Journal on Interactive Systems, 2023, 14:1, doi: 10.5753/jis.2023.3238

Exploring user reviews to identify accessibility problems in applications for autistic users

Mara Taynar Santiago [**Federal University of Ceara** | ***taynarsantiago@alu.ufc.br***]
Anna Beatriz Marques [**Federal University of Ceara** | ***beatriz.marques@ufc.br***]

Abstract

The Google Play Store provides various user reviews that can provide information about user experience, usability, and accessibility. Despite multiple studies addressing these reviews' importance and contributions to improving interactive systems, accessibility for users with Autism Spectrum Disorder (ASD) is still little discussed in this context. Considering the potential of user reviews, this article presents a textual analysis of reviews extracted from eight educational applications available in Portuguese with a focus on autistic children, namely: "ABC Autismo", "Aprendendo com Biel e seus amigos", "AutApp Autismo", "Autismo projeto integrar", "Jade Autismo", "Matraquinha", "OTO (Olhar Tocar Ouvir)" and "Teacch.me". We conducted an analysis based on the Guidelines for Accessible Interfaces for People with Autism (GAIA) and the BBC Mobile Accessibility Guidelines to classify user reviews.

Keywords: *user review, accessibility, mobile educational apps, Autism Spectrum Disorder*

1 Introduction

The market for mobile educational applications is constantly growing, especially in app stores such as the Play Store and the Apple Store (Koyani et al., 2004). Accessibility is essential in applications, ensuring everyone can use them without limitations. For this, guidelines and techniques that promote accessibility must be included in software development. It means that people with disabilities, including hearing, cognitive, neurological, physical, and visual impairments, should be able to use the system effortlessly, as the software and hardware will be adapted to their needs.

According to Acosta-Vargas et al. (2021), mobile application developers do not usually prioritize accessibility in their products, and the most commonly explored type of disability is the sensory disability, which includes visual and hearing disabilities, while cognitive and motor disabilities are little explored in accessibility evaluations. Britto and Pizzolato (2016) identified a gap in software development professionals' knowledge about accessibility for people with Cognitive, Neuronal, or Learning Disabilities (CNLD). This means that these types of users can use software with low accessibility. Autistic Spectrum Disorder (ASD) is within the CNLD classification. It is crucial to analyze the accessibility for Autistic Spectrum Disorder (ASD) because, when identifying gaps in the knowledge of professionals who develop the software, users with autism may face difficulties interacting with systems.

More and more educational applications aimed at the autistic public have been developed to support the development of skills (such as literacy and communication) and help in performing daily life activities (Branco et al., 2020). Some research has been conducted to investigate whether the quality of use criteria are met by these apps (Magaton and Bim, 2017; Abdul Aziz et al., 2015; Branco et al., 2020).

Marques and da Silva Monte (2021) concluded through a systematic mapping on the evaluation of technologies for autistic users that observation, questionnaires, and interviews are the most adopted methods in evaluating these technologies. Such techniques usually require selecting and recruiting users to share their usage experiences.

Branco et al. (2020) claim that if the technology is not adequately designed, it can generate interaction barriers preventing the user from using it autonomously or, in extreme cases, cause discomfort and unnecessary stress to the individual. Therefore, it is necessary to understand whether the software complies with accessibility guidelines for autism.

User feedback can provide valuable information about how an application works and help to improve it. User reviews are the considerations they make about an application, in which users can explain why they like or dislike an item based on their usage experiences (Koyani et al., 2004). Because they reflect the user's experience, we can use reviews to identify how to improve the software. It is natural for software to evolve, and developers understand the importance of using user feedback to drive improvements. Therefore, app stores promote a favorable environment for users to discuss their experience with the application, thus helping the development team understand which problems must be solved. The improvement of accessibility in an application is a valuable measure for all stakeholders, including developers and end users, who will enjoy a more efficient and inclusive experience.

The present study aims to investigate whether user reviews can indicate accessibility problems in educational applications for autistic children. The research problem lies in identifying any accessibility-related issues in user textual reviews. User reviews were extracted and manually analyzed from the Play Store to conduct the research. We classified the user reviews according to the categories defined by the accessibility guidelines from (1) the Web Interface Accessibility Guide focused on Autism aspects (GAIA) and (2) the mobile accessibility guidelines from the BBC (British Broadcasting Corporation, but abbreviated here simply as BBC). We used a keyword bank to assist in the review classification process.

This research is innovative by adopting an evaluation method few explored in evaluating technologies for autistic users (Marques and da Silva Monte, 2021). In addition, this

research combines the textual analysis method with two accessibility guidelines, one specific to autism.

This manuscript is an extended version of Santiago and Marques (2022) published in IHC 2022 (XXI Brazilian Symposium on Human Factors in Computing Systems). This paper is structured as follows: section 2 presents the theoretical foundation of this research. Section 3 summarizes the main related works. Section 4 describes the methodology adopted in this research. Sections 5 and 6 respectively present the quantitative and qualitative results obtained. A discussion of the results is presented in Section 7. Section 8 discusses conclusions and future work.

