button, that have an inappropriate description.	(MATEUS et al., 2020; YAN; RAMACHANDRAN, 2019)
67	Target Size	Font size, button.	(MATEUS et al., 2020; YAN; RAMACHANDRAN, 2019)
9	Duplicate information	Content that presents duplicate textual information, such as alternative text for non-text content.	(MATEUS et al., 2020)
12	Insufficient contrast	Bad contrast ratio.	(MATEUS et al., 2020; YAN; RAMACHANDRAN, 2019)
14	Incompatibility of technologies	Incompatible content with screen readers, such as flash	(YAN; RAMACHANDRAN, 2019)
17	Inappropriate description in controls	Controls, such as a link or button, that have an inappropriate description.	(MATEUS et al., 2020)
20	Inadequate navigation sequence	Content that does not allow an adequate navigation sequence by screen readers.	(MATEUS et al., 2020)
41	Visible Focus	The user cannot understand what the system expects him to do.	(MATEUS et al., 2020)
34	Inappropriate title	Page title that does not correctly describe the content.	(MATEUS et al., 2020; YAN; RAMACHANDRAN, 2019)
51	Spacing	spacing between images, text, forms	(MATEUS et al., 2020; YAN; RAMACHANDRAN, 2019)

Source: Mateus et al. (2021)

the studies involving mobile platforms. , It shows the types of systems evaluated, the number of participants and their profiles, and the assistive technologies used.

Users encountered issues such as lack of resources to expand content, inaccessible help links, lack of feedback, inappropriate textual content, and inappropriate title. The Chart 4.5 presents the five problems with the highest number of instances, Absence of alternative text,

Table 4.2 – Characteristics of accessibility inspections by specialists

Name study	# of tools	Tools used - Mobile	# of specialists	Profile of the specialists	# of aplications	Smartphone	Study involved
Accessibility of the smart home for users with visual disabilities: an evaluation of open source mobile applications for home automation.	1	TalkBack	3	2 under-graduate student specialists 1 IHC teacher	6	Android	(OLIVEIRA; BETTIO; FREIRE, 2016)
Accessible smart cities? inspecting the accessibility of brazilian municipalities' mobile applications	1	TalkBack	4	specialist	10	Android	(CARVALHO et al., 2016)

Source: Mateus et al. (2021)

Language not set, Insufficient contrast, Absence of titles, Inadequate navigation sequence. The Table 3 in appendix presents all the problems found.

4.2.4 Problems encountered by different methods

In the analysis of the selected studies, several accessibility problems were raised, resulting from the use of different types of accessibility assessments. There are cases where an accessibility issue was identified by a single method, but there are situations where two or three methods identified accessibility issues. These singular problems do not characterize problematic instances as in the Carvalho et al. (2018) or Power et al. (2012)' studies, as such information was not available in all studies. Thus, considering the accessibility problems identified by the three evaluation methods, we have a total of seventy-one types of problems, fifty six problems were found by a single method. User testing found forty-four types of problems, automated testing found three unique problems, expert inspections nine. The Table Table 4.4 shows the distribution of problems found by the evaluation methods and also those that were found by

Chart 4.4 – Accessibility five problems encountered by expert inspections

Barrier code	Barrier	Description	Studies
1	Absence of labels	Form fields that have no labels on their purpose.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
6	Absence of alternative text	Non-text content that does not have alternative text.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
12	Insufficient contrast	Bad contrast ratio.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
41	Visible Focus	The user cannot understand what the system expects him to do.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
45	Consistent Identification	Components that have the same functionality in a set of web pages are identified consistently.	(CARVALHO et al., 2016)

Source: Own author

more than one method. The Figure 4.2 shows the types of problems encountered in a Venn diagram.

4.3 Discussion

The results obtained show the characteristics of each type of accessibility assessment method for the mobile platform. The benefits and limitations of inspections and tests are presented, providing greater knowledge about the use of applications by people with visual impairments. This corroborates with RQ1 of the dissertation. Following, the text answers the research questions of this chapter.

4.3.1 Problems identified by different methods

Question Q1 was defined as "Among the problems identified in accessibility assessments, what are the problems encountered by some combination of methods?". To answer this question, it is necessary to observe the results presented in Table 4.4 on mobile. Only four

Table 4.3 – Characteristics of user evaluations on mobile apps

Study involved	Number of AT	AT used	# of participants	Profile of the specialists	Apps evaluated	System Smartphone
(CARVALHO et al., 2018)	2	TalkBack and VoiceOver	10	Six blind and four with normal vision	4	Android and iOs
(SILVA; FERREIRA; RAMOS, 2016b)	unavailable	unavailable	5	blind	1	unavailable
(ACOSTA-VARGAS et al., 2020b)	unavailable	unavailable	unavailable	low vision	4	unavailable
(ACOSTA-VARGAS et al., 2020a)	unavailable	unavailable	5	low vision	5	unavailable
(PARK; SO; CHA, 2019)	1	VoiceOver	3	blind graduate student	3	iOs
(SILVA; FERREIRA; SACRAMENTO, 2018b)	1	TalkBack	5	blind	1	Android
(KULPA; AMARAL, 2014)	1	Tablet	5	low vision	1	Android and iOs
(KIM et al., 2016)	1	TalkBack	20	7 blind, 7 visual impaired, 6 normal vision	1	Android

Source: Mateus et al. (2021)

accessibility issues were identified by automated assessments, expert inspections, and user reviews.

Responding to Q1, the results presented in Table 4.4 provide information on the types of problems encountered. Seven accessibility issues were identified by expert inspections and user reviews.

Chart 4.5 – Accessibility five barriers encountered on mobile apps by user evaluations

Barrier code	Barrier	Description	Studies
6	Absence of alternative text	Non-text content that does not have alternative text.	(CARVALHO et al., 2018; SILVA; FERREIRA; RAMOS, 2016b; ACOSTA-VARGAS et al., 2020b; PARK; SO; CHA, 2019; SILVA; FERREIRA; SACRAMENTO, 2018b)
8	Language not set	Content that has no language defined.	(SILVA; FERREIRA; SACRAMENTO, 2018b; KULPA; AMARAL, 2014)
12	Insufficient contrast	Bad contrast ratio.	(KIM et al., 2016; SILVA; FERREIRA; SACRAMENTO, 2018b; KULPA; AMARAL, 2014; ACOSTA-VARGAS et al., 2020a; ACOSTA-VARGAS et al., 2020b; SILVA; FERREIRA; RAMOS, 2016b)
15	Absence of titles	Pages that do not have an identifying title.	(SILVA; FERREIRA; RAMOS, 2016b; SILVA; FERREIRA; SACRAMENTO, 2018b)
20	Inadequate navigation sequence	Content that does not allow an adequate navigation sequence by screen readers.	(CARVALHO et al., 2018; SILVA; FERREIRA; RAMOS, 2016b; ACOSTA-VARGAS et al., 2020b; SILVA; FERREIRA; SACRAMENTO, 2018b; KULPA; AMARAL, 2014)

Source: Own author

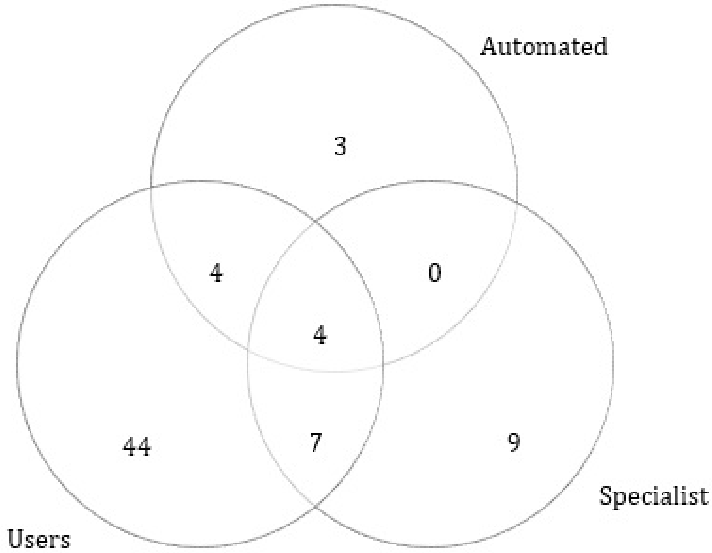
Table 4.4 – Problems found on the mobile platform

Identification	# of Barriers	Percentage
All	4	5.63%
Automated	3	4.23%
Automated and specialist	0	0%
Automated and user	4	5.63%
User and specialist	7	9.86%
Specialist	9	12.68%
User	44	61.97%
Total	71	100.00%

Source: Mateus et al. (2021)

The problems presented in Tables 4.3, 2, and 3 show that the most common violations were: Insufficient contrast, Inadequate navigation sequence, and Visible focus. It is important to

Figure 4.2 – Distribution of coverage of problems on the mobile platform.



Source: Mateus et al. (2021)

highlight that even with few studies using automated tools in the mobile context, the tools identified relevant problems, considering that they had a shorter evolution time than the automated web accessibility assessment tools.

The number of problems found by all methods on the mobile platform is low, this reason is directly linked to the small number of studies that made accessibility assessments through automated inspections in mobile applications, so knowing about the size of the coverage of automated tools is limited.

4.3.2 Problems identified by two methods

Q2 was phrased as "What are the benefits and limitations of each accessibility assessment method on mobile platforms?". According to results presented in Table 4.4, (i) four problems were found by automated inspections and user tests, and (ii) seven problems were identified by specialized inspections and user tests.

The problems encountered by users and experts are related to; Too much information, Lack of resources for expansion, Location, Keyboard, Images, Inappropriate navigation sequence, and Absence of titles.

Thus, results with problems found only by expert inspections and user testing corroborate other results found in the (VIGO; BROWN; CONWAY, 2013) literature, which highlighted the disadvantages of using only automatic assessments and considering the relevance of problems found only with the involvement of users and experts.

4.3.3 Benefits and Limitations of Different Methods

Accessibility assessment tools in mobile applications are characterized by dynamic verification of components, being able to find a more significant number of problems, compared to tools that perform the verification statically (QUISPE; ELER, 2018; ELER et al., 2018). The mobile accessibility assessment tools performed better (ELER et al., 2018) when finding a greater number of instances of violations.

Knowing the limitations of the tool to detect only the presence and absence of (QUISPE; ELER, 2018; ELER et al., 2018), inspections by experts, and inspections by users, (VIGO; BROWN; CONWAY, 2013; BRAJNIK, 2008) becomes necessary. For example, expert inspections help identify accessibility issues that might go unnoticed in user reviews that might not explore specific parts of large systems, and issues that automated tools cannot identify. In the analysis of this study, for example, specialized inspections identified duplicate links and difficulties in finding "help" pages, and inadequate description of content. In addition, specialized inspections can also be applied earlier in the (LAZAR, 2005; FREIRE, 2012) development process, as organizations can organize consultation demands or internal inspections, with less difficulty than the logistics of user reviews.

4.4 Limitations

The work did not identify the use of different accessibility assessment methods in the same application, the types of problems encountered by the different assessment methods have the same characteristics as the problems found in the literature. The number of studies that used automated tools on mobile platforms is still low and with that, we do not have the dimension of what the tools find problems.

4.5 Final Remarks

The results of this chapter show that automated tools have limitations on what types of problems they can encounter, they were able to detect three unique types of problems, users found 44 unique types of problems, and specialists 9 unique problems. As a result, inspection methods with users and specialists are the most appropriate to be used in the final moments of application development. In the next chapter, we will present the results of comparison methods of comparison with users and automated methods through mobile applications.

5 COMPARISON BETWEEN USER EVALUATION OF THE ACCESSIBILITY OF MOBILE APPLICATIONS AND RESULTS FROM AUTOMATED TOOLS

This chapter presents a study comparing data from user evaluation of mobile applications and automated tools of the same applications. The study had been previously published in a conference paper, with the title "Accessibility of mobile applications: evaluation by users with visual impairment and by automated tools" (MATEUS et al., 2020).

The understanding of the comparison helps in the understanding of the relationship between the results from the different methods, how they can be employed in evaluation processes and the limitations and strengths of each method.

This chapter is organized as follows: Section 5.1 presents the methods used in this study; Section 5.2 presents all the results and discussions; Section 5.3 presents the limitations of the study, and Section 5.4 presents the final considerations of this study.

5.1 Methodology

This section presents the methods used to compare accessibility assessment methods for users with disabilities and automated tools. Data collection, categorization of problems and devices used.

5.1.1 Study Design

This study involved comparing problems encountered by users with visual disabilities and violations of accessibility recommendations encountered by automated tools in mobile applications. Several tools use different strategies to automatically check for accessibility violations in mobile applications (SILVA; ELER; FRASER, 2018b): static and dynamic analysis. When it comes to dynamic analysis, the app exploration can be either manual, script-based or automatic.

In this study, two automated tools were employed to perform the automated verifications: Google's Accessibility Scanner and MATE (ELER et al., 2018). Google's Accessibility Scanner is the official tool for mobile accessibility evaluation of the Android platform; it checks for accessibility violations while the user navigates through the app screens (manual exploration). MATE is an open-source tool that covers the majority of accessibility guidelines currently addressed by automated tools; it can be used based on both manual and automatic exploration, in which the tool generates inputs to navigate through the app screens to simulate actual user interaction. This study aimed to perform an initial evaluation to compare results from automated tools with results from user evaluation. Thus, the study employed these two tools due to the possibility of using different approaches and covering more accessibility guidelines during the automated testing. MATE also demonstrated superior performance in coverage of problems in previous studies, compared to other existing tools (ELER et al., 2018).

The evaluations were performed on legacy versions of four mobile apps that had been evaluated by users with visual impairment in a previous study (DIAS, 2018; CARVALHO et al., 2018): Caixa, Receita Federal, Decolar and Saraiva. We aimed to evaluate the same versions that were evaluated in the study published in 2018, to allow for fair comparisons. The previous study that selected such apps considered the variability in terms of public (governmental) and private apps and apps with more and fewer violations of accessibility guidelines, as assessed using manual inspections by specialists. Problems encountered by users were compared to those obtained in a previous study (DIAS, 2018; CARVALHO et al., 2018), with 415 instances of accessibility problems encountered by blind and partially-sighted users. For each problem category, we analyzed whether each tool and users had encountered the same problems.

5.1.2 Participants in the User Study

The six blind users group included five males and one f emale. The age of the participants ranged between 23 and 63 years old, with an average age of 42 years old. The average experience with mobile devices, on a scale from 1 (none) to 7 (a lot of experience) was 6. Regarding the level of education, one participant had a postgraduate degree, two had a higher education degree, two were undergraduate students and one completed high school (DIAS, 2018; CARVALHO et al., 2018). Half of the participants had previous experience with accessibility evaluations.

Low-vision participants included 3 females and two males. The age varied between 20 and 42 years with an average of 31 years. Following the same scale of experience as blind users, the average experience was 5. Regarding the level of education, two completed high school, two were undergraduate students, and one was a graduate student (DIAS, 2018; CARVALHO et al., 2018).

The user evaluation study (DIAS, 2018; CARVALHO et al., 2018) that produced the data used in the present study was approved by the Research Ethics Committee of the Federal

University of Lavras, with CAAE 49781115.9.0000.5148. This research only used processed data from that study, without any access to participants identity and raw data.