2 Background

This section presents concepts and theoretical bases adopted in this research: Autism Spectrum Disorder, user reviews, and GAIA and the BBC Mobile Accessibility Guidelines.

2.1 Autistic Spectrum Disorder (ASD)

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects communication, socialization, interest, and imagination skills (Association et al., 2014). It is recommended that parents and guardians seek professional guidance for diagnosis and stimulation to mitigate the disorder's negative impacts if confirmed.

Because it affects different abilities, ASD manifests itself in various degrees of severity: (i) Light Grade or Level 1 - needs few supports to live and perform basic tasks; (ii) Moderate degree or Level 2 - needs only some support to perform basic tasks and have problems with organization and planning; (iii) Severe Grade or Level 3 - have a severe deficit in verbal and non-verbal communication skills, have a narrow behavior profile, difficulty dealing with changes and difficulty in social interactions, and have reduced cognition.

Designing technologies appropriate for people with autism is necessary because an interaction and user experience designed inappropriately for people with autism can increase the effort of use and cause unnecessary stress, irritation, and anxiety (Britto and Pizzolato, 2016). Thinking about the difficulties faced by autistic people, there are currently several software programs aimed at this audience, especially for children, ranging from educational applications to aid in communication to therapeutic applications. Therefore, developers who see this new market niche as an opportunity must recognize the needs of their target audience and follow design guidelines that address the characteristics of autism to ensure the proper construction of these applications.

2.2 User reviews

Eler et al. (2019) studied how accessibility is treated in user reviews of Android system applications and classified the evaluations on application accessibility into two categories, they were: request, for when a user mentions a problem or requests a feature, and praise, for when agreeing that the application is accessible.

According to Vu et al. (2015), some factors can get in the way when analyzing the reviews. User reviews may have typos, acronyms, abbreviations, emojis, etc. In addition, 60% of the reviews do not contain helpful opinions. In the user reviews, there may be fraudulent reviews and spam, which can damage the image of an application within the store, as consumers rely on user reviews to decide whether or not to purchase the application.

Da Silva et al. (2019) argue that textual analysis of user reviews can account for many spontaneously provided user reports. The user reports are expressed without anyone's influence, which is essential to obtain user information. According to the same authors, user reviews can indicate: a) what is good or bad in the system, b) what needs to be modified, and c) obtain evolutionary requirements for the system.

Considering these points, the textual analysis of user reviews is a promising option for evaluating educational applications for autistic children. It is because this public may be inaccessible to researchers and professionals (Marques et al., 2021), and unfamiliar people in test environments can influence their behaviors and reactions throughout user tests (Melo et al., 2017). Therefore, obtaining spontaneous reports about the user experience is a potential source of relevant information.

2.3 Web Interface Accessibility Guide Focused on Aspects of Autism (GAIA)

The Web Interface Accessibility Guide focused on aspects of Autism (GAIA) was proposed by Britto and Pizzolato (2016) as a tool to aid in developing applications suitable for autistic people. GAIA provides design recommendations for interfaces and interaction projects that consider the needs of individuals within the autistic spectrum.

With the use of GAIA, it is possible to further reduce accessibility problems for autistic users by adapting application interfaces to their specific needs, which promotes better interaction between them and the available technology. The guide was developed by collating contributions in the literature of accessibility recommendations in designing interfaces for autism or other CNLD. It is freely available to the community of developers and software designers to democratize the knowledge gathered about accessibility for autistic people in WEB systems.

GAIA is divided into ten categories of recommendations. Each recommendation has a title, a description, characteristics of autism related to it, its importance, how to do it, and some examples of how to implement it. We briefly describe each category according to GAIA below:

- Visual and textual vocabulary: an approach on the appropriate use of texts and images, considering the specific needs of people with Autism Spectrum Disorder (ASD).
- Customization: Guidance for including application interface tuning options, allowing users to customize according to their preferences.
- Engagement: presentation of guidelines on maintaining focus and attention, as well as strategies to assist users in interacting with the system.

- redundant representation: refers to recommendations reinforcing that information should not be linked exclusively to a presentation format (text, image, or audio).
- Multimedia: details the appropriate use of multimedia resources in web interfaces to working memory, attention, visual and textual comprehension, and sensory integration of people with ASD, especially children.
- Responses to actions: explanation of why incomplete feedback or its absence is critical for people with ASD, particularly children, due to the commonly presented difficulties in retaining attention, dealing with changes, and understanding verbal instructions.
- Affordance: clarification of issues related to the design of interface elements. These must identify their operation without an investigation or cognitive effort.
- Navigability: recommendations include simplifying navigation between pages or screens so that the user can always know where he is in the application. In addition, it is crucial to allow the user to have complete control over navigation.
- System state visibility: The application must report the progress of tasks performed by the user, including clearly providing error information, help instructions, and information related to changes in the state of elements.
- Interactions with touch screens: it recommends adjusting the sensitivity of the interface to avoid accidental touches on elements.

More information about GAIA (available in Portuguese) can be found on the website: `https://gaia.wiki.br/`.

2.4 BBC Mobile Accessibility Guidelines

BBC (British Broadcasting Corporation) Guidelines are technology-agnostic best practices for mobile web content, Android and iOS. The English radio and television broadcaster BBC developed these guidelines.

The guidelines are divided into 11 categories and categorized as "Must" or "Must Not" for best practices that can be easily tested. Practices considered less testable but which can be regarded as fundamental for accessibility are categorized as Should or Should Not. Topics are Audio and Video, Design, Editorial, Focus, Forms, Images, Links, Notifications, Scripts and Dynamic Content, Structure, and Text Equivalences. Below is a brief description of each of the guidelines. However, the full report can be found in: `https://www.bbc.co.uk/accessibility/forproducts/guides/mobile/summary`.

- Audio and video: The guidelines cover visual and auditory content, autoplay, metadata, volume control, and audio conflict resolution.
- Design: Guidelines include colors, color meanings, styles and readability, size of touch targets, spacing, content resizing, interactive elements, focus, consistency in experience, choice in interaction, media adjustability, and avoiding flickering or blinking content.
- Editorial: The guidelines include the need for consistent labeling, indicating language changes, and providing supplemental instructions.
- Focus: The guidelines require that all focusable elements be interactive, that there are no keyboard traps, that navigation follows a meaningful sequence, that interactions are appropriate for the type of user, and that alternative input methods are available.
- Forms: The guidelines state that you must strip all form controls, properly group form elements, and maintain constant focus during user input.
- Images: The guidelines require that background images that convey meaning have an additional accessible alternative.
- Links: link and navigation content must uniquely describe the destination or function of the link, and repeated links to the same resource must be combined into a single link.
- Notifications: the guidelines recommend that notifications are visible and audible and that error and correction messages should provide clear information. In addition, feedback and assistance should be provided where appropriate.
- Scripts and dynamic content: Applications must be developed progressively, and the media must have an alternative to pause, stop or hide. Pages may not automatically refresh without prior notice, among other guidelines.
- Structure: All pages or screens must be clearly and uniquely identified. Headings should follow a logical and hierarchical structure, allowing users to understand the organization of the content.
- Text equivalents: Providing alternatives for non-text content, such as images, is necessary. Decorative images must be marked as decorative to be ignored by assistive technologies. Elements must have accessibility properties appropriately set to ensure they are accessible to users with disabilities.

3 Related works

This section briefly presents some works that explored the textual analysis of user reviews.

Freitas et al. (2016) investigated user reviews of the Spotify streaming app. The authors explored Usability and User Experience (UUX) through user reviews using the MALTU method (Model to Evaluating Interaction in Social Systems Based on the User's Textual Language). As a result, the authors found that 76.1% of user rewrites were critical, while 16.8%at were related to doubts, praise represented 5.4%, and 1.8% were suggestions. Many reviews of the required type were because user reviews were extracted from the Reclame Aqui website. This platform allows consumers to register their complaints about the experience of buying products or services.

Genc-Nayebi and Abran (2017) conducted a systematic literature review on user reviews in app stores o find solutions for mining user opinions. User behavior trends were analyzed, including how users give credibility to app feedback and ratings and how there are experts who conduct fraudulent reviews and spam to either harm or boost the reputation of specific software. The authors concluded that it is necessary

Table 1. Summary of related works

Related work	Evaluated quality criterion	Type of evaluation	Target audience
Freitas et al. (2016)	Usability and User experience	Manual	Spotify Users
Genc-Nayebi and Abran (2017)	User experience	Manual	General public
Eler et al. (2019)	Accessibility	Manual	General public
Alqahtani and Orji (2019)	Usability	Both	People with mental health issues
Wang et al. (2021)	Relevance	Manual	Spotify users
Diniz et al. (2022)	Usability	Manual	General public
Aljedaani et al. (2022)	Accessibility	Automatic	General public
This research	**Accessibility**	**Manual**	**Autistic users**

Table 2. Evaluated Educational Apps

Name	Number of app reviews	General App Note	Release date
ABC Autismo	1380	4,0	October 25, 2013
Aprendendo com Biel e seus amigos	70	4,0	April 26, 2018
AutApp Autismo	39	4,3	October 15, 2017
Autismo projeto integrar	31	4,5	May 4, 2016
Jade Autismo	216	3,4	May 24, 2018
Matraquinha	484	4,2	July 25, 2018
OTO (olhar tocar ouvir)	94	4,7	June 11, 2015
Teacch.me	5	5,0	June 13, 2016

to develop an exclusively specific model for app store analysis, given the nature of review texts, and that user reviews can be a source for extracting information on user experience and usability and used for removing new requirements.

Eler et al. (2019) conducted a study on accessibility on user reviews on the Play Store. The authors sought to understand whether analyzing user reviews could reveal accessibility problems in the application. Seven hundred-one general public applications were selected to carry out the study. The authors used keywords that referred to accessibility issues, which might be able to return reviews about accessibility issues. They used a keyword bank obtained from the BBC Mobile Accessibility Guidelines. The authors found that it was possible to find reports about accessibility in 1.24% of the user reviews. The authors concluded that the results found about the applications should be taken more into account by the development team.