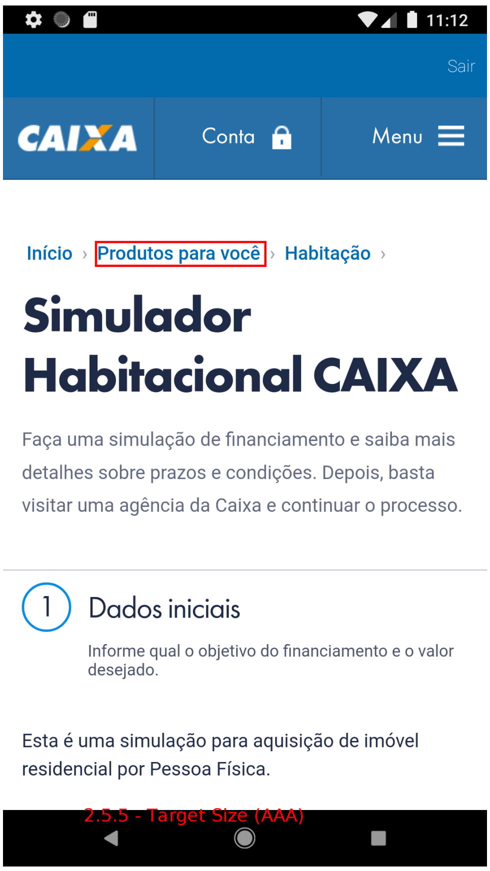
5.1.3 Procedures for the Automated Evaluations

Automatic tests run on a notebook with the Linux Mint operating system, with the Android Studio program with Android 8.0, and the MATE and Accessibility Scanner tools. The manual execution test offered by the MATE tool when performing tasks in the applications allowed the identification of accessibility errors, generating a file containing the images and errors found.

Following are described the procedures used in each tool employed in the study.

MATE (ELER et al., 2018): This is an automated tool for verifying accessibility dynamically for Android applications. When using the tool, each screen accessed from the application under evaluation is checked according to accessibility guidelines. In the end, a file is generated containing the problems found. For example, Figure 5.1 shows where the problem is and the problem description.

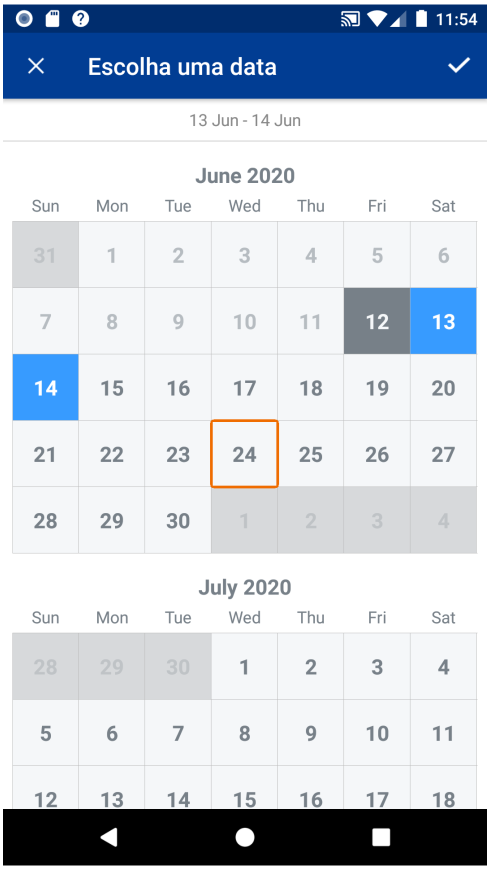
Figure 5.1 – Example of problem found by MATE with a field without alternative description



Source : Mateus et al. (2020)

Accessibility Scanner (GOOGLE, 2015): It is a mobile application on the Android platform, which, when printing the screen of an application in use, generates recommendations for improving accessibility. For example, Figure 5.2 shows where the problem is.

Figure 5.2 – Image of the Acessibility Scanner tool with the identification of the problem and where the problem



Source : Mateus et al. (2020)

5.1.4 Applications Evaluated

The study selected four applications that allowed electronic services to be performed. Two of the applications were from the Brazilian government and two from the commercial sector (DIAS, 2018; CARVALHO et al., 2018):

a) Caixa Econômica Federal: application of the government-owned Brazilian bank, which carries out operations to query the lottery, housing simulation, debt negotiation, and banking services;

b) Receita Federal: application of the Federal Revenue that belongs to the Brazilian government that allows the consultation of the CPF (Individual Taxpayer Registry), consultation of the income tax, a second copy of the CPF, CPF services, and consultation of the income tax;

c) Saraiva: Brazilian trade application that sells products, such as books and others;
d) Decolar: application in the hotel and tourism industry, can be found, in travel packages, car rentals, and hotel reservations, among others.

The versions of the applications are the same ones used in the study of Dias (2018), Carvalho et al. (2018), which can be downloaded for free from the online repository ApkPure[1].

5.1.5 Scenarios for Evaluations

The test scenario that was performed in all assessments was obtained from the study(DIAS, 2018; CARVALHO et al., 2018).

Saraiva Application:

a) Scenario 1: Your godson's 13th birthday is near. Knowing that he loves to read, you want to give a book from the Harry Potter saga of gift for him. Your maximum budget for this gift is R$ 60, including transportation fees. More than one item can be purchased if the budget allow. You should look for books that are in Portuguese whose copy is printed. Find out how much this gift will cost you in total and the deadline;
b) Scenario 2: After choosing the book, you would like to know the ways to payment that Saraiva accepts (for example: which cards, boleto, between others).

Decolar app:

a) Scenario 1: You are planning to travel when the pandemic is over. Look for a travel package that includes a flight (from Belo Horizonte) and hotel for you and an adult companion from today. You want the accommodation to have at least breakfast, internet and swimming pool. Find out how much this trip will cost you, ways of payment, other accommodations / benefits provided by the hotel and its evaluation;
b) Scenario 2: After researching the price of the trip, you realize that you will need a car to walk with your companion. Look for one medium sized car to be used during your stay. find it out how much it will cost you.

Caixa app:

[1] https://m.apkpure.com/br/

a) Scenario 1: New year is coming and extra money would do you good. You decide that luck is with you and think about betting on MegaSena. Caixa Lotteries hold the draw from time to time. Discover the date, prize and the accumulated amount of the next drawing. Discover also the numbers from the last draw so you don't repeat the sequence and how the division of the money collected for each drawing works (who receives the what and how much). note: If MegaSena has no draw, open Quina or similar (main draw);

b) Scenario 2: If you win at MegaSena, you intend to use part of money to finance a new home for you approximately R$400,000.00 in Cabo Frio-RJ. Make a simulation of housing financing Caixa, where you do not have or wish to have link with the box, to learn more about terms and conditions of the financing (input, installments, term and interest). Use the following data: CPF: xxx.xxx.xxx-xx; Gross income: R$ 6,000.00.

Federal Receita app:

a) Scenario 1 (application): You got a new job and want to know as is your situation with the IRS. Make an appointment with your CPF and find out the regularity of it. Use the CPF xxx.xxx.xxx-xx. Note: Service goes down a few times;

b) Scenario 2: You would like to estimate the Income Tax of 2015 to be paid. Whereas you have a monthly taxable income of R$ 4,500.00, have 2 dependents and pay R$ 500.00 of pension find out how much your tax deduction will be for the months.

5.1.6 Comparison of Accessibility Problems Encountered by Tools and by Users

The analysis compared the problems encountered by the tools with the problem instances obtained in tests with users with visual impairment. The categorization of accessibility criteria followed the guidelines in WCAG, with a mapping of the problems encountered by the tools. However, some problems were not directly related to the guidelines. Those problems were classified in the previous study with user tests (DIAS, 2018; CARVALHO et al., 2018), which was adapted from Power et al. (2012).

The problems encountered by users were mapped into issues in the Android version of the apps. Problems were mapped into interface components, enabling a comparison of whether the problems encountered by tools and users referred to the same issue.

5.2 Results and Discussion

This section presents the main results obtained in the study, with the analysis of the problems encountered by the automated evaluation tools and the comparison of such problems with the problems encountered by blind and partially-sighted users in a previous study (DIAS, 2018; CARVALHO et al., 2018).

5.2.1 Accessibility Problems Identified by the Automated Evaluation Tools

Table 5.1 shows that both tools found the highest number of instances of the problem in the Decolar app and found lower instances of problems in the Receita app.

Table 5.1 – Problem instances per app found by automated tools.

	Tools	
App	**MATE**	**Accessibility Scanner**
Caixa	154	103
Receita	65	31
Decolar	466	260
Saraiva	98	142

Source: Mateus et al. (2020)

The MATE tool identified 13 types of problems, as shown in Table 5.2. The problem with the highest number of instances was minimum contrast, with 204 instances corresponding to 25.86% of the problems. The page title problem had the lowest number of instances, 4 in total (0.12%).

The tests performed by the Accessibility Scanner tool found five types of problems, totalling 536 instances. Of these, 38.8% were related to the size of the clickable area and 0.75% of the instances related to minimal contrast (Table 5.3).

5.2.2 Relationship between problems encountered by tools and user tests

This section describes different types of problems encountered only by tools, problems encountered only by users and problems encountered both by users and automated tools.

5.2.2.1 Problems found only by the tools

The tests performed by the Accessibility Scanner and MATE tools found 11 types of unique problems, totalling 638 instances, of which 41% were related to the improved contrast

Table 5.2 – Problems encountered by the MATE tool.

Problems	Caixa	Decolar	Saraiva	Receita	Total
WCAG 2.1 - 1.1.1 Non-text Content	23	48	2	9	82
WCAG 2.1 - 1.3.4 Orientation	7	9	5	4	25
WCAG 2.1 - 1.4.11 Non-text-contrast	11	14	8	3	36
WCAG 2.1 - 1.4.3 Contrast (Minimum)	45	118	28	13	204
WCAG 2.1 - 1.4.6 Contrast (Enhanced)	27	30	17	5	79
WCAG 2.1 - 2.4.2 Page Titled	-	2	1	1	4
WCAG 2.1 - 2.5.5 Target Size	20	44	20	7	91
GENERAL - Phantom Element	5	2	6	-	13
BBC - LINK - Descriptive link (repeated)	3	57	3	4	67
BBC - Label is not defined	4	1	2	1	8
BBC - FOCUS - Managing Focus	4	1	2	1	9
BBC – DESIGN Spacing	2	114	4	12	132
BBC - DESIGN - Actionable Element	3	25	1	4	33

Source: Mateus et al. (2020)

Table 5.3 – Problems found by the Accessibility Scanner tool.

Problems	Caixa	Decolar	Saraiva	Receita	Total
WCAG 2.1 - 2.5.3 Label in Name	21	-	28	1	50
WCAG 2.1 - 3.3.2 Labels or Instructions	8	71	11	2	92
WCAG 2.1 - 1.4.3 Contrast (Minimum)	-	4	-	-	4
WCAG 2.1 - 1.4.6 Contrast (Enhanced)	20	144	1	17	182
WCAG 2.1 - 2.5.5 Target Size	54	41	102	11	208

Source: Mateus et al. (2020)

and 1% to the page title (Table 5.4). The problems with the highest frequencies were: improved contrast, spacing design, descriptive link, accessible label, textural contrast. They were followed by: actionable element design, orientation, phantom element, managing focus, the label is not defined and page with the title.

The tools check for accessibility problems, which include contrasts, non-text content, alternative touch area, and text, among others (SILVA; ELER, 2018). The tools have a limited degree of coverage of only certain types of accessibility problems (SILVA; ELER, 2018; BACH, 2009).

Table 5.4 presents a summary of the problems that were encountered only by the two tools, but not by users.

Table 5.4 – Problems found only by the tools.

Problems	MATE	Acc. Scan.	Total
WCAG 2.1 - 1.4.6 Contrast (Enhanced)	79	182	261
BBC - DESIGN - Spacing	132	-	132
BBC - LINK - repeated link	67	-	67
WCAG 2.1 - 2.5.3 Label in Name	-	50	50
WCAG 2.1 - 1.4.11 Non-text-contrast	36	-	36
BBC - DESIGN - Actionable Element	33	-	33
WCAG 2.1 - 1.3.4 Orientation	25	-	25
GENERAL - Phantom Element	13	-	13
BBC - FOCUS - Managing Focus	9	-	9
BBC - Label is not defined	8	-	8
WCAG 2.1 - 2.4.2 Page Titled	4	-	4

Source: Mateus et al. (2020)

5.2.2.2 Problems encountered only by users

Tests performed with users found 36 types of unique problems, with 305 instances. Of these, 19% were related to headings and labels (Table 5.5). The other problems with the most instances were: name and function and value, inadequate feedback, consistent navigation, focus order, server not working, slow system, resizing text and meaningful sequence.

Table 5.5 – Problems encountered only by users (continuation)

Problems	Blind users	PS users	Med. severity	Total
WCAG 2.1 - 2.4.6 Headings and Labels	45	13	1	58
WCAG 2.1 - 4.1.2 Name, Role, Value	34	22	1	56
3.4.1 Inadequate feedback	33	16	1	49
WCAG 2.1 - 3.2.3 Consistent Navigation	10	7	1	17
WCAG 2.1 - 2.4.3 Focus Order	11	4	3	15
6.1.1 Web Server not working properly	6	9	4	15
6.1.4 System too slow	1	12	2	13
WCAG 2.1 - 1.4.4 Resize text	-	12	1	12
WCAG 2.1 - 1.3.2 Meaningful Sequence	3	3	1	6

Table 5.5 – *Problems encountered only by users (conclusion)*

Problems	**Blind users**	**PS users**	**Med. severity**	**Total**
WCAG 2.1 - 1.3.1 Info and Relationships	4	1	1	5
WCAG 2.1 - 1.4.1 Use of Color	1	3	1	4
WCAG 2.1 - 2.1.2 No Keyboard Trap	-	4	4	4
WCAG 2.1 - 2.4.4 Link Purpose	3	1	1	4
WCAG 2.1 - 3.1.2 Language of Parts	4		1	4
WCAG 2.1 - 3.1.4 Abbreviations	3	1	1	4
WCAG 2.1 - 3.3.1 Error Identification	1	3	1	4
WCAG 2.1 - 4.1.1 Parsing	4		1	4
4.1.5 Inconsistent navigation	4	-	1	4
WCAG 2.1 - 2.4.8 Location	3	-		3
6.1.3 System issues with the assistive technology	3		3	3
WCAG 2.1 - 1.2.3 Audio Description or Media Alternative (Prerecorded)	2	-	2	2
WCAG 2.1 - 1.4.10 Reflow	1	1	2	2
WCAG 2.1 - 3.1.1 Language of Page	-	2	2	2
1.1.2 Irrelevant content	1	1	2	2
2.2.1 No alternative		2	1	2
WCAG 2.1 - 2.1.1 Keyboard	1	-	4	1
WCAG 2.1 - 2.2.2 Pause, Stop, Hide	-	1	1	1
WCAG 2.1 - 3.3.4 Error Prevention (Legal, Financial, Data)	-	1	1	1
WCAG 2.1 - 3.3.5 Help	1	-	1	1
3.3.4 Lack of alternative to data	-	1		1
2.1.5 Default presentation not adequate	1	-	1	1
1.1.3 Users can't make sense of content	-	1	1	1
3.4.3 Element not reachable using assistive technology	-	1	4	1
3.4.5 Functionality not clear	1		1	1
3.4.7 No alternative to functionality	-	1	3	1
4.1.3 No way to return to home page	-	1	4	1

Source: Mateus et al. (2020)

We observed that tests with users found fewer repetitions of problems. However, they can find a higher number of problem types, differently from automatic tools (SILVA; ELER, 2018; BACH, 2009).

Due to the breadth of sensory abilities, user tests are important to find problems from a diversity of users (SILVA; ELER, 2018; PARK; SO; CHA, 2019). We observed that the problems found by users were not fully covered by WCAG 2.1 (W3C, 2018) or the BBC guidelines (BBC, 2014b)

5.2.2.3 Problems encountered by both tools and users

Comparing the problems described by the tools and users, only three types of problems are found by both methods, totalling 642 instances (including those encountered by the tools) (Table 5.6). The problems encountered were related to label or instruction, clickable area size, and minimum contrast.

Table 5.6 – Problems encountered by automated tools and by users.

Problems	MATE	Acc. Scanner	Tools	Blind users	PS users	Median severity	Total - users	Total
WCAG 2.1 - 3.3.2 Labels/ Instructions	-	92	92	23	12	1	35	127
WCAG 2.1 - 2.5.5 Target Size	91	208	299	1	3	1	4	303
WCAG 2.1 - 1.1.1 Non-text Content	82	-	82	7	2	1	9	91
WCAG 2.1 - 1.4.3 Contrast	204	4	208	-	4	1	4	212

Source: Mateus et al. (2020)

Automated tools were more effective to identify problems related to improved contrast, spacing design, descriptive link, accessible label, non-textual contrast, actionable element design, orientation (content is not restricted), phantom element, focus, label not defined and page with the title. They may have failed to identify problems, such as the adequacy of alternative texts, for example (SILVA et al., 2019).