Alqahtani and Orji (2019) conducted an assessment of user reviews from mobile app stores, Play Store and App Store. The research aimed to identify usability problems in mental health applications. To select the apps, searches were performed on the Play Store using words referring to mental health problems (such as “anxiety” and “depression”) and found 106 apps related to the topic. The authors found 1236 user reviews on usability. The analyses were classified into six categories: bugs, poor interface design, data loss, battery, and memory usage problems, lack of guidance, explanation, and problem with the Internet connection. The vast majority of users did reviews to report bugs found in the application (820 reviews), and to report poor inter-research aimed at other types of studies were less frequently found in studies made by users; none totaled above one hundred cases. The research found several problems in the applications and highlighted that the mental health area is a sensitive subject, and usability found 1236 user reviews on usability. The analyses were classified into six categories.

Wang et al. (2021) explored the relevance of user reviews in app updates based on release notes. They collected user reviews and release notes of Spotify in the Apple App Store and manually determined the relevance of the app reviews concerning the release notes. Word2Vec calculation techniques were applied to determine the correlation between user reviews and release notes. Their results showed that more than 60% of user reviews corresponding to release notes are irrelevant.

Diniz et al. (2022) aimed to investigate the presence of heuristic usability issues through user reviews. They collected and analyzed 200 reviews, from 10 Android and iOS apps. Three researchers analyzed the reviews individually and then discussed obtaining a common classification result. The results pointed out that 25% of the user reviews indicate heuristic issues. Most problems were related to user error recovery and match between the system and the real world.

Aljedaani et al. (2022) used an automated approach based on supervised learning techniques for classifying accessibility app reviews into four categories of BBC guidelines: Principles, Audio/Video, Design, and Focus. They applied the supervised learning techniques: Extra Tree Classifier (ETC), Random Forest, Support Vector Classification, Decision Tree, KNearest Neighbors (KNN), and Logistic Regression. 2,663 Android app reviews were classified, and the results have shown that the ETC classifier produces the best results in the automated classification of accessibility app reviews with 93% accuracy.

Table 1 summarizes some characteristics of the related workers aiming to understand this research’s differential. Regarding the evaluated quality criterion, this study is similar to the analysis of Eler et al. (2019) and Aljedaani et al. (2022) by focusing on accessibility aspects. However, the target audience of this research is autistic users; while the study of the above has no delimited public, they are users of general public applications. Therefore, the difference in this research lies in adopting the textual analysis method to evaluate technologies for a specific target audience.

4 Methodology

The study methodology had the following steps: selection of applications, extraction of user reviews, use of keywords, classification of user reviews, and analysis of results. We adopted two guidelines to study user reviews: GAIA and BBC. We selected the GAIA guidelines because it was used to design and evaluate technologies for autistic users and consider aspects specific to ASD as content, customization, and engagement (Britto and Pizzolato, 2016). However, GAIA is not specific to mobile applications. Thus, we adopted the BBC guidelines, which focus on general accessibility recommendations for mobile applications related to images, structure, and notifications (Eler et al., 2019).

4.1 Selection of educational apps for autistic children

Educational apps aimed at autistic people were selected based on research from Branco et al. (2020), in which the researchers surveyed mothers of autistic children and their respective children. While conducting this research, all apps were accessible on the Play Store and in Portuguese. The educational apps evaluated in the mentioned work were: learning with Biel and his friends, ABC Autismo, Jade Autismo, OTO (look, touch, listen), Matraquinha, AutApp Autismo, Autismo projeto integrar, and Teacch.me. Table 2 provides some information about the applications. The information presented refers to the period in which the extraction of user reviews began. It is essential to point out that some information, such as the number of evaluations and grades, may constantly change, making the information presented outdated.

The authors' research aimed to evaluate the user experience and accessibility of selected applications for the education of autistic children. Such applications used different educational methods for this audience of children; the authors used the GAIA and the BBC guide to assess the accessibility of the selected applications. To add evidence about the accessibility of the educational applications evaluated in their work, it was decided to analyze user reviews of the same applications considered in the research. Due to the Branco et al. (2020) research being conducted in Brazil with Brazilian children, both the applications and user reviews considered in our research are in Portuguese.

4.2 Extracting user reviews

User reviews of the selected apps were manually extracted from the Play Store, and this was possible as the number of studies found across all apps needed to be more significant. The first extraction took place on June 4, 2020. This first extraction was applied to the Jade Autismo app. The extractions of the other reviews were completed on June 25, 2020. All extracted reviews were stored in a spreadsheet for later analysis. In all, 777 user reviews related to the eight apps were removed. No specific criteria were established for collecting reviews; all reviews available in each application were collected. Information about the authors of the reviews should not have been collected to preserve their identities.

4.3 Use of Keywords

At this stage, we used GAIA to select keywords that could refer to accessibility problems in educational applications. We defined keywords following the GAIA guidelines, stipulating terms related to each category of accessibility guidelines. As a result of this step, we generated a GAIA keyword database. Table 3 presents the keywords defined for each GAIA category.