Table 5.7 describes the problems that were found by the methods, with a summary in the Venn diagram in Figure 5.3.

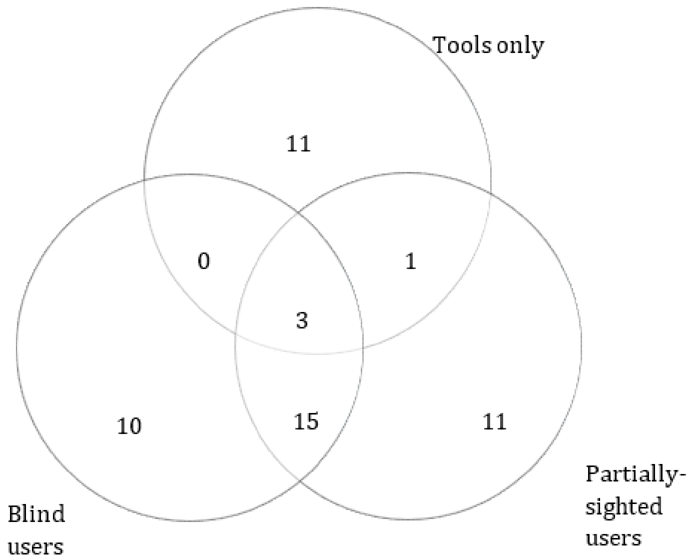
Asserting the frequency of problems encountered by tools and users is still a research gap. For example, clickable area size has many instances from the tools and few of the users. Instances of user problems accounted for only 1.34% of tool instances. Therefore, the question remains whether some of these problems have not impacted users or whether this is a result of the tool's efficiency, as it sweeps the entire system without depending on tests that would spend more time being done by users.

Table 5.7 – Total problems found by each method.

Evaluation Method	Unique Problems - N (%)
Tools only	11 (21.5%)
Tools and blind users	0
Tools and partially-sighted users	1 (1.9%)
Blind users and partially-sighted users	15 (29.6%)
Blind users	10 (19.7%)
Partially-sighted users	11 (21.5%)
All methods and user groups	3 (5.8%)
Total	51 (100%)

Source: Mateus et al. (2020)

Figure 5.3 – Venn diagram with intersections between number of problem types found by different methods and users.



Source: Mateus et al. (2020)

The study also compared the difference between the severity ratings of problems encountered only by users and problems encountered by users that were also covered by the automated tools. The severity ratings were assigned according to the impact the problem had during user evaluations. A Mann-Whitney test was performed comparing the severity ratings of problems encountered only by users and problems encountered by users that were covered by tools, but no significant difference was found (p-value = 0.2219). It is worth noting that the number of problems compared was still limited, so it would be important to deepen the study to establish whether there is a difference between severity ratings. Further to this, it is important to highlight that the types of problems in common that were encountered were still limited, which also restrained the comparison.

5.3 Limitations

Some limitations can occur during automated tests, such as false positives and false negatives. Although the problems were related to the guidelines, there was no verification with a test method by specialists where problems could be found.

5.4 Final Remarks

The main observations of this chapter are: the evaluation methods, user tests, and automated tests found a total of 51 problems. We also observed that the tools found 11 unique problems and that users with low vision also found 11 unique problems. Blind users encountered ten problems. We can see that each method covered a range of problems, reinforcing RQ1's answer. The methods cover different types of problems.

The next chapter advances in the comparison between accessibility evaluation methods comparing tests with users and inspections by professionals. The professionals' group was divided into two: the first is development experts, and the second group is experts in human-computer interaction.

6 COMPARISON BETWEEN USER EVALUATION AND MANUAL INSPECTIONS

This chapter presents an empirical study of the differences between the results of user tests and manual inspections of mobile applications carried out by professionals. Composing the group of professionals, we have specialists in software development (DEV) and human-computer interaction (HCI). Comparing the groups DEV, HCI, and users allows us to understand which limitations each group has and which inspection methods are adequate in various software development cycles.

For the comparison, we used the same user data that was reported in Chapter 5, from a previous study with visually-impaired users (CARVALHO et al., 2018). Users found 38 types of problems, of which the WCAG 2.1 guidelines cover only 26. We created a checklist from these types of problems and distributed it to both groups (DEV and HCI) to guide the evaluation of two mobile apps.

This chapter is organized as follows: Section 6.1 presents the research questions of this study; Section 6.2 presents the study design; Section 6.3 presents the results obtained and discussions; Section 6.4 presents the study's limitations and Section 6.5 presents the final considerations.

6.1 Goals and Research Questions

The goal of this study is to compare the violations found by developers (DEV), HCI professionals (HCI), and users (USER) through qualitative and quantitative analysis. To achieve this objective, we recruited software developers and HCI professionals. The characterization and recruitment strategy of participants are described in Section 6.2.1. We further derived this goal into two Research Questions:

Q1. Is there a difference between the violations found by users, software developers and HCI professionals? - The objective of this research question is to assess whether there is a difference in the results of inspections carried out by the participants of the groups DEV, HCI and tests performed by users (USER).

Q2. What factors impact manual inspections performed by specialists in software development and in HCI? - The objective of this research question is to identify possible factors that impact the inspections performed by participants of the groups DEV and HCI.

6.2 Study Design

In this section, we discuss the data collection process, recruitment, and method used. First, we analyzed the results of the user evaluations of the apps Caixa and Rereceita Federal based on the data discussed in Chapter 5. We classified the issues identified by users as per WCAG 2.1. We then designed a manual inspection task and recruited software developers and HCI professionals for an empirical study. We observed and interviewed these participants as they performed manual inspections. We compared the results of these inspections with the results of user tests and used observation and interview data to identify possible factors that impacted the inspections performed. This study was approved by the Federal University of Lavras' Research Ethics Committee under the code CAAE 41956121.7.0000.5148.

The following subsections describe the recruitment process and profile of participants (Section 6.2.1), inspection procedures (Section 6.2.2) and strategies for data collection and analysis (Section 6.2.3).

6.2.1 Participants and Recruitment

For this study, we considered data from three groups of participants: users (USER), software developers (DEV), and HCI specialists (HCI). We considered the data from user tests described in Chapter 5 and in the user study conducted in the ALCANCE research group (CARVALHO et al., 2018; DIAS, 2018). Therefore, the USER group is composed of eleven participants, six who were blind and five who had low vision.

The recruitment of participants for the DEV and HCI groups took place through the researcher's social circle and social networks. Contact was made by e-mail, WhatsApp and telephone. Participants received the Free and Informed Consent Term (ICF) and all the guidelines and clarifications about the process. Participants who agreed to complete the evaluation responded via e-mail, expressing their acceptance. Given the social distancing, the realization was done remotely through Google Meet. The following inclusion and exclusion criteria were applied:

Inclusion criteria:

a) Be 18 years of age or older;

b) Have knowledge in Human-Computer Interaction OR;

c) Have professional experience in the area of mobile development.

Exclusion Criteria:

a) Participants who do not attempt all assessment tasks will be excluded from the study.

We recruited 27 professionals, 17 software developers and 10 HCI professionals. The DEV group was composed of developers of three professional expertise: Front-end developers, Fullstack developers, and testers.

a) Front-end developers: responsible for creating interfaces;
b) Tester: are responsible for carrying out tests on the system;
c) Full-stack: which can act as a back-end, front-end, and tester.

The following list describes the profile of the DEV group, with the characteristics of each participant of the DEV group.

a) **DEV-01**: Front End junior, 24-year-old woman, graduated in Information Systems, front-developer for 1 year in a private company;
b) **DEV-02**: Front End junior, 38-year-old woman, graduated as a Computer Networks Technologist course worked as a front-end developer for 1 year in a private company;
c) **DEV-03**: Front End junior, a 44-year-old man, who graduated in Physics with a master's degree in statistics and a Ph.D. student in Computational Modeling, working with a front end for 1 year in a public company;
d) **DEV-04**: Mid-senior Front End, the 40-year-old man, who graduated in Information Systems, front-end developer for 6 years in a private company;
e) **DEV-05**: Front End full, the 35-year-old man, graduated as a Technologist in Systems Analysis and Development, and works as a front-end developer for 6 years in a private company;
f) **DEV-06**: Front End full, the 31-year-old man, graduated in Computer Science, works as a front-end developer for 5 years in a public company;
g) **DEV-07**: Tester junior, a 30-year-old man, graduated in Computer Engineering, I work with software testing for 1 year in a private company;
h) **DEV-08**: Tester full, a 25-year-old woman, who graduated in Information Systems. Worked with software testing for 2 years in a private company;
i) **DEV-09**: FullStack junior, the 24-year-old man, graduated in Information Systems, with approximately 1 year as a full-stack in the public sector;

j) **DEV-10**: FullStack junior, the 28-year-old man, graduated in Information Systems and Master's student in Software Engineering, with approximately 1 year as a full-stack in the private sector;

k) **DEV-11**: FullStack junior, the 28-year-old man, graduated in Computer Science and Master's student in Software Engineering, with approximately 1 year as a full-stack in the private sector;

l) **DEV-12**: Mid-senior fullStack, 35-year-old man, graduated in Information Systems with a postgraduate in data science, 5 years in the public sector;

m) **DEV-13**: Mid-senior FullStack, 37-year-old man, graduated in Technologist in Systems Analysis and Development, 5 years in the public sector;

n) **DEV-14**: Mid-senior FullStack full, 31-year-old man, graduated in Information Systems, a graduate student in Big Data, 6 years in the private sector;

o) **DEV-15**: Mid-senior FullStack, 34-year-old woman, Information Systems, master's and Ph.D. in Computational Ontology, 8 years in the public sector;

p) **DEV-16**: Senior FullStack, 52-year-old man, graduated in Business with a postgraduate degree in Web Development, 12 years in the public sector;

q) **DEV-17**: Senior FullStack, 37-year-old man, graduated in Computer Engineering, 9 years in the public sector.

The profiles of the participants of the HCI group, regarding age, gender, academic and profesional background, group are presented as follows.

a) **HCI-01:** 33-year-old woman, who graduated in Information Systems, a postgraduate in IT Management, works as a Development Analyst, and HCI researcher. She gives lectures on web accessibility, has already carried out tests with blind users, and programs websites with accessibility;

b) **HCI-02:** A 50-year-old man who graduated in Information Systems works as an Infrastructure Analyst and an HCI researcher who has already carried out tests with blind users and developed devices for blind users;

c) **HCI-03:** Man, 42 years old, graduated in Technologist in Systems Analysis and Development, works as a Team leader, and has experience with users without disabilities and with blind users in usability evaluation and software tests;

d) **HCI-04:** A man of approximately 40 years, a Professor of HCI and doctoral candidate in HCI, has experience with testing with blind users, and elderly people with low vision, and has published articles in the area of accessibility;

e) **HCI-05:** 27-year-old male, Master in Computer Science IHC, works as an Accessibility Analyst and Teacher, has experience with testing elderly users, and has published articles in the area of accessibility;

f) **HCI-06:** A woman of approximately 40 years, an IHC teacher, and a Ph.D. candidate in IHC. Sh has experience with low literacy user tests and has published articles in the area of accessibility;

g) **HCI-07:** A woman of approximately 40 years, an IHC teacher, and a Ph.D. candidate in IHC has experience with low literacy user tests and has published articles in the area of accessibility;

h) **HCI-08:** A woman who was approximately 35 years old, a Ph.D. student in HCI, has experience with testing elderly users and has published articles in the area of accessibility;

i) **HCI-09:** A woman of approximately 30 years, a Master's student in HCI, has experience with tests of elderly users and children and has published articles with a focus on usability;

j) **HCI-10:** 30-year-old male, Master in robotics, has experience with testes with users without disabitilies, has experience implementing accessibility in projects such as dashboards and websites.

6.2.2 Procedure for Inspections

Based on the problems identified in user tests (DIAS, 2018; CARVALHO et al., 2018), we designed a WCAG checklist (Table B.1 in Appendix B). Participants were assigned with a task that consisted of performing a manual inspection for each application and filling out the checklist provided. We provided the participants with guidelines for filling out the checklist, and we also provided them with links to the BBC mobile and WCAG 2.1 guidelines. Evaluation techniques were not provided. Each participant was supposed to use whichever technique they were familiar with.

Before starting the evaluations, the following guidelines were given to developers: (i) to use the Talkback screen reader on a mobile phone to perform the tests; and (2) participants were strongly encouraged to comment on their actions. For filling out the checklist, participants were instructed to assign an “X” in the columns“Yes” or “No” for each problem category they identify during their inspection. In the “Screen” column, participants should write down 1 for the first scenario and 2 for the second scenario, and in the “Severity” column, participants should mark the severity category of the problem according to those defined by Molich, ranging from 1 to 4 (1 = cosmetic, 2 = minor, 3 = major, 4 = catastrophic).

For these tasks, the participants were asked to evaluate the Receita Federal and the Saraiva apps. They are the same apps used for the study described in Chapter 5, and described in Section 5.1.4. We selected only two of the four apps used, because of limitations to the use of electronic services in the older versions of the apps. The versions of the applications are the same ones used in the study of Dias (2018), Carvalho et al. (2018) which can be downloaded for free from the online repository ApkPure [1].

The test scenario that was performed in all assessments was obtained from the studies of the user study (DIAS, 2018; CARVALHO et al., 2018). The scenarios were the same as the tests compared with the automated tools in Chapter 5 described in Section 5.1.5.

6.2.3 Data Gathering and Analysis

We monitored the inspections performed by each participant individually. Therefore, inspections were carried out remotely using the Google Meet [2] online meeting platform and the Teamviewer[3] remote area sharing software. All professionals performed the inspections using the same device: a notebook with the Linux Mint operating system, and the Android Studio program with Android 8.0. During this process, the professionals were encouraged to report the techniques and strategies used to verify violations, and some questions were asked such as, “which violations were easier to find?”, “Which violations were most difficult to find?”, “ Are the guidelines usable?”. We recorded all interactions for qualitative analysis. The professionals were also free to report violations they found that were not present in the checklist.

For addressing Q1 (**“Is there a difference between the violations found by users, software developers and HCI professionals?”**), we compared the violations found by the HCI and DEV participants to the violations identified in user tests (USER) for the same two apps. Therefore, the violations identified by the USER group consisted of 24 violations, mapped in

---

[1] https://m.apkpure.com/br/
[2] https://meet.google.com/
[3] https://www.teamviewer.com/

Chart 6.1. The analysis consisted of quantitatively comparing the violations found by the HCI and DEV groups, to the violations mapped from the USER group. Considering that the DEV group was composed of participants with different expertise levels, we also mapped the specific violations identified by testers, front-end developers, and fullstack developers.

For addressing Q2 ("What factors impact manual accessibility inspections of mobile apps performed by specialists?"), we considered the qualitative data obtained from observation, from the commentaries of the participants in the execution of the inspections, and from interviews. We applied open coding to synthesize the qualitative data into factors and categories of factors that may have impacted the inspection process performed by the professionals. We applied open coding to synthesize the qualitative data into factors and categories of factors that may have impacted the inspection process performed by the professionals. The lead researcher analyzed the responses and marked relevant segments with "codes" (keyword tagging). Later, the codes were presented to two other researchers and, by consensus, we grouped these codes into relevant categories. With this, it was possible to count the number of occurrences of the codes and the number of items in each category to understand the factors that influence the accessibility inspections pointed out by the participants.

6.3 Results and Discussion

This section presents the results obtained in accessibility inspections carried out by DEV and HCI specialists. Both expert groups completed the checklist stating whether each guideline was violated during the guideline review process. In addition, the qualitative analysis of the data obtained during the tests is presented. In this way, the data is presented to indicate which violations were found and which profiles found them.

Chart 6.1 – Violation related to problems encountered by users in the Caixa and Receita Federal apps

WCAG 2.1 violations	Receita Federal	Saraiva
1.1.1 Non-text content	YES	YES
1.2.3 Audio description or media alternative (pre-recorded)	No	YES
1.3.1 Information and relationships	YES	YES
1.3.2 Significant Sequence	No	YES
1.4.1 Use of Color	No	YES
1.4.10 Reflux	No	YES
1.4.3 Contrast (minimum)	No	YES
1.4.4 Resize Text	YES	YES
2.1.1 Keyboard	No	YES
2.1.2 No keyboard trap	No	YES
2.2.2 Pause, Stop, Hide	No	YES
2.4.3 Focus Order	YES	YES
2.4.4 Purpose of the link (in context)	No	YES
2.4.6 Headers and Labels	No	YES
2.4.8 Location	YES	YES
2.5.5 Target Size	YES	No
3.1.1 Language of the page	YES	YES
3.1.4 Abbreviations	No	YES
3.2.3 Consistent Navigation	No	YES
3.3.1 Error Identification	YES	YES
3.3.2 Labels or instructions	YES	YES
3.3.5 Help	YES	YES
4.1.1 Parsing	YES	No
4.1.2 Name, Function, Value	YES	YES

Source: Own author

6.3.1 Results of the different DEVs groups, HCI group, and User groups (Q1)

Table 6.1 presents all violations encountered by all groups concentrated at the principles perceivable, operable, understandable and robust. Success criteria from principle “Perceivable” start with number 1, principle “Operable” with number 2, principle “Understandable” with number 3 and “Robust” with number 4. Only the HCI and User groups identify violations not covered by the WCAG guidelines. These violations are categorized as “Other”.

The six front-end developers detected 20 types of WCAG-related problems. Table 6.1 concentrated on the principles perceivable (8), understandable (7) and operable (5). It was surprising to see that no problems were encountered related to guidelines concerning the robust principles. The problems found according to the success criterion from principle “Perceivable”

Table 6.1 – Violations by WCAG principles encountered by all groups

Principles	Violated guidelines - Front	Violated guidelines - Tester	Violated guidelines - Fullstack	Violated guidelines - HCI	Violated guidelines - User
Perceivable	8	6	8	8	8
Operable	7	6	8	8	8
Understandable	5	7	7	8	6
Robust	0	2	2	2	2
Other	0	0	0	8	14

Source: Own author

start with number 1, non-text content, audio description or media alternative (pre-recorded), information and relationships, significant sequence, use of colour, reflux, contrast (minimum), resize text. The principle “Operable” has criteria starting with number 2: keyboard, no keyboard trap, pause, stop, hide, focus order, focus order, headers and labels, location, and target size. The principle “Understandable” has criteria starting with number 3, the language of the page, the language of parts, error identification, and error prevention (legal, financial, data). The principle “Robust” has criteria starting with number 4. No participants from the Front-end DEV group encountered problems with this principle. The front-end developers did not encounter the following violations: the purpose of the link (in context), consistent navigation, labels or instructions, help, parsing, and name, function, and value. Table 5 in the appendix has complete information about what each DEV front-end encountered types of problems.

The two testers were able to detect 21 types of violations related to WCAG. Table 6.1, focused on perceptible (6), operable (7), understandable principles (6) and robust (2). The problems found according to the success criteria from principle “Perceivable” principle has criteria starting with number 1: non-text content, significant sequence, use of colour, reflux, contrast (minimum), and resize text. The “Operable” principle has criteria starting with number 2: no keyboard trap, focus order, purpose of the link (in context), headers and labels, location, and target size. The “Understandable” principle has criteria starting with number 3: language of the page, language of parts, consistent navigation, error identification, labels or instructions, error prevention (legal, financial, data), and help. The "Robust" has criteria starting with number 4: parsing, and name, function, and value. The DEV tester did not encounter the following violations: abbreviations, no keyboard trap, resize text Information and relationships, audio description or media alternative (pre-recorded). Table 6 in the appendix has complete information about what types of problems were encountered by each DEV tester.

The nine full-stack developers were able to detect 25 types of WCAG-related problems. Table 6.1, concentrated at the WCAG principles perceivable (8), operable (8), understandable (7), and robust (2). It was surprising to observe that no problems were encountered at understandable and robust levels. The problems found according to the success criteria from principle "Perceivable" principle starts with number 1: non-text content, audio description or media alternative (pre-recorded),information and relationships, meaningful sequence, use of colour, reflux, contrast (minimum), and resize text. The "Operable" principle has criteria with number 2, keyboard, no keyboard trap, pause, stop, hide, focus order, purpose of the link (in context), headers and labels, location, and target size. The "Understandable" principle has criteria starting with number 3: language of the page, language of parts, consistent navigation, error identification, labels or instructions, error prevention (legal, financial, data), and help. The "Robust" principle has criteria starting with number 4: parsing, and name, function, value. The DEV fulsstack did not encounter the following: abbreviations. Table 7 in the appendix has complete information about what types of problems were encountered by each DEV fullstack.

The ten professional HCI were able to detect 26 types of WCAG-related problems. Table 6.1 concentrated at the WCAG principles perceivable (8), operable (8), understandable (8), and robust (2). It was surprising to observe that no violations were encountered at understandable and robust levels. Success criteria from principle "Perceivable" start with number 1, principle "Operable" with number 2, principle "Understandable" with number 3 and "Robust" with number 4. The violations found in principle (1) were: non-text content, audio description or alternative media (pre-recorded), information and relationships, significant sequence, use of colour, reflux, contrast (minimum), and resize text. Violations found in principle (2) were: keyboard, no keyboard trap, pause, stop, hide, focus order, purpose of the link (in context), headers and labels, location, and target size. Violations found in principle (3) were: language of the page, language of parts, abbreviations, consistent navigation, error identification, labels or instructions, error prevention (legal, financial, data), and help. Violations found in principle (4) were: parsing, and name, function, and value. Table 8 in the appendix has complete information about what types of problems were encountered by each professional HCI.

It is worth mentioning that HCI professionals encountered problems that were not listed in the checklist, which are: Broken links, Duplicate links, layout problems, wrong form entries, lack of masking in numeric fields, lack of identification of items per page, lack of adequacy of calendars and lack of descriptive icons. In addition, it was the only group that defined the severity for each problem found.

All violations found by the DEV group, HCI group, and User group are present in Chart 6.2. We observed that, in an aggregation of the results by group, only the HCI group found all expected violations, followed by the DEV full-stack. These two groups of professionals found approximately the same amount of violation that users. On the other hand, front-end DEVs found fewer violations. An interesting observation is just the HCI group found the Abbreviations violation.

Responding to **Q1**, the front-end DEV professionals found violations in three principles, mainly in the “Perceivable” principle. These violations were directly linked to their area of expertise. They did not find any violations related to the principle “Robust”. Only two DEV testers found one more violation than the front-end DEVs. They found violations in all principles. Full-stack DEVs encountered the highest number of violations, totalling 25. Table 6.1 shows the number of violations by principles.

The User group found 38 types of violations and the HCI group found 34 of them. Some violations were not covered by WCAG 2.1. Table 6.1 presents the number of violations by WCAG principles and others are those that are not covered by the guideline. These results corroborate with other studies that (RØMEN; SVANÆS, 2012; POWER et al., 2012; CARVALHO et al., 2018) these two groups are able to identify violations not covered by the WCAG Guideline.

Chapter 4 helps to confirm the results of this section. The 5-year sample of the literature confirmed that HCI specialists and users with disabilities encounter a wider range of accessibility problems.

6.3.2 Factors that influence the inspection process (Q2)

This section presents a qualitative analysis involving DEV and HCI professionals, on the factors that influence the accessibility inspection process. The factors were divided into five

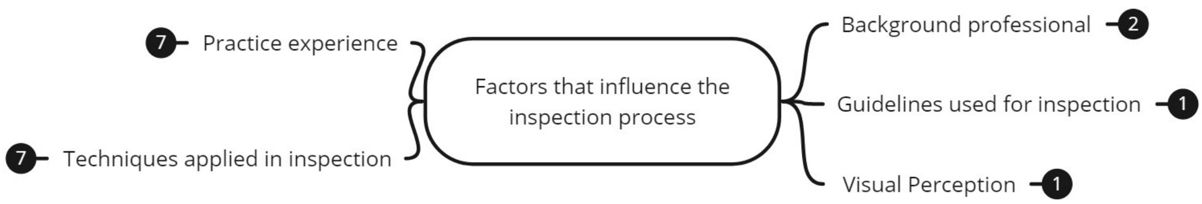
Chart 6.2 – Violation encountered by WCAG, DEV group and IHC group

WCAG 2.1 violations	**Front**	**Fullstack**	**HCI**	**Tester**	**User**
1.1.1 Non-text content	Yes	Yes	Yes	Yes	Yes
1.2.3 Audio description or media alternative (pre-recorded)	Yes	No	Yes	No	Yes
1.3.1 Information and relationships	Yes	Yes	Yes	Yes	Yes
1.3.2 Significant Sequence	Yes	Yes	Yes	Yes	Yes
1.4.1 Use of Color	Yes	Yes	Yes	Yes	Yes
1.4.10 Reflux	No	Yes	Yes	Yes	Yes
1.4.3 Contrast (minimum)	Yes	Yes	Yes	Yes	Yes
1.4.4 Resize Text	Yes	Yes	Yes	No	Yes
2.1.1 Keyboard	Yes	Yes	Yes	Yes	Yes
2.1.2 No keyboard trap	No	Yes	Yes	No	Yes
2.2.2 Pause, Stop, Hide	No	Yes	Yes	Yes	Yes
2.4.3 Focus Order	Yes	Yes	Yes	Yes	Yes
2.4.4 Purpose of the link (in context)	Yes	Yes	Yes	Yes	Yes
2.4.6 Headers and Labels	No	Yes	Yes	Yes	Yes
2.4.8 Location	No	Yes	Yes	Yes	Yes
2.5.5 Target Size	Yes	Yes	Yes	Yes	Yes
3.1.1 Language of the page	Yes	Yes	Yes	No	Yes
3.1.2 Language of Parts	No	Yes	Yes	No	No
3.1.4 Abbreviations	No	No	Yes	No	Yes
3.2.3 Consistent Navigation	No	Yes	Yes	Yes	Yes
3.3.1 Error Identification	No	Yes	Yes	Yes	Yes
3.3.2 Labels or instructions	No	Yes	Yes	Yes	Yes
3.3.4 Error prevention (legal, financial, data)	No	Yes	Yes	Yes	No
3.3.5 Help	No	Yes	Yes	Yes	Yes
4.1.1 Parsing	No	No	Yes	Yes	Yes
4.1.2 Name, Function, Value	No	Yes	Yes	Yes	Yes

Source: Own author

themes, namely: Professional background; Guidelines used for inspection; Practice experience; Visual perception; and Techniques applied in inspection. Figure 6.1 shows an overview of these factors with their categories. In the following subsections, the five factors are detailed. The categories used in the coding that were grouped under each theme are highlighted in the description.

Figure 6.1 – The five factors that interfere in the inspection process

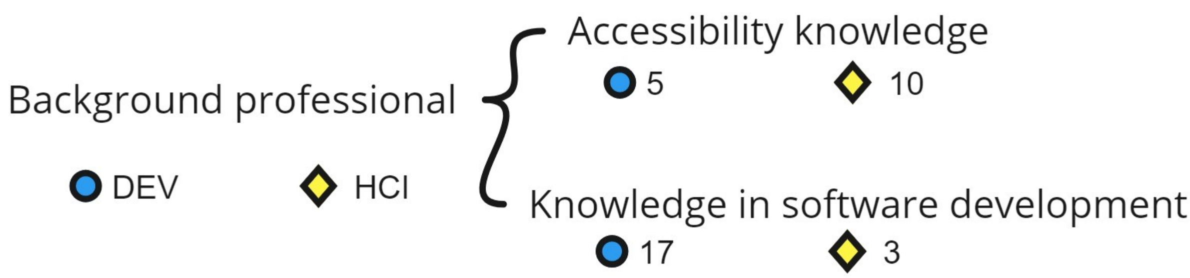


Source: Own author

6.3.3 Influence of professional background

Vocational training is linked to tacit and explicit knowledge acquired along the professional trajectory. This knowledge is linked to accessibility and software development. Figure 6.2 shows the categories linked to the professional background.

Figure 6.2 – Professional Background Theme



Source: Own author

DEV professionals reported having knowledge as **partial in accessibility**. Some understood that accessibility was in having buttons to increase and decrease the font, or having an auto contrast button. HCI professionals reported having mastery in **accessibility**, through academic research or professional experience. DEVs also demonstrated knowledge of software development. Following are excerpts from the responses of professionals DEV-15 and HCI-07 regarding their backgrounds.

> ... I work for the state administration and until then we intended that accessibility on the site was to put the A+ (increase letter) and A- (decrease letter) in the letter, make a contrast and such putting a little thing and another that is what we usually see. (DEV-15).

> ... Today I work with users with low literacy in my doctoral research, I am a professor of HCI. (HCI-07).

Another factor of professional background is knowledge in software development, DEVs and HCI reported having theoretical and practical knowledge in **software development**. This knowledge helped them in the inspection process.

6.3.4 Do the guidelines used for inspection influence the results?

In the inspection process, the guidelines used in the inspection interfered with the results. The professionals reported the complexity of the guidelines. Figure 6.3 presents the category in this theme.

Figure 6.3 – Usability of guidelines

Source: Own author

Professional DEVs reported that the **guidelines are complex**, especially WCAG. The items seemed to have the same meaning. Participants also reported that the BBC had **practical examples**, and the understanding was very clear regarding the problem. Following is the response from participants on the usability of the DEV-15 guidelines.

> ... Correctly identify the guideline items within the application, I didn't understand sometimes, because it was not very clear understanding of that item, sometimes they clashed with other items. (DEV-15).

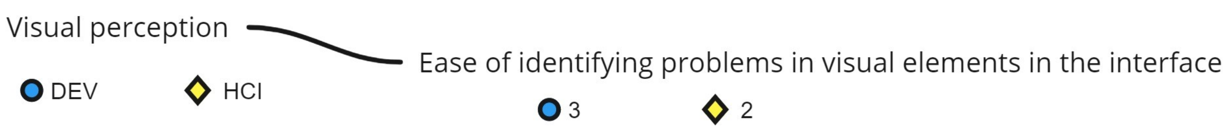
The HCI professionals also reported that the description of the guideline items are similar, bringing **confusion to the understanding of the item**. Following is the report of the professional HCI-05.

> ...the WCAG guideline I found a bit confusing, it has similar terms with others. (HCI-05).

6.3.5 Does visual perception contributes to problem detection?

Some problems can be identified visually, and professionals reported problems regarding visual elements. Figure 6.4 presents the categories linked to this theme.

Figure 6.4 – Visual perception



Source: Own author

DEV professionals reported having trouble spotting visual interface elements (e.g, **large banner without alt text**), **lack of number of items per page**, and lack of **masking in numeric fields** problems. Following are reports from professionals DEV-08 and DEV-03.