The definition of the keywords happened as follows, the descriptions of the accessibility recommendations were analyzed, and from this analysis, some words were extracted, words that could summarize the guideline. Below is an example of how it happened: the keyword "Feedback" was removed from recommendation 6.1: "Provide feedback confirming correct actions or warning about possible errors and use audio, text, and images to represent the message, avoiding icons that involve emotions or facial expressions". Morphological variations were not considered, as keywords were not a determinant for ranking user reviews. The keyword search process for the BBC guide was simplified, as the researchers used a set of keywords existing in the literature used in Eler et al. (2019).

We noticed that for this research, manual classification would be more effective. Although the current base is not comprehensive enough, the base is a research contribution and can be evolved in the future.

4.4 User reviews classification

The collected reviews went through the initial stage of keyword search, where manually, each keyword was searched for all the extracted sets of reviews, and this step was carried out in Excel. Each review that had a keyword was marked. In this stage, we adopted two sets of keywords: one for GAIA and another for the BBC guide.

In the next stage, we read all the reviews, and we tried to identify if the study had any report of accessibility. We validated this classification regardless of whether the reviews had been returned in the keyword search stage. At this point, it is essential to highlight that using keywords alone was inefficient in the final classification of particular keywords. False positives were returned, and many reviews about accessibility were not identified. Therefore, a manual analysis was used to understand the nature of the accessibility reports better. They were classified into four categories: praise, criticism, suggestion, and bugs. It is important to note that a single review can be classified into multiple categories.

Next, we created a new spreadsheet with only accessibility reviews so that the reviews could be reanalyzed, now observing if we could identify any GAIA guidelines within the accessibility report. Again, each review was classified and validated by both researchers. In case of disagreement in the review classification, the researchers discussed reaching a consensus based on the definition of GAIA categories.

In order to broaden the scope of identifying accessibility issues beyond those specific to autism, it was decided to utilize the Mobile Accessibility Guidelines from the BBC. The inclusion of these guidelines aimed to provide additional evidence regarding the accessibility status of the applications.

Table 3. Bank of keywords referring to GAIA categories

G1 - Visual and textual vocabulary colors, contrast, content, distinguish, simple language, long text, lots of text, symbol, reading, structure, icons, vocabulary
G2 - Customization increase, text size, change, font, color, sound, customize, change, personalize, layout, elements, preferences, quantity, functionalities, reading mode, printing mode
G3 - Engagement distraction, distract, attention, focus, simplicity, clarity, understanding, understand, simple interface, content, instructions, guidelines, motivate, engage
G4 - Redundant representations caption, instructions, representations, representation, understanding, symbols, pictograms, image, audio, video, figure, icon
G5 – Multimedia different representations, text, content, understanding, video, audio, image, attention, visualization, amplification, sound, disturbing, explosive
G6 - Responses to action response, feedback, right actions, wrong actions, right, wrong, alert, confirmation, emotion, expression
G7 – Affordance similar, consistent, predictable, seem clickable, adequate, instruction, feedback
G8 - Navigability simple navigation, consistent navigation, bookmarks, page, exit, return, home page, help, time, redirect, control
G9 - System state visibility instruction, message, error, resolve, cancel, undo, commit, revert, restart, attempt
G10 - Interactions with touch screens screen, touch, sensitivity, sensitive, error, selection, accidental touch

It is a consensus among the researchers that the best guideline is the GAIA, which was built for the autistic spectrum. However, we considered it valid to use a second source of accessibility guidelines. Analyzing and classifying the BBC guideline reviews occurred similarly to GAIA. The researchers read and tried to identify the reviews that addressed accessibility and attempted to identify if the review addressed any of the issues described in the BBC guideline.

4.5 Results analysis

The results obtained from the analysis of user reviews were analyzed quantitatively and qualitatively. In the quantitative analysis, we counted the accessibility issues reported by users. We conducted a qualitative analysis to diagnose the main aspects of accessibility that were affected by the applications. The analysis of user reviews (available in Portuguese) can be consulted at `https://bit.ly/3e6wIiM`.

5 Quantitative Results

The following presents the results discovered in the study of the accessibility of educational applications, we generated graphs of this data, and in this section, we describe the results.

5.1 User reviews and accessibility reports

Figure 1 shows the number of reviews each app pulled from the Play Store. The graph shows that the ABC Autismo application is notably prevalent. Of the selected educational applications, it is the one that had the most reviews extracted, while Teacch.me being the one with the least extractions (only two reviews).

After extracting the reviews, we found that, as predicted by Eler et al. (2019), many of them were 'noisy'. That is, they contained elements that added little to the main message, such as typos, acronyms, and emojis, or the review did not present contributions to the state of the application's accessibility. Among the extractions performed, 125 reviews (16%) presented accessibility reports, as shown in Figure 2, which summarizes the initial identification process of these reports.

Figure 1 illustrates that the apps with the most reviews containing accessibility reports were Matraquinha (73 reviews being accessibility) and ABC Autismo (33 reviews being accessibility). In user reviews of the Teacch.me app, there were no reports of accessibility.

When analyzing the nature of the reports in accessibility Figure 3, it is possible to notice that most users complained about bugs in the application. This type of report corresponded to more than half of the total number of reports (57%), which the urgency can explain that users felt in solving these problems, as this could be preventing the use of the application. Then, we observed that users often left suggestions for improvements that impacted accessibility (28%), 14% of users criticized some aspects of the application's accessibility, and finally, only 1% of the reviews praised accessibility.