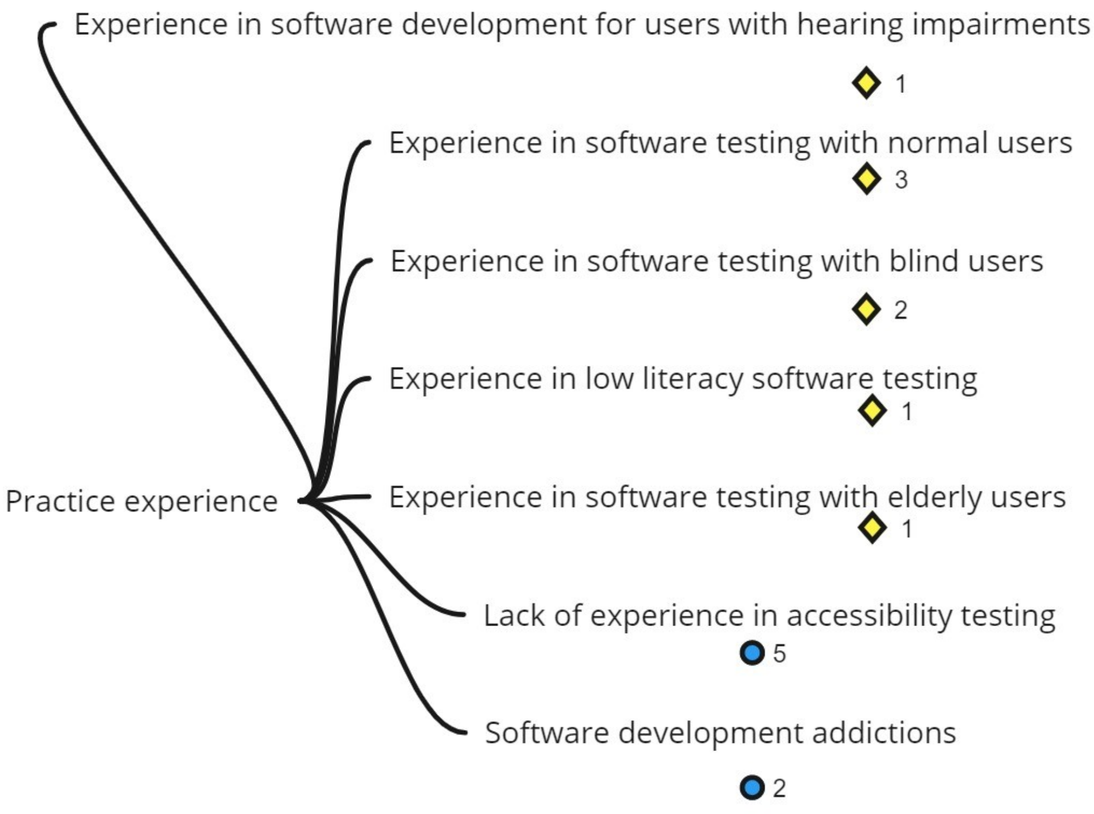
> ... The non-text content was very easy to identify as it had the banner passing through. (DEV-08).
>
> ...The easiest problem I found was the non-text content, it has banner passing. (DEV-03)

HCI professionals also reported problems related to visual elements, **layout breakage**, and **lack of mask** in CPF field.

6.3.6 Hands-on experience in inspection processes or software development

During the inspection, professionals report having practical experience in software development, in accessibility tests with blind users, the elderly and others. Figure 6.5 presents the categories linked to the topic of practical experience.

Figure 6.5 – Experience pratice



Source: Own author

Only the HCI group reported having experience in software development for elderly and **hearing-impaired** users. They also reported having experience in testing with **blind users**, **elderly users**, and users with **low literacy**. These experiences came through academic research and professional experiences in public and private sector organizations. Below are reports from professionals HCI-07, and HCI-09.

> ... I have already carried out tests with users, and today I work with users with low literacy in my doctoral research, I am a professor of HCI. (HCI-07).

> ... I have experience in HCI, I have publications in the area, and I work with usability for children and the elderly. (HCI-09).

Some DEV professionals reported having **no experience** with accessibility before the inspection in the study. They had a macro view of usability problems, but it was **not aimed at users with disabilities**. Following is the report of the professional DEV-16.

> ... The tests were aimed at users as a whole, I didn't have this view of the limitation of some users with accessibility issues. (DEV-16).

Professionals also reported having experience in **software development**. With this experience they reported having **greater power** in identifying problems related to **poorly developed coding**, bugs caused by **improper error handling**, and **lack of error handling**. Following is the report of the professional DEV-17.

> ... Some crashes and bugs caused by improper error handling and system boot errors, due to poorly developed coding. (DEV-17).

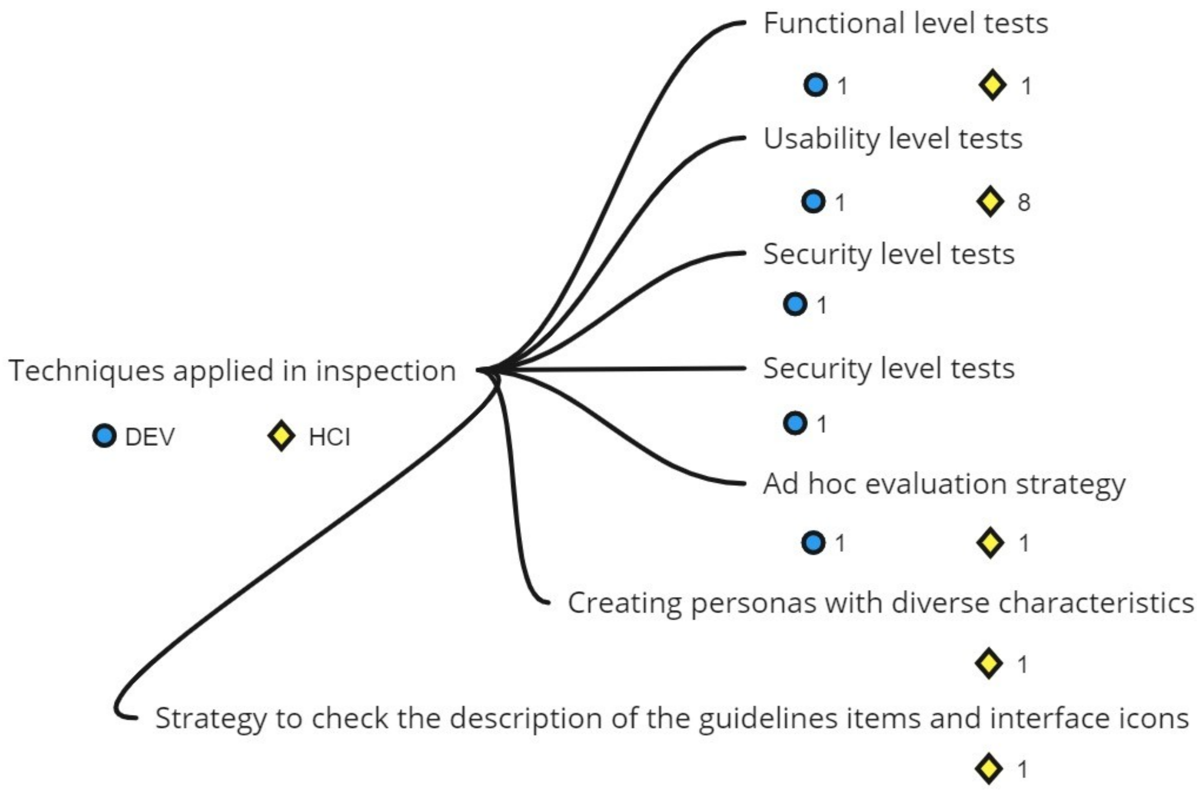
Professionals also **reported frustration** in finding so many accessibility problems in applications, they also reported satisfaction in acquiring **practical experience** in the subject after inspections. Below is the reports of the professional DEV-15.

> ... It was stressful because nothing works right. (DEV-15).

6.3.7 Techniques applied in the inspection

Professionals reported having several techniques to carry out inspections. Figure 6.6 shows the techniques used in the inspection process.

Figure 6.6 – Techniques applied in inspection



Source: Own author

Several professionals use techniques of their domain to carry out inspections in this context. There was the **Ad hoc evaluation strategy**. Professionals reported that they **do not have a defined technique**, that as they go through the task, they identify problems. Following is the report from the professional HCI-07.

> ... I don't have a strategy to test, I'm doing the task and identifying the problems. (HCI-07).

Most professionals used usability techniques to check navigation. They reported turning **off the monitor** and navigating using the screen reader. There were also those who tried to navigate using the **TAB key to jump** between contents, this same technique was used to identify problems related to focus. Following is the report from professional HCI-02.

> ... I activated the screen reader and turned off the monitor, unable to navigate in Saraiva's application, I didn't know where the element was, finally turned on the monitor and continued the tests. (HCI-02).

Professionals reported that some of the accessibility violations can be identified by performing functional tests, for example the calculate shipping **button does not work**, the **back button does not work**. Following is a quote from the professional HCI-05.

> ... The calculate shipping button is not working, it does not return the calculation. (HCI-05)

The HCI professional reported the **creation of personas** to carry out the inspections, each persona has a different characteristic, and with that it brings the obligation to **inspect with a different look**, making it find **different problems** whenever **testing the usability** of the applications.

> ... I create personas to test usability in ways that I can identify issues with looking different. (HCI-05)

Completing the **Q2** answer, the qualitative analysis showed that the difference between the DEV group and the HCI group is in the strategies used to inspect the applications, the DEVs focused on the code and techniques that analyze the written code. The HCI professionals showed a greater repertoire of usability techniques, especially those that interfere with the user experience. For example, almost all of which identified navigation problems in the tested applications were identified by them.

6.4 Limitations

The COVID-19 pandemic (OMS, 2021) interfered with the process of recruiting and inspecting professionals. The study did not compare the instances of problems, and the number of DEVs and experience was not entirely balanced, so that the results may interfere with the comparison between the DEV group. Another limitation is that the study only addressed the violations covered by the WCAG guideline.

6.5 Final Remarks

The main findings of this chapter were the techniques used by the HCI group in the inspection process navigation strategies. Only the HCI and User groups identified violations not covered by the WCAG guidelines. The other fact is the low amount of violations found by the front-ends, even with four more professionals than the groups formed by the tester. Doing justice to their profession's skills, two tests found 21 violations of guidelines, two more than the front-end ones.

7 CONCLUSIONS AND FUTURE WORK

Considering the growth in the use and importance of mobile applications, the study reported in this dissertation aimed to identify if there is a difference between assessment methods, namely: automated tools, user testing, and inspections by professionals. It also aimed to understand if there is a difference between expert inspections by specialists in software development and expert inspections with HCI specialists.

The dissertation defined two main research questions:

RQ1: Is there a difference between the accessibility evaluation of mobile applications performed by automated tools, inspections by professionals and user tests?

RQ2: Is there a difference between accessibility inspections of mobile applications performed by software development specialists and HCI specialists?

7.1 Findings

To answer the research questions, the research design employed three studies: 1) a systematic mapping of the literature covering different types of accessibility evaluations of mobile applications; 2) an empirical study comparing the results of automated tests with the results of a previous study that conducted tests with users with visual impairment (CARVALHO et al., 2018), and 3) an empirical study comparing the results from inspections by professionals (specialists in software development and specialists in HCI) with the results from the same user evaluation study.

The findings from the systematic mapping of the literature reported in Chapter 4 pointed out that the different accessibility evaluation methods interfere with obtaining inspection results. The results from the literature provided initial evidence that automated tests make it possible to perform tests quickly. However, it has limitations concerning the number of types of problems found.

The literature also confirmed that, in mobile applications, manual inspection methods make it possible to identify a greater number of issues that are not covered by accessibility guidelines. They help to identify problems by specialists, for example, in the graphical interface and code. However, if we compare automated evaluations, they have a higher cost and still cannot reveal all problems that user testing with users with disabilities could yield.

User tests are performed with actual tasks and real users with disabilities who have specific mental models and perceptions of how to use certain interfaces. Users can identify real problems in the application, ensuring good robustness. This method requires a higher cost and is sometimes chosen as the last test to be performed.

Despite the preliminary evidence, the investigated literature reported in Chapter 4 did not provide depth in understanding the contexts in which such problems were found and direct mapping comparing problems encountered in the same applications evaluated. These issues were further investigated in the studies reported in Chapter 5 and Chapter 6.

The relationship between the accessibility problems encountered by automated tools and user testing was investigated in Chapter 5. The study compared data from user evaluations of four mobile applications Caixa, Receita Federal, Saraiva, and Decolar, and the results from tools Accessibility Scanner and MATE. The results showed 51 types of problems of which 11 were found only by the tools, ten only by blind users and 11 by low vision users. Only three problems were found by blind users, low vision users, and tools.

When comparing the severity ratings of issues found only by users and issues encountered by users that were also covered by the automated tools, no significant numerical difference was found, considering the limitations of this study.

Chapter 6 presented a comparison between software development (DEV) professionals (front-end, tester, and full-stack), HCI professionals, and user groups. In addition, the chapter presented a qualitative analysis between the DEV groups and the HCI group on the factors that influence the way they perform accessibility inspections. The professionals performed the guideline review process based on problems of 24 WCAG 2.1 guideline violations.

Among the DEV groups, front-end professionals could not find violations related to the “Robust” principle (4) Parsing Name, Function, and Value. The group formed by tester professionals found violations in all principles but did not find all the expected violations. The group formed by full-stack found all the expected violations but did not find any violations other than those present in the checklist, the group formed by HCI professionals managed to find all the expected violations, and in addition, they also found issues not covered by the WCAG 2.1 guidelines, as well as the user group.

Concerning the factors that interfere with the accessibility inspection process, few DEV professionals showed partial or higher knowledge of accessibility. Participants from the DEV group also based their entire inspection process on functional and security level tests. Few

DEVs used usability testing strategies. For many, this was their first contact with accessibility guidelines. The HCI professionals demonstrated their knowledge of accessibility through academic research and professional careers. These professionals demonstrated several guidelines verification strategies and a wide range of usability inspection skills. Some used persona creation for testing, even when they were not provided in the empirical study.

Chapters 4, 5, and 6 provided different types of evidence to address the main objective of analyzing the types of problems found by automated tools, inspection by professionals (involving developers and specialists in human-computer interaction (HCI)) and user tests of mobile application accessibility, focusing on visually impaired users.

As a specific objective, an investigation of the types of problems encountered by different automated accessibility assessment tools in a mobile context was carried out, resulting in the identification of 34.61% of problems that can be automated from the WCAG 2.1 guidelines. This identification will depend on the type of tool used. Of the tools used in this study, MATE had the best performance concerning the number of instances and types of problems encountered. In addition, it has verification in the WCAG 2.1 and BBC guidelines.

In the second specific objective, the study showed differences between the types of accessibility problems that inspections by professionals (front-end and full-stack developers, testers, and HCI specialists with accessibility experience) can find. Front-end professionals found fewer violations of guidelines linked to the graphical interface. Testers found only one more violation than front-end professionals, and full-stack professionals performed worse than HCI professionals. The biggest highlight regarding domino in accessibility techniques was HCI professionals finding a wide range of violations that are not covered by the WCAG 2.1 guidelines.

The study had a third specific objective, comparing the problems found by automated tools and specialized inspections with those found by visually impaired users. The automated tools detected only presence and absence, limiting their performance against violations found by the experts, from 24 violations expected by the tools to WCAG-related violations. Of the nine violations, six were not found by specialists. The performance of these experts was more assertive than the tools concerning the number of types of problems.

Answering the research question **RQ1**: Is there a difference between accessibility evaluation methods in the mobile context? The study showed a difference between accessibility evaluation methods in the mobile context. Comparing the different accessibility inspection methods, the automated tools found fewer types of problems. Of the guideline items, approximately 34.61% are automatable (ELER et al., 2018; SILVA; ELER, 2018), tests with users with disabilities, inspections, and automated assessments have different coverage (JAEGER, 2006). Thus different ways of evaluating impact the results (BRAJNIK; YESILADA; HARPER, 2011). Of the 31 WCAG guideline violations, the tools find eight types of violations, non-text content, orientation, non-text-contrast, contrast (minimum), contrast (Enhanced), page title, label in name, target size, and labels or instructions. Experts found 27 violations, and users revealed 24 violations.

Experts and users found greater numbers of problems, according to the results from the systematic mapping of the literature reported in Chapter 4. The empirical results from this study confirmed results from previous studies (MARTINS et al., 2022; MATEUS et al., 2022; LOPES; FAÇANHA; VIANA, 2022; AIZPURUA et al., 2014; VIGO; BROWN; CONWAY, 2013; SOUZA; CARDOSO; PERRY, 2019) and provided deeper insight into the types and contexts in which problems were encountered.

Responding to **RQ2**: Is there a difference between accessibility inspections of mobile applications performed by development specialists and HCI specialists? The study showed that there are notable differences. Like in the study by Mankoff, Fait e Tran (2005), the DEV group demonstrated less mastery of accessibility inspections, and only scarce knowledge in accessibility. Therefore, they found fewer instances of violations. In the study by Mankoff, Fait e Tran (2005), they recruited developers with no experience in accessibility. This study did not recruit developers based on their knowledge of accessibility. However, the sample has similar characteristics compared to the sample recruited by Mankoff, Fait e Tran (2005).

7.2 Limitations

The new coronavirus pandemic contributed to some limitations of the dissertation. The inspection procedures with the DEV and HCI professionals had to happen remotely. However, because they were in isolation, the professionals did not allocate enough time for the inspections to have found the number of instances of problems. In this way, they found only the types of problems. For this reason, it was not possible to obtain severity ratings for each problem encountered to perform comparisons such as those performed by Harrison e Petrie (2007).

The second limitation was related to the group formed by software developers. The DEV group was not fully balanced in terms of experience and areas of development, even if we consider the group as a whole. For this reason, it was not possible to perform statistical tests.

The third limitation was not testing the inspection methods in practice in the development cycle.

Considering that there was a gap between the time when the applications were evaluated by visually impaired users (CARVALHO et al., 2016), some versions of the applications were not functional any more. This prevented the study from using the whole dataset from the user evaluations in the comparison with inspections by professionals.

7.3 Contribution

The studies conducted as part of this dissertation resulted in several publications in national and international conferences and journals, described as follows:

a) **Mateus, D. A.**, Silva, F. A. C. D., Silva, T. S. D., Freire, A. P. (2022). Evaluation methods in legal procedures concerning digital accessibility in Brazil: an analysis of cases investigated by the federal public ministry. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems* (pp. 1-12). (*honourable mention*);

b) **Mateus, D. A.**, Silva, C. A., de Oliveira, A. F., Costa, H., Freire, A. P. (2021). A Systematic Mapping of Accessibility Problems Encountered on Websites and Mobile Apps: A Comparison Between Automated Tests, Manual Inspections and User Evaluations. *Journal on Interactive Systems*, 12(1), 145-171;

c) **Mateus, D. A.**, Aparecido da Silva, F., Freire, A. P. (2021). Pandemic Crisis Brings More Digital Governmental Services to Mobile Devices-But Are They Accessible to People with Disabilities?. In The 39th *ACM International Conference on Design of Communication* (pp. 197-204);

d) **Mateus, D. A.**, Silva, C. A., Eler, M. M., Freire, A. P. (2020). Accessibility of mobile applications: evaluation by users with visual impairment and by automated tools. In Proceedings of the 19th *Brazilian Symposium on Human Factors in Computing Systems* (pp. 1-10).

The following papers are also in preparation for submission:

a) **Mateus, D. A.**, Souza, M. R. A. Freire, A. P. Comparison between accessibility problems encountered by users with visual impairment and inspections by software development and accessibility specialists. To be submitted to the 19th *International Conference of Technical Committee 13 (Human- Computer Interaction) (INTERACT 2023)*;

b) **Mateus, D. A.**, Silva, F. A. C., Rigatto, S. H., Silva, T. S., Souza, M. R. A. Freire, A. P. The Legal Handling of Digital Accessibility: a Comparison of Evaluation and Policy Approaches in Federal-Level Cases in Brazil and the United States. To be submitted to the *Journal of the Brazilian Computer Society*, (2023).

7.4 Future work

As future work, we intend to conduct studies to verify the impacts caused in the applications that used the automated tests and inspections with specialists in the initial stages of the software development cycle, in the final stage of the software development cycle, carry out the tests with users.

We also intend to conduct studies with users with different types of disabilities, different types of mobile applications and a wider group of specialists. This way, we intend to perform statistical analysis as follow-up studies from the initial findings revealed in this study.

REFERENCES

ABDOLRAHMANI, A.; KUBER, R. Should i trust it when i cannot see it? credibility assessment for blind web users. In: **Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility**. New York, NY, USA: Association for Computing Machinery, 2016. (ASSETS '16), p. 191–199. ISBN 9781450341240. Disponível em: <https://doi.org/10.1145/2982142.2982173>.

ACOSTA-VARGAS, P. et al. Accessibility assessment of mobile meteorological applications for users with low vision. In: **International Conference on Applied Human Factors and Ergonomics**. Springer, 2020. p. 199–205. Disponível em: <https://doi.org/10.1007/978-3-030-51369-6_27>.

ACOSTA-VARGAS, P. et al. Heuristic method of evaluating accessibility of mobile in selected applications for air quality monitoring. In: NUNES, I. L. (Ed.). **Advances in Human Factors and Systems Interaction**. Springer International Publishing, 2020. p. 485–495. ISBN 978-3-030-20040-4. Disponível em: <https://doi.org/10.1007/978-3-030-20040-4_44>.

ACOSTA-VARGAS, P.; SALVADOR-ULLAURI, L. A.; LUJáN-MORA, S. A heuristic method to evaluate web accessibility for users with low vision. **IEEE Access**, v. 7, p. 125634–125648, 2019.

AIZPURUA, A. et al. Are users the gold standard for accessibility evaluation? In: **Proceedings of the 11th Web for All Conference**. New York, NY, USA: Association for Computing Machinery, 2014. (W4A '14). ISBN 9781450326513. Disponível em: <https://doi.org/10.1145/2596695.2596705>.

AIZPURUA, A.; ARRUE, M.; VIGO, M. Prejudices, memories, expectations and confidence influence experienced accessibility on the web. **Computers in Human Behavior**, Elsevier, v. 51, p. 152–160, 2015.

AKRAM, M.; SULAIMAN, R. B. An empirical study to evaluate the accessibility of arabic websites by low vision users. In: **2020 8th International Conference on Information Technology and Multimedia (ICIMU)**. IEEE, 2020. p. 206–211. Disponível em: <10.1109/ICIMU49871.2020.9243565>.

ALAJARMEH, N. The extent of mobile accessibility coverage in wcag 2.1: Sufficiency of success criteria and appropriateness of relevant conformance levels pertaining to accessibility problems encountered by users who are visually impaired. **Universal Access in the Information Society**, Springer, p. 1–26, 2021.

ARCHAMBAULT, D. et al. Web widgets barriers for visually impaired users. **Studies in Health Technology and Informatics**, v. 242, p. 836–842, 2017.

BACH, C. F. **Avaliação de acessibilidade na web: estudo comparativo entre métodos de avaliação com a participação de deficientes visuais (Web accessibility assessment: a comparative study between evaluation methods with the participation of the visually impaired.)**. Dissertação (Mestrado) — Universidade Federal do Estado do Rio de Janeiro, 2009.

BBC, B. B. C. **HTML Accessibility Standards v2.0.** 2014. BBC. Retrieved February 15, 2021 from http://www.bbc.co.uk/guidelines/futuremedia/accessibility/html/.

BBC, B. B. C. **Mobile Accessibility Standards and Guidelines v1.0**. 2014. BBC. Retrieved February 15, 2021 from http://shorturl.at/wCPX3.

BigDataCorp. **Big Data Corp**. 2019. Retrieved September 22, 2021 from https://bigdatacorp.com.br/acessibilidade-dos-sites-brasileiros/.

BRAJNIK, G. A comparative test of web accessibility evaluation methods. In: **Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility**. New York, NY, USA: Association for Computing Machinery, 2008. (Assets '08), p. 113–120. ISBN 9781595939760. Disponível em: <https://doi.org/10.1145/1414471.1414494>.

BRAJNIK, G.; YESILADA, Y.; HARPER, S. The expertise effect on web accessibility evaluation methods. **Human–Computer Interaction**, Taylor & Francis, v. 26, n. 3, p. 246–283, 2011.

CALVO, R.; SEYEDARABI, F.; SAVVA, A. Beyond web content accessibility guidelines: Expert accessibility reviews. In: **Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion**. New York, NY, USA: Association for Computing Machinery, 2016. (DSAI 2016), p. 77–84. ISBN 9781450347488. Disponível em: <https://doi.org/10.1145/3019943.3019955>.

CARVALHO, L. P.; DIAS, F. S.; FREIRE, A. P. An analysis of five different native and web-hybrid platforms for building android apps and their accessibility for screen readers. **SBC Journal on Interactive Systems**, v. 9, n. 3, p. 20–33, 2018.

CARVALHO, L. P.; FREIRE, A. P. Native or web-hybrid apps? an analysis of the adequacy for accessibility of android interface components used with screen readers. In: **Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2017. (IHC 2017). ISBN 9781450363778. Disponível em: <https://doi.org/10.1145/3160504.3160511>.

CARVALHO, L. P. et al. Accessible smart cities? inspecting the accessibility of brazilian municipalities' mobile applications. In: **Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2016. (IHC '16). ISBN 9781450352352. Disponível em: <https://doi.org/10.1145/3033701.3033718>.

CARVALHO, M. C. N. et al. 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**. New York, NY, USA: Association for Computing Machinery, 2018. (SAC '18), p. 2022–2029. ISBN 9781450351911.

DIAS, F. S. **Análise de causas técnicasde problemas de acessibilidade encontrados em aplicativos nativos para dispositivos móveis (Analysis of technical causes of accessibility problems found in native mobile apps)**. Lavras: Trabalho de Conclusão do Curso de Ciência da Computação, Universidade Federal de Lavras, 2018.

DOUSH, I.; ALMERAJ, Z. Evaluating the accessibility of kuwaiti e-government websites. **Jordanian Journal of Computers and Information Technology (JJCIT)**, v. 5, n. 03, 2019.

ELER, M. M. et al. Automated accessibility testing of mobile apps. In: **2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST)**. IEEE, 2018. p. 116–126. Disponível em: <10.1109/ICST.2018.00021>.

FERNANDES, N. et al. Optimus web: Selective delivery of desktop or mobile web pages. In: **Proceedings of the 12th International Web for All Conference**. New York, NY, USA: Association for Computing Machinery, 2015. (W4A '15). ISBN 9781450333429. Disponível em: <https://doi.org/10.1145/2745555.2746668>.

FOESCH, C. K. et al. S.t.c: Sistema de transporte cooperativo para deficientes visuais. **Revista ComInG - Communications and Innovations Gazette**, v. 6, n. 1, p. 108–116, nov. 2022. Disponível em: <https://periodicos.ufsm.br/coming/article/view/40046>.

FREIRE, A. P. **Disabled people and the Web: User-based measurement of accessibility**. Tese (Doutorado) — University of York, 2012.

GADDE, P.; BOLCHINI, D. From screen reading to aural glancing: Towards instant access to key page sections. In: **Proceedings of the 16th International ACM SIGACCESS Conference on Computers Accessibility**. New York, NY, USA: Association for Computing Machinery, 2014. (ASSETS '14), p. 67–74. ISBN 9781450327206. Disponível em: <https://doi.org/10.1145/2661334.2661363>.

GALKUTE, M.; P., L. A. R.; M., V. A. S. Improving the web accessibility of a university library for people with visual disabilities through a mixed evaluation approach. In: MEISELWITZ, G. (Ed.). **Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis**. Springer International Publishing, 2020. p. 56–71. ISBN 978-3-030-49570-1. Disponível em: <https://doi.org/10.1007/978-3-030-49570-1_5>.

Ghidini, E. et al. Developing apps for visually impaired people: Lessons learned from practice. In: **2016 49th Hawaii International Conference on System Sciences (HICSS)**. [S.l.: s.n.], 2016. p. 5691–5700.

GONÇALVES, R. et al. Evaluation of e-commerce websites accessibility and usability: an e-commerce platform analysis with the inclusion of blind users. **Universal Access in the Information Society**, Springer, v. 17, n. 3, p. 567–583, 2018.

GOOGLE. **Accessibility Scanner**. 2015. Retrieved February 28, 2023. Disponível em: <https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.auditor&hl=pt_BR>.

Governo Brasileiro. **eMAG - Modelo de Acessibilidade em Governo Eletronico.** 2014. Retrieved February 16, 2023 from http://emag.governoeletronico.gov.br/.

Governo Brasileiro. **Lei Nº 13.146, DE 6 de Julho de 2015**. 2015. Retrieved March 02, 2023. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm>.

Governo Brasileiro. **Lei Nº 14.129, DE 29 DE MARÇO DE 2021**. 2021. Retrieved March 02, 2023. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14129.htm>.

Governo Brasileiro. **Acessibilidade em aplicativos de dispositivos moveis - Requisitos**. 2022. Retrieved November 30, 2022. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?Q=SURUWmJCcTFwcExjK2txS2wyYmpOYStTS0p4b1VnMklzcDB3WXdORlJKQT0=>.

HARRISON, C.; PETRIE, H. Severity of usability and accessibility problems in ecommerce and egovernment websites. In: BRYAN-KINNS, N. et al. (Ed.). **People and Computers**

XX — Engage. London: Springer London, 2007. p. 255–262. ISBN 978-1-84628-664-3. Disponível em: <https://doi.org/10.1007/978-1-84628-664-3_19>.

ISMAIL, A.; KUPPUSAMY, K.; NENGROO, A. S. Multi-tool accessibility assessment of government department websites: a case-study with jkgad. **Disability and Rehabilitation: Assistive Technology**, Taylor & Francis, v. 13, n. 6, p. 504–516, 2018.

ISO. **ISO 9241-11. Usability: Definitions and concepts**. 2018. Retrieved February 16, 2023 from https://www.iso.org/obp/ui/iso:std:iso:9241:-11:ed-2:v1:en.

JAEGER, P. T. Assessing section 508 compliance on federal e-government web sites: A multi-method, user-centered evaluation of accessibility for persons with disabilities. **Government Information Quarterly**, Elsevier, v. 23, n. 2, p. 169–190, 2006.

KIM, H. K. et al. The interaction experiences of visually impaired people with assistive technology: A case study of smartphones. **International Journal of Industrial Ergonomics**, Elsevier, v. 55, p. 22–33, 2016.

KULPA, C. C.; AMARAL, F. G. Evaluation of tablet pc application interfaces with low vision users: Focusing on usability. In: MARCUS, A. (Ed.). **Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience**. Cham: Springer International Publishing, 2014. p. 273–284. ISBN 978-3-319-07668-3.

LAZAR, J. **Web usability: A user-centered design approach**. Addison-Wesley Longman Publishing Co., Inc., 2005. ISBN 978-0-321-32135-0. Disponível em: <https://dl.acm.org/doi/10.5555/1088946>.

LOPES, R.; FAÇANHA, A. R.; VIANA, W. I can't pay! accessibility analysis of mobile banking apps. In: **Proceedings of the Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: Association for Computing Machinery, 2022. (WebMedia '22), p. 253–257. ISBN 9781450394093. Disponível em: <https://doi.org/10.1145/3539637.3558048>.

LOUREIRO, J. R.; CAGNIN, M. I.; PAIVA, D. M. B. Web accessibility in social networking services. In: MURGANTE, B. et al. (Ed.). **Computational Science and Its Applications – ICCSA 2014**. Cham: Springer International Publishing, 2014. p. 586–601. ISBN 978-3-319-09156-3. Disponível em: <https://doi.org/10.1007/978-3-319-09156-3_41>.

MANKOFF, J.; FAIT, H.; TRAN, T. Is your web page accessible? a comparative study of methods for assessing web page accessibility for the blind. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2005. (CHI '05), p. 41–50. ISBN 1581139985. Disponível em: <https://doi.org/10.1145/1054972.1054979>.

MARTINS, R. S. et al. Where is the description? investigating accessibility issues in portuguese versions of smart home apps. In: **Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2022. (IHC '22). ISBN 9781450395069. Disponível em: <https://doi.org/10.1145/3554364.3559133>.

MATEUS, D. A. et al. Accessibility of mobile applications: Evaluation by users with visual impairment and by automated tools. In: **Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2020. (IHC '20). ISBN 9781450381727. Disponível em: <https://doi.org/10.1145/3424953.3426633>.

MATEUS, D. A. et al. A systematic mapping of accessibility problems encountered on websites and mobile apps: A comparison between automated tests, manual inspections and user evaluations. **Journal on Interactive Systems**, v. 12, n. 1, p. 145–171, Nov. 2021. Disponível em: <https://sol.sbc.org.br/journals/index.php/jis/article/view/1778>.

MATEUS, D. A. et al. Evaluation methods in legal procedures concerning digital accessibility in brazil: An analysis of cases investigated by the federal public ministry. In: **Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2022. (IHC '22). ISBN 9781450395069. Disponível em: <https://doi.org/10.1145/3554364.3559130>.