5.2 Classification of user reviews by the guidelines used

After being classified as accessibility-related, the review was subjected to a new manual review to verify whether it fits any of the GAIA or BBC guidelines for accessibility. The manual classification was fundamental since, in some cases, the simple presence of a keyword was not enough to identify reviews with accessibility reports. In addition, there were situations in which the reviews mentioned aspects of accessibility but did not include any previously defined keywords. Therefore, manual analysis proved to be essential for the proper classification of reviews.

Figure 4 illustrates the classification of user reviews of applications according to GAIA guideline categories. We found all GAIA guidelines categories in the textual analysis. However, we observed that user reviews more frequently pointed to recommendations from the categories: G6 - Responses to actions (76 times), G9 - Visibility of the system state (71 times), G2 - Customization (35 times), and G4 - Representations redundant (9 times). Notably, a review could be classified in more than one GAIA category.

The classification of the reviews in the BBC guidelines proved less comprehensive than that of the GAIA. Only 47 reviews were related to some of the guidelines of the BBC guide. This number was reduced because the BBC accessibility guidelines were aimed at the general public. However, as the GAIA was developed with the specific needs of people with autism in mind, it was able to be more comprehensive and inclusive. Figure 5 shows the rankings results for

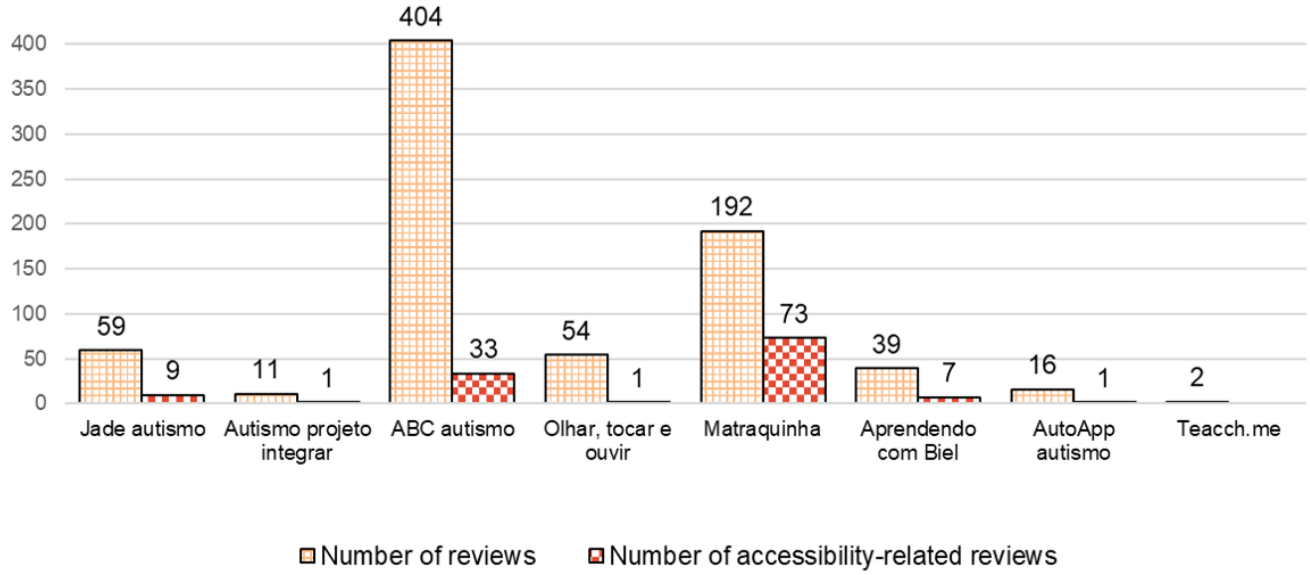

Figure 1. Number of user reviews by app

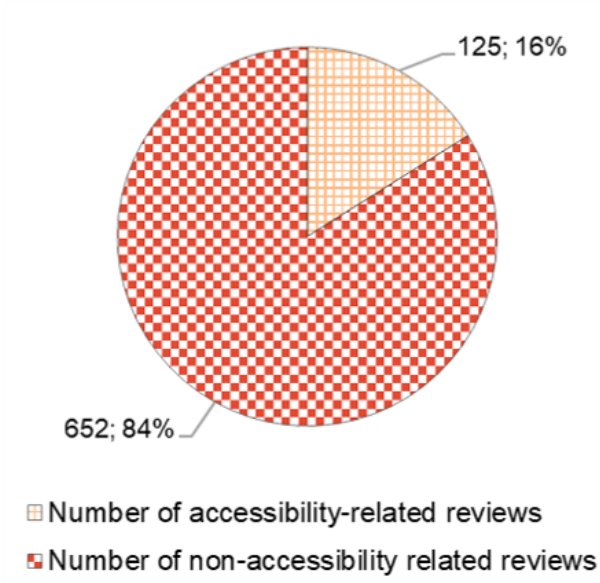

Figure 2. Overview of extracted user reviews

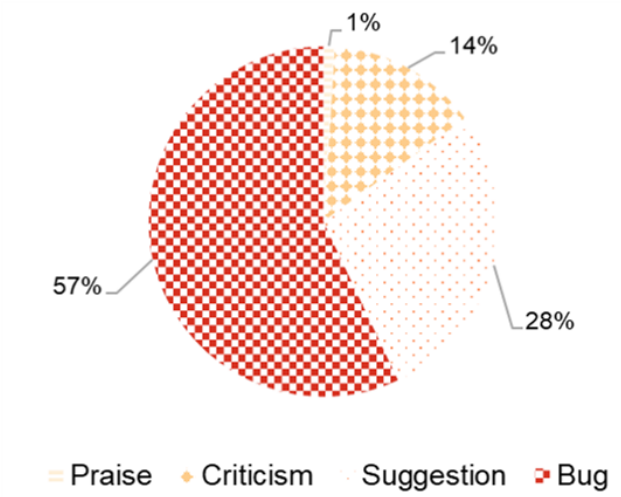

Figure 3. Type of accessibility reviews.

the BBC guide.

Below, we explain each app's results, followed in order, GAIA classification results, followed by BBC classification results.

6 Qualitative Results

This section describes the accessibility issues reported in user reviews of each application. We organize the results in subsections describing the issues identified through GAIA and BBC.

6.1 ABC Autismo

6.1.1 GAIA

The ABC Autismo educational app is very popular with its target audience. It was the second with the highest number of reviews with accessibility reports. In Figure 3, we can see that accessibility reports are related to almost all GAIA categories. However, we can also see that two categories appeared in most reports: G9 - Visibility of the system state (19 reviews) and G6 - responses to actions (22 reviews).

The numbers of reports for the two categories are similar, as it was prevalent for the same user review to be classified in both guidelines, as in the example of the review below:

"I'm so sad it is not opening on my phone. My daughter loves it." - This review falls under guideline 6.1 of the G6 category (action response): *Provide feedback confirming correct actions or warning about possible errors and use audio, text, and images to represent the message, avoiding icons that involve emotions or facial expressions*. In this case, the application should provide feedback on what is happening in the application, and this does not happen. This same review is classified in guideline 9.1 of category G9 (system status visibility): *Present proper instructions for interacting with page elements, provide clear messages about errors and mechanisms for resolving errors*. The app is not delivering a clear message about the error preventing its use.