MÁTRAI, R. K. How to make an electronic library accessible. **The electronic library**, Emerald Publishing Limited, 2018.

MOUNIKA, P. et al. Accessibility evaluation of three important indian websites. In: CHAKRABARTI, A. (Ed.). **Research into Design for a Connected World**. Singapore: Springer Singapore, 2019. p. 243–254. ISBN 978-981-13-5974-3. Disponível em: <https://doi.org/10.1007/978-981-13-5974-3_21>.

NOGUEIRA, T. C. et al. Evaluating responsive web design's impact on blind users. **IEEE MultiMedia**, IEEE, v. 24, n. 2, p. 86–95, 2017.

OLIVEIRA, G. A. A. de; BETTIO, R. W. de; FREIRE, A. P. Accessibility of the smart home for users with visual disabilities: An evaluation of open source mobile applications for home automation. In: **Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2016. (IHC '16). ISBN 9781450352352. Disponível em: <https://doi.org/10.1145/3033701.3033730>.

OMS. **World Health Organization**. 2021. Acessado em 20 de Junho, 2021 from https://www.who.int/. Disponível em: <https://www.who.int/es>.

ONCINS, E. Accessibility in online user testing. v. 4, p. 6–22, Oct. 2021. Disponível em: <https://jatjournal.org/index.php/jat/article/view/176>.

PARK, K.; SO, H.-J.; CHA, H. Digital equity and accessible moocs: Accessibility evaluations of mobile moocs for learners with visual impairments. **Australasian Journal of Educational Technology**, v. 35, n. 6, p. 48–63, 2019.

PEREIRA, L. S.; ARCHAMBAULT, D. Correlating navigation barriers on web 2.0 with accessibility guidelines. In: MIESENBERGER, K.; KOUROUPETROGLOU, G. (Ed.). **Computers Helping People with Special Needs**. Cham: Springer International Publishing, 2018. p. 13–21. ISBN 978-3-319-94277-3.

PETERSEN, K. et al. Systematic mapping studies in software engineering. In: **Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering**. Swindon, GBR: BCS Learning Development Ltd., 2008. (EASE'08), p. 68–77.

PETRIE, H.; BEVAN, N. The evaluation of accessibility, usability, and user experience. **The universal access handbook**, v. 1, p. 1–16, 2009.

PETRIE, H.; POWER, C. What do users really care about? a comparison of usability problems found by users and experts on highly interactive websites. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2012. (CHI '12), p. 2107–2116. ISBN 9781450310154. Disponível em: <https://doi.org/10.1145/2207676.2208363>.

POWER, C. et al. Guidelines are only half of the story: Accessibility problems encountered by blind users on the web. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2012. (CHI '12), p. 433–442. ISBN 9781450310154. Disponível em: <https://doi.org/10.1145/2207676.2207736>.

POWER, J. A. Ebsco information services usability study on accessibility. **Reference Services Review**, Emerald Publishing Limited, 2018.

Prefeitura Paulista. **Comissão Permanente de Acessibilidade**. 2018. Retrieved March 02, 2022. Disponível em: <https://www.prefeitura.sp.gov.br/cidade/secretarias/pessoa_com_deficiencia/sao_paulo_mais_inclusiva/?p=203752>.

PRIBEANU, C.; FOGARASSY-NESZLY, P.; PĂTRU, A. Municipal web sites accessibility and usability for blind users: preliminary results from a pilot study. **Universal access in the information society**, Springer, v. 13, n. 3, p. 339–349, 2014.

QUISPE, F. E. M.; ELER, M. M. Accessibility recommendations for mobile applications: A contribution to the brazilian digital government standards. In: **Proceedings of the XIV Brazilian Symposium on Information Systems**. New York, NY, USA: Association for Computing Machinery, 2018. (SBSI'18). ISBN 9781450365598. Disponível em: <https://doi.org/10.1145/3229345.3229415>.

RODRIGUES, L. A. C.; PRIETCH, S. S. Analysis, redesign and validation of accessibility resources applied to an official electronic journal for the promotion of equal access to public acts. In: **Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2018. (IHC 2018). ISBN 9781450366014. Disponível em: <https://doi.org/10.1145/3274192.3274200>.

RØMEN, D.; SVANÆS, D. Validating wcag versions 1.0 and 2.0 through usability testing with disabled users. **Universal Access in the Information Society**, Springer, v. 11, n. 4, p. 375–385, 2012.

RYSAVY, M. D.; MICHALAK, R. Assessing the accessibility of library tools & services when you aren't an accessibility expert: Part 1. **Journal of Library Administration**, Taylor & Francis, v. 60, n. 1, p. 71–79, 2020.

SIEBRA, C. et al. **Mobile Accessibility, guide for developing accessible mobile apps**. 2017. Retrieved June 30, 2021 from http://www.sidi.org.br/guiadeacessibilidade/en/index.html.

SILVA, C.; ELER, M. Um estudo sobre a avaliação automática de diretrizes de acessibilidade para dispositivos móveis. In: **Anais Estendidos do XIV Simpósio Brasileiro de Sistemas de Informação**. Porto Alegre, RS, Brasil: SBC, 2018. p. 111–113. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/6208>.

SILVA, C.; ELER, M. M.; FRASER, G. A survey on the tool support for the automatic evaluation of mobile accessibility. In: **Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion**. New York, NY, USA: Association for Computing Machinery, 2018. (DSAI 2018), p. 286–293. ISBN 9781450364676.

SILVA, C.; ELER, M. M.; FRASER, G. A survey on the tool support for the automatic evaluation of mobile accessibility. In: **Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion**. New York, NY, USA: Association for Computing Machinery, 2018. (DSAI 2018), p. 286–293. ISBN 9781450364676.

SILVA, C. A. et al. Types of problems encountered by automated tool accessibility assessments, expert inspections and user testing: a systematic literature mapping. In: **Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems**. [S.l.: s.n.], 2019. p. 1–11.

SILVA, C. F. da; FERREIRA, S. B. L.; RAMOS, J. a. F. M. Whatsapp accessibility from the perspective of visually impaired people. In: **Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2016. (IHC '16). ISBN 9781450352352. Disponível em: <https://doi.org/10.1145/3033701.3033712>.

SILVA, C. F. da; FERREIRA, S. B. L.; RAMOS, J. a. F. M. Whatsapp accessibility from the perspective of visually impaired people. In: **Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2016. (IHC '16). ISBN 9781450352352. Disponível em: <https://doi.org/10.1145/3033701.3033712>.

SILVA, C. F. da; FERREIRA, S. B. L.; SACRAMENTO, C. Mobile application accessibility in the context of visually impaired users. In: **Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2018. (IHC 2018). ISBN 9781450366014. Disponível em: <https://doi.org/10.1145/3274192.3274224>.

SILVA, C. F. da; FERREIRA, S. B. L.; SACRAMENTO, C. Mobile application accessibility in the context of visually impaired users. In: **Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2018. (IHC 2018). ISBN 9781450366014. Disponível em: <https://doi.org/10.1145/3274192.3274224>.

SOUZA, N.; CARDOSO, E.; PERRY, G. T. Limitações da avaliação automatizada de acessibilidade em uma plataforma de moocs: Estudo de caso de uma plataforma brasileira. **Revista Brasileira de Educação Especial**, SciELO Brasil, v. 25, p. 603–616, 2019.

STEPHANIDIS, C. **The universal access handbook**. 1st edition. ed. CRC Press, 2009. ISBN 9780429135309. Disponível em: <https://doi.org/10.1201/9781420064995>.

VALENCIA., X. et al. Interdependent components for the development of accessible xul applications for screen reader users. In: INSTICC. **Proceedings of the 10th International Conference on Web Information Systems and Technologies - Volume 1: WEBIST,**.

SciTePress, 2014. p. 65–73. ISBN 978-989-758-024-6. ISSN 2184-3252. Disponível em: <https://doi.org/10.5220/0004855200650073>.

VIGO, M.; BROWN, J.; CONWAY, V. Benchmarking web accessibility evaluation tools: Measuring the harm of sole reliance on automated tests. In: **Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility**. New York, NY, USA: Association for Computing Machinery, 2013. (W4A '13). ISBN 9781450318440.

W3C. **Web Content Accessibility Guidelines 1.0, Web Accessibility Initiative**. 1999. Retrieved February 28, 2023 from https://www.w3.org/TR/WCAG10/.

W3C. **W3C – World Wide Web Consortium. Web Content Accessibility Guidelines (WCAG) 2.0**. 2008. Retrieved February 28, 2023 from https://www.w3.org/TR/WCAG20/.

W3C. **W3C. Web Content Accessibility Guidelines (WCAG) 2.1**. 2018. Retrieved February 28, 2023 from https://www.w3.org/TR/WCAG21/.

W3C. **W3C. Web Content Accessibility Guidelines (WCAG) 2.2**. 2020. Retrieved February 28, 2023 from https://www.w3.org/TR/WCAG22/.

W3C. **W3C. Web Content Accessibility Guidelines (WCAG) 3.0**. 2021. Retrieved September 10, 2022 from https://www.w3.org/TR/wcag-3.0/.

WENTZ, B. et al. Documenting the accessibility of 100 us bank and finance websites. **Universal Access in the Information Society**, Springer, v. 18, n. 4, p. 871–880, 2019.

WPT. **Movimento Web Para Todos**. 2019. Retrieved September 22, 2021 from https://mwpt.com.br/3o-estudo-de-acessibilidade-do-movimento-web-para-todos-nos-sites-brasileiros/.

YAN, S.; RAMACHANDRAN, P. The current status of accessibility in mobile apps. **ACM Transactions on Accessible Computing (TACCESS)**, ACM New York, NY, USA, v. 12, n. 1, p. 1–31, 2019.

YI, Y. J. Web accessibility of healthcare web sites of korean government and public agencies: a user test for persons with visual impairment. **Universal Access in the Information Society**, Springer, v. 19, n. 1, p. 41–56, 2020.

ZITKUS, E. et al. Accessibility and usability of websites intended for people with disabilities: A preliminary study. In: MARCUS, A. (Ed.). **Design, User Experience, and Usability: Novel User Experiences**. Cham: Springer International Publishing, 2016. p. 678–688. ISBN 978-3-319-40355-7. Disponível em: <https://doi.org/10.1007/978-3-319-40355-7_66>.

APPENDIX A – Results of Systematic Mapping

Table 1 – List of papers (continuation)

ID	Title	Citation
A1	Accessibility and usability of websites intended for people with disabilities: A preliminary study	(ZITKUS et al., 2016)
A2	Accessibility and usability problems encountered on websites and applications in mobile devices by blind and normal-vision	(CARVALHO et al., 2018)
A3	Analysis of web accessibility in social networking services through blind users' perspective and an accessible prototype	(LOUREIRO; CAGNIN; PAIVA, 2014)
A4	Analysis, redesign and validation of accessibility resources applied to an official electronic journal for the promotion of equal access to public acts	(RODRIGUES; PRIETCH, 2018)
A5	Are users the gold standard for accessibility evaluation?	(AIZPURUA et al., 2014)
A6	Correlating navigation barriers on web 2.0 with accessibility guidelines	(PEREIRA; ARCHAMBAULT, 2018)
A7	Documenting the accessibility of 100 US bank and finance websites	(WENTZ et al., 2019)
A8	EBSCO information services usability study on accessibility	(POWER, 2018)
A9	Evaluating Responsive Web Design's Impact on Blind Users	(NOGUEIRA et al., 2017)
A10	Evaluation of e-commerce websites accessibility and usability: an e-commerce platform analysis with the inclusion of blind users	(GONÇALVES et al., 2018)
A11	From screen reading to aural glancing: Towards instant access to key page sections	(GADDE; BOLCHINI, 2014)
A12	How to make an electronic library accessible	(MÁTRAI, 2018)
A13	Multi-tool accessibility assessment of government department websites: a case-study with JKGAD	(ISMAIL; KUPPUSAMY; NENGROO, 2018)
A14	Municipal web sites accessibility and usability for blind users: Preliminary results from a pilot study	(PRIBEANU; FOGARASSY-NESZLY; PĂTRU, 2014)

Table 1 – *List of papers (continuation)*

ID	Title	Citation
A15	Prejudices, memories, expectations and confidence influence experienced accessibility on the Web	(AIZPURUA; ARRUE; VIGO, 2015)
A16	Should I trust it when I cannot see it? Credibility assessment for blind web users	(ABDOLRAHMANI; KUBER, 2016)
A17	Web accessibility in social networking services	(LOUREIRO; CAGNIN; PAIVA, 2014)
A18	Web accessibility of healthcare Web sites of Korean government and public agencies: a user test for persons with visual impairment	(YI, 2020)
A19	Web Widgets Barriers for Visually Impaired Users	(ARCHAMBAULT et al., 2017)
A20	Interdependent components for the development of accessible XUL applications for screen reader users	(VALENCIA. et al., 2014)
A21	Optimus web: Selective delivery of desktop or mobile web pages	(FERNANDES et al., 2015)
A22	WhatsApp accessibility from the perspective of visually impaired people	(SILVA; FERREIRA; RAMOS, 2016b)
A23	Accessibility of the smart home for users with visual disabilities: An evaluation of open source mobile applications for home	(OLIVEIRA; BETTIO; FREIRE, 2016)
A24	Accessible smart cities?: Inspecting the accessibility of Brazilian municipalities' mobile applications	(CARVALHO et al., 2016)
A25	Improving the web accessibility of a university library for people with visual disabilities through a mixed evaluation approach	(GALKUTE; P.; M., 2020)
A26	Heuristic method of evaluating accessibility of mobile in selected applications for air quality monitoring	(ACOSTA-VARGAS et al., 2020b)
A27	Accessibility Assessment of Mobile Meteorological Applications for Users with Low Vision	(ACOSTA-VARGAS et al., 2020a)
A28	Accessibility evaluation of three important Indian websites	(MOUNIKA et al., 2019)

Table 1 – *List of papers (conclusion)*

ID	Title	Citation
A29	Assessing the Accessibility of Library Tools & Services When You Aren't an Accessibility Expert: Part 1	(RYSAVY; MICHALAK, 2020)
A30	A Heuristic Method to Evaluate Web Accessibility for Users with Low Vision	(ACOSTA-VARGAS; SALVADOR-ULLAURI; LUJáN-MORA, 2019)
A31	Digital equity and accessible MOOCs: Accessibility evaluations of mobile MOOCs for learners with visual impairments	(PARK; SO; CHA, 2019)
A32	Evaluating the accessibility of Kuwaiti e-government websites	(DOUSH; ALMERAJ, 2019)
A33	Mobile Application Accessibility in the Context of Visually Impaired Users	(SILVA; FERREIRA; SACRAMENTO, 2018b)
A34	Evaluation of tablet PC application interfaces with low vision users: Focusing on usability	(KULPA; AMARAL, 2014)
A35	The interaction experiences of visually impaired people with assistive technology: A case study of smartphones	(KIM et al., 2016)
A36	An Empirical Study to Evaluate the Accessibility of Arabic Websites by Low Vision Users	(AKRAM; SULAIMAN, 2020)
A37	Accessibility of mobile applications: Evaluation by users with visual impairment and by automated tools	(MATEUS et al., 2020)
A38	The Current Status of Accessibility in Mobile Apps	(YAN; RAMACHANDRAN, 2019)

Source: Mateus et al. (2021)

Table 2 – Accessibility problems encountered by expert inspections (continuation)

Barrier code	Barrier	Description	Studies
27	Absence of shortcuts	Absence of shortcuts to access main content.	(OLIVEIRA; BETTIO; FREIRE, 2016)
5	Absence of headers	Pages that do not have headings to indicate main content or sections.	(CARVALHO et al., 2016)
18	Absence of resources for expansion	Absence of resources to expand textual content.	(OLIVEIRA; BETTIO; FREIRE, 2016)
1	Absence of labels	Form fields that have no labels on their purpose.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
6	Absence of alternative text	Non-text content that does not have alternative text.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
15	Absence of titles	Pages that do not have an identifying title.	(CARVALHO et al., 2016)
12	Insufficient contrast	Bad contrast ratio.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
41	Visible Focus	The user cannot understand what the system expects him to do.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)
45	Consistent Identification	Components that have the same functionality in a set of web pages are identified consistently.	(OLIVEIRA; BETTIO; FREIRE, 2016)
48	Images	If the technologies being used can provide visual presentation, text is used to convey information instead of images of text except for the following.	(CARVALHO et al., 2016)
46	Error identification	If an input error is automatically detected, the item with an error is identified and the error is described to the user in text.	(OLIVEIRA; BETTIO; FREIRE, 2016)
8	Language not set	Content that has no language defined.	(PARK; SO; CHA, 2019)