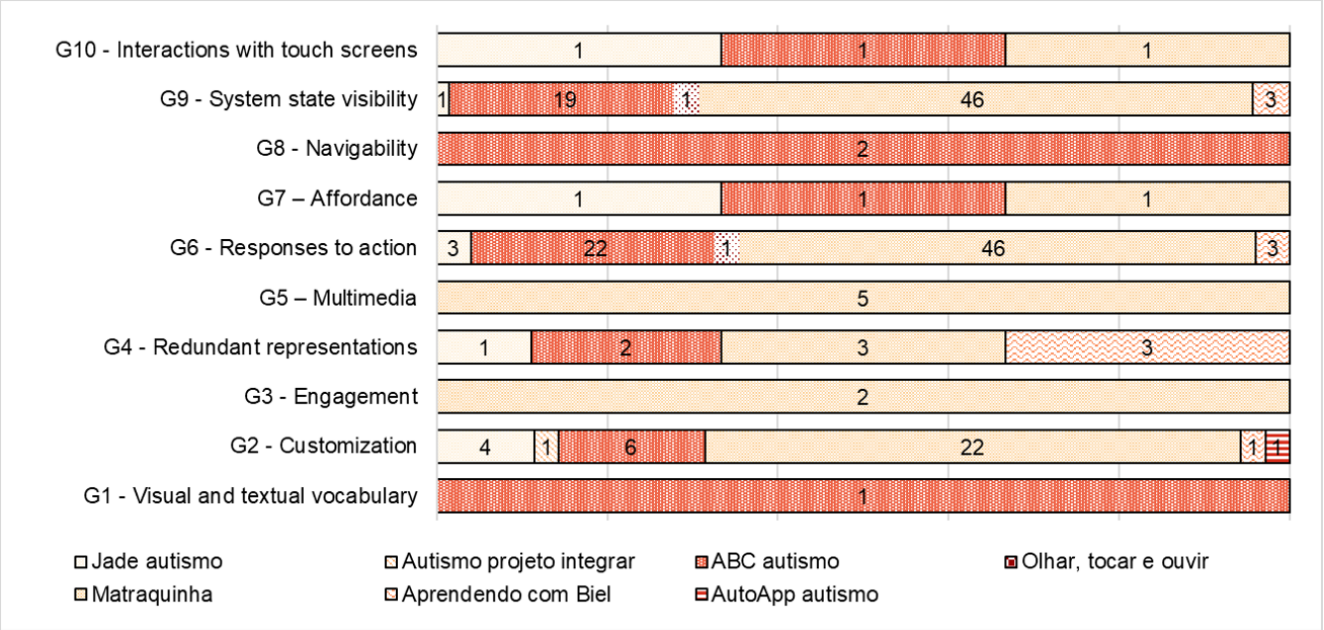

Figure 4. User review classification by GAIA category

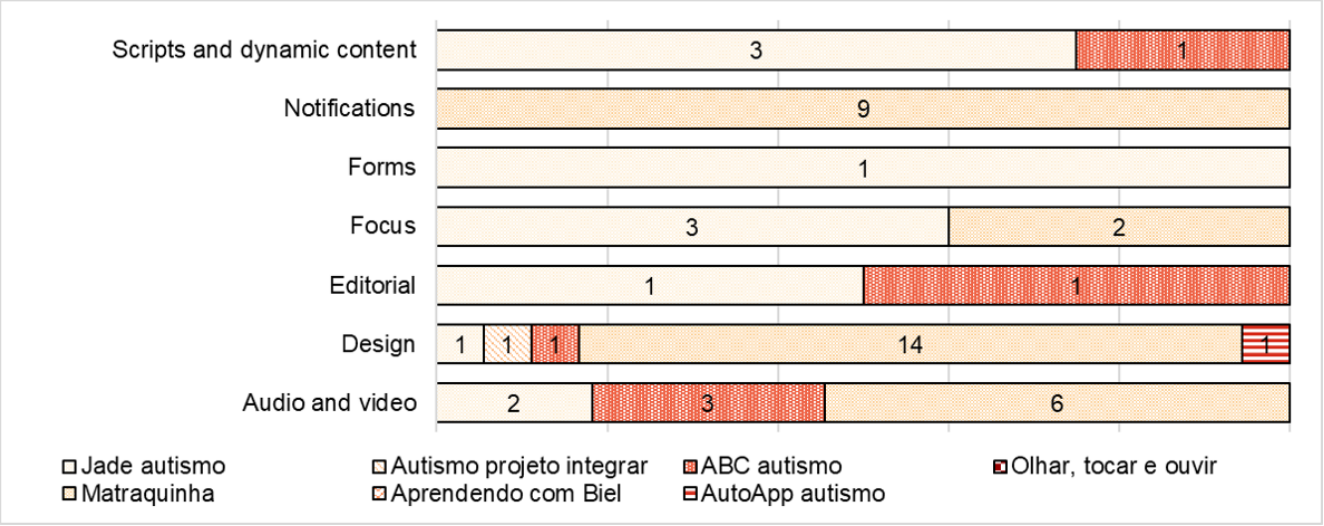

Figure 5. User review classification by BBC guide category.

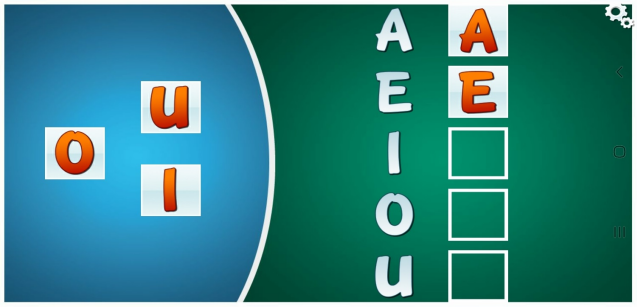

Figure 6. Accessibility report in the app ABC Autismo

Figure 6 demonstrates an accessibility issue identified by a user in the following review: *"It could have options to return to the main menu at any time after selecting the game. The options for building words with syllables are few. They could have words more common to the context of children."* This issue is associated with guideline 8.1 of category G8 (navigability): *Provide simplified and consistent navigation between pages, using location indicators, progress indicators, and presenting global navigation buttons on all pages.* The app should provide global navigation options on the mentioned screen, but performing any navigation action before completing the task is impossible. The interface has a gear icon, but nothing happens when selected. This issue is related to guideline 8.1 of category G8 (navigability).

All reviews in the responses to actions (G6) category were classified under guideline 6.1. This guideline aims to use immediate feedback as an essential role in retaining attention, understanding instructions, and reducing incorrect actions by autistic users. Based on the analysis results of user reviews, we can consider that the app has difficulty capturing the attention of autistic children, conveying necessary instructions for use, and that children are likely to make mistakes when using the application.

Similarly, all reviews associated with system status visibility (G9) were classified under guideline 9.1. This guideline describes that interactive elements and functionalities should guide how the interaction should be performed and provide

solutions for errors when they occur, always with simple language to facilitate the user's life. Based on the result of the accessibility analysis of the app, it was not providing adequate guidance for its functionalities and elements or not satisfactorily providing error resolution for the user.

6.1.2 BBC

Analyzing the reviews from the perspective of a generic mobile accessibility guideline, we found accessibility problems in the following topics: notifications, editorial, audio and video, design, scripts and dynamic content. Figure 5 shows the results found.

A report that may show audio and video problems in the application is at *"I gave it five stars because it is really good, but to make it better, it could have audio options like drag the letter a and then say the name of the letter and then say what was formed with the letters."*. This review has been categorized as Alternatives to Visual and Audio Content, and this is justified as the guideline establishes audio description as embedded media.

6.2 Aprendendo com Biel e seus amigos

6.2.1 GAIA

In this application, we identified accessibility problems in four GAIA categories: G2 - customization, G6 - responses to actions, G4 - redundant representations