Table 2 – *Accessibility problems encountred by expert inspections (conclusion)*

Barrier code	**Barrier**	**Description**	**Studies**
42	Location	It is not possible to know where it is within the system	(CARVALHO et al., 2016)
59	Time limits	Users are advised of the duration of any user inactivity that may cause data loss, unless data is preserved when the user does not take any action for more than 20 hours.	(CARVALHO et al., 2016)
39	Keyboard	All mouse operations have an accessible keyboard equivalents	(CARVALHO et al., 2016)
20	Inadequate navigation sequence	Content that does not allow an adequate navigation sequence by screen readers.	(OLIVEIRA; BETTIO; FREIRE, 2016; CARVALHO et al., 2016)

Source: Mateus et al. (2021)

Table 3 – Accessibility barriers encountered apps by user evaluations (continuation)

Barrier code	Barrier	Description	Studies
1	Absence of labels	Form fields that have no labels on their purpose.	(CARVALHO et al., 2018; ACOSTA-VARGAS et al., 2020a)
2	Inappropriate link destination	Links that do not direct the user to the page they are intended for.	(FERNANDES et al., 2015)
4	Too much information	Pages that contain a lot of data and elements.	(CARVALHO et al., 2018; PARK; SO; CHA, 2019)
6	Absence of alternative text	Non-text content that does not have alternative text.	(CARVALHO et al., 2018; SILVA; FERREIRA; RAMOS, 2016b; ACOSTA-VARGAS et al., 2020b; PARK; SO; CHA, 2019; SILVA; FERREIRA; SACRAMENTO, 2018b)
7	Empty links	Links that do not have a description of their purpose.	(SILVA; FERREIRA; RAMOS, 2016b)
8	Language not set	Content that has no language defined.	(SILVA; FERREIRA; SACRAMENTO, 2018b; KULPA; AMARAL, 2014)
12	Insufficient contrast	Bad contrast ratio.	(KIM et al., 2016; SILVA; FERREIRA; SACRAMENTO, 2018b; KULPA; AMARAL, 2014; ACOSTA-VARGAS et al., 2020a; ACOSTA-VARGAS et al., 2020b; SILVA; FERREIRA; RAMOS, 2016b)
14	Incompatibility of technologies	Content inaccessible by screen readers, such as flash.	(CARVALHO et al., 2018)

Table 3 – *Accessibility barriers encountered apps by user evaluations (continuation)*

Barrier code	Barrier	Description	Studies
15	Absence of titles	Pages that do not have an identifying title.	(SILVA; FERREIRA; RAMOS, 2016b; SILVA; FERREIRA; SACRAMENTO, 2018b)
17	Inappropriate description in controls	Controls, such as a link or button, that have an inappropriate description.	(CARVALHO et al., 2018; ACOSTA-VARGAS et al., 2020a)
18	Absence of resources for expansion	Absence of resources to expand textual content.	(FERNANDES et al., 2015)
20	Inadequate navigation sequence	Content that does not allow an adequate navigation sequence by screen readers.	(CARVALHO et al., 2018; SILVA; FERREIRA; RAMOS, 2016b; ACOSTA-VARGAS et al., 2020b; SILVA; FERREIRA; SACRAMENTO, 2018b; KULPA; AMARAL, 2014)
22	Unreachable help link	Help link is not easy to find.	(SILVA; FERREIRA; RAMOS, 2016b)
26	Absence of feedback	When using any means of access, such as a link or button, the user does not receive feedback.	(CARVALHO et al., 2018; SILVA; FERREIRA; RAMOS, 2016b)
32	Inappropriate textual content	Lists, paragraphs or other textual elements that are not correctly identified by screen readers.	(SILVA; FERREIRA; RAMOS, 2016b)
34	Inappropriate title	Page title that does not correctly describe the content.	(KIM et al., 2016)
37	Inconsistent content organization	Content that is not well organized.	(CARVALHO et al., 2018)
39	Keyboard	All mouse operations have an accessible keyboard equivalents	(KIM et al., 2016; SILVA; FERREIRA; RAMOS, 2016b; AKRAM; SULAIMAN, 2020)

Table 3 – *Accessibility barriers encountered apps by user evaluations (continuation)*

Barrier code	Barrier	Description	Studies
40	Violated header	Violated header structures.	(PARK; SO; CHA, 2019)
41	Visible Focus	The user cannot understand what the system expects him to do.	(SILVA; FERREIRA; RAMOS, 2016b; PARK; SO; CHA, 2019)
42	Location	It is not possible to know where it is within the system	(SILVA; FERREIRA; RAMOS, 2016b; KIM et al., 2016)
43	Change upon request	Can't see pop-up.	(SILVA; FERREIRA; RAMOS, 2016b)
47	Order Focus	If a web page can be navigated sequentially and the navigation sequences affect the meaning or the operation, the components that can be focused are focused on in an order that preserves meaning and operability	(PARK; SO; CHA, 2019)
48	Images	If the technologies being used can provide visual presentation, text is used to convey information instead of images of text except for the following.	(ACOSTA-VARGAS et al., 2020b)
49	Use of colors	Color is not used as the only visual means of transmitting information, indicating an action, asking for an answer or distinguishing a visual element	(ACOSTA-VARGAS et al., 2020b)
50	Resize text	Increase text	(KIM et al., 2016; ACOSTA-VARGAS et al., 2020a)

Table 3 – *Accessibility barriers encountered apps by user evaluations (continuation)*

Barrier code	**Barrier**	**Description**	**Studies**
51	Spacing	spacing between images, text, forms.	(KIM et al., 2016)
55	Pause, Stop, Hide	For information in motion, in intermittent mode, in displacement or in automatic update, all the following statements are green.	(KIM et al., 2016)
56	Audio description or Alternative Media (Pre-recorded)	An alternative to media based or an audio description of pre-recorded video content is provided for synchronized media, except when the media is an alternative to text and is clearly identified as such.	(PARK; SO; CHA, 2019)
57	Analyze	In content implemented using markup languages, the elements have complete start and end tags, the elements are nested according to the respective specifications, the elements do not contain duplicate attributes, and any IDs are unique, except when the specifications allow these characteristics.	(PARK; SO; CHA, 2019)
58	Visual Presentation	Foreground and background colors can be selected by the user.	(SILVA; FERREIRA; SACRAMENTO, 2018b)
60	Difficulty using horizontal mode	-	(KIM et al., 2016)

Table 3 – *Accessibility barriers encountered apps by user evaluations (continuation)*

Barrier code	**Barrier**	**Description**	**Studies**
61	Difficulty finding the menu	-	(KIM et al., 2016)
62	Speak very slow.	-	(KIM et al., 2016)
63	Face recognition performance is lower than I expected.	-	(KIM et al., 2016)
64	Recording the TalkBack sound with my voice.	-	(KIM et al., 2016)
65	Difficulties to understand the location of a face on the screen.	-	(KIM et al., 2016)
66	Correct pronunciation.	-	(KIM et al., 2016)
67	Target Size	Font size, button.	(KIM et al., 2016)
68	Hostile voice screen reader.	-	(KIM et al., 2016)
69	The beep sound lasts a long time.	-	(KIM et al., 2016)
70	Difficulties in memorizing the interface layout.	-	(KIM et al., 2016)
71	Difficulty understanding the direction to move the camera.	-	(KIM et al., 2016)
72	Difficulties to infer the word of the big and small face.	-	(KIM et al., 2016)
73	Speak very loudly	-	(KIM et al., 2016)
74	Difficulty using a touch-sensitive interface	-	(KIM et al., 2016)
75	Source shape	-	(KIM et al., 2016)

Table 3 – *Accessibility barriers encountered apps by user evaluations (conclusion)*

Barrier code	**Barrier**	**Description**	**Studies**
76	Difficulties in recognizing the meaning of icon designs	-	(KIM et al., 2016)
77	Complex interface layout	-	(KIM et al., 2016)
78	Very loud beep sound	-	(KIM et al., 2016)
79	Stroke width	-	(KIM et al., 2016)
80	Needs higher recording quality	-	(KIM et al., 2016)

Source: Mateus et al. (2021)

APPENDIX B – Checklist for evaluation expert

Table 4 – CheckList WCAG

Problem category	Yes	No	Screen	Severity
WCAG – 1.1.1 Non-text content				
WCAG – 1.2.3 Audio description or media alternative (pre-recorded)				
WCAG – 1.3.1 Information and relationships				
WCAG – 1.3.2 Significant Sequence				
WCAG – 1.4.1 Use of Color				
WCAG – 1.4.10 Reflux				
WCAG – 1.4.3 Contrast (minimum)				
WCAG – 1.4.4 Resize text				
WCAG – 2.1.1 Keyboard				
WCAG – 2.1.2 No keyboard trap				
WCAG – 2.2.2 Pause, Stop, Hide				
WCAG – 2.4.3 Focus Order				
WCAG – 2.4.4 Purpose of the link (in context)				
WCAG – 2.4.6 Headers and Labels				
WCAG – 2.4.8 Location				
WCAG – 2.5.5 Target Size				
WCAG – 3.1.1 Language of the page				
WCAG – 3.1.2 Language of the Parties				
WCAG – 3.1.4 Abbreviations				
WCAG – 3.2.3 Consistent Navigation				
WCAG – 3.3.1 Error Identification				
WCAG – 3.3.2 Labels or instructions				
WCAG – 3.3.4 Error prevention (legal, financial, data)				
WCAG – 3.3.5 Help				
WCAG – 4.1.1 Analysis				
WCAG – 4.1.2 Name, Function, Value				

Source: Own author

APPENDIX C – Results of inspections performed with developers

Table 5 – Problems encountered by DEV front-end using WCAG

	Receita Federal		Saraiva	
Found problems	**# expert**	**user**	**# expert**	**user**
1.1.1 Non-text content	2	YES	4	YES
1.2.3 Audio description or media alternative (pre-recorded)			3	YES
1.3.1 Information and relationships	4	YES	5	YES
1.3.2 Significant Sequence	4		4	YES
1.4.1 Use of Color	5		4	YES
1.4.10 Reflux			1	YES
1.4.3 Contrast (minimum)	4		3	YES
1.4.4 Resize Text	3	YES	3	YES
2.1.1 Keyboard	1		1	YES
2.1.2 No keyboard trap	2		1	YES
2.2.2 Pause, Stop, Hide	2		1	YES
2.4.3 Focus Order	2	YES	1	YES
2.4.4 Purpose of the link (in context)				YES
2.4.6 Headers and Labels	2		3	YES
2.4.8 Location	1	YES	2	YES
2.5.5 Target Size		YES	1	
3.1.1 Language of the page	1	YES	2	YES
3.1.2 Language of Parts			1	
3.1.4 Abbreviations	2			
3.2.3 Consistent Navigation				YES
3.3.1 Error Identification	3	YES	1	YES
3.3.2 Labels or instructions		YES		YES
3.3.4 Error prevention (legal, financial, data)			1	
3.3.5 Help		YES		YES
4.1.1 Parsing		YES		
4.1.2 Name, Function, Value		YES		YES

Source: Own author

Table 6 – Problems encountered by DEV Tester using WCAG

	Receita Federal		Saraiva	
Found problems	**# expert**	**user**	**# expert**	**user**
1.1.1 Non-text content	1	YES	2	YES
1.2.3 Audio description or media alternative (pre-recorded)				YES
1.3.1 Information and relationships		YES		YES
1.3.2 Significant Sequence	2		2	
1.4.1 Use of Color	1		2	YES
1.4.10 Reflux	2		2	YES
1.4.3 Contrast (minimum)	1		1	YES
1.4.4 Resize Text	1		1	YES
2.1.1 Keyboard		YES		YES
2.1.2 No keyboard trap	1			YES
2.2.2 Pause, Stop, Hide				YES
2.4.3 Focus Order	1		1	YES
2.4.4 Purpose of the link (in context)	2	YES		YES
2.4.6 Headers and Labels	2		2	YES
2.4.8 Location	2	YES	2	YES
2.5.5 Target Size	1	YES		
3.1.1 Language of the page		YES	1	YES
3.1.2 Language of Parts	1	YES		
3.1.4 Abbreviations				YES
3.2.3 Consistent Navigation	2	YES	2	YES
3.3.1 Error Identification	1			Yes
3.3.2 Labels or instructions	1	YES	1	YES
3.3.4 Error prevention (legal, financial, data)			1	
3.3.5 Help	2	YES		YES
4.1.1 Parsing	1	YES	2	
4.1.2 Name, Function, Value	1	YES		YES

Source: Own author

Table 7 – Problems encountered by DEV front-end using WCAG

	Receita Federal		**Saraiva**	
Found problems	**# expert**	**user**	**# expert**	**user**
1.1.1 Non-text content	3	YES	4	YES
1.2.3 Audio description or media alternative (pre-recorded)			2	YES
1.3.1 Information and relationships	6	YES	5	YES
1.3.2 Significant Sequence	5		6	YES
1.4.1 Use of Color	5		5	YES
1.4.10 Reflux	2		4	YES
1.4.3 Contrast (minimum)	4		5	YES
1.4.4 Resize Text	2	YES	5	YES
2.1.1 Keyboard	5		3	YES
2.1.2 No keyboard trap	2			YES
2.2.2 Pause, Stop, Hide	2		2	YES
2.4.3 Focus Order	4	YES	2	YES
2.4.4 Purpose of the link (in context)	3		3	YES
2.4.6 Headers and Labels	4		6	YES
2.4.8 Location	2	YES		YES
2.5.5 Target Size	3	YES	6	
3.1.1 Language of the page	2	YES	4	YES
3.1.2 Language of Parts	2		3	
3.1.4 Abbreviations				
3.2.3 Consistent Navigation	2		4	YES
3.3.1 Error Identification	2	YES		YES
3.3.2 Labels or instructions		YES	4	YES
3.3.4 Error prevention (legal, financial, data)			5	
3.3.5 Help	1	YES		YES
4.1.1 Parsing		YES	1	
4.1.2 Name, Function, Value	1	YES	3	YES

Source: Own author

Table 8 – Problems encountered by DEV front-end using WCAG

	Receita Federal		**Saraiva**	
Found problems	**# expert**	**user**	**# expert**	**user**
1.1.1 Non-text content	1	YES	6	YES
1.2.3 Audio description or media alternative (pre-recorded)			1	YES
1.3.1 Information and relationships	8	YES	6	YES
1.3.2 Significant Sequence	8		7	YES
1.4.1 Use of Color	7		8	YES
1.4.10 Reflux	5		4	YES
1.4.3 Contrast (minimum)	6		7	YES
1.4.4 Resize Text	4	YES	5	YES
2.1.1 Keyboard	6		5	YES
2.1.2 No keyboard trap	3		2	YES
2.2.2 Pause, Stop, Hide	3		4	YES
2.4.3 Focus Order	3	YES	5	YES
2.4.4 Purpose of the link (in context)	5		6	YES
2.4.6 Headers and Labels	5		6	YES
2.4.8 Location	4	YES	5	YES
2.5.5 Target Size	5	YES	4	
3.1.1 Language of the page	3	YES	1	YES
3.1.2 Language of Parts	2		4	
3.1.4 Abbreviations	3		2	
3.2.3 Consistent Navigation	6		9	YES
3.3.1 Error Identification	5	YES	6	YES
3.3.2 Labels or instructions	7	YES	6	YES
3.3.4 Error prevention (legal, financial, data)	4		7	
3.3.5 Help	6	YES	7	YES
4.1.1 Parsing	1	YES		
4.1.2 Name, Function, Value	3	YES	6	YES

Source: Own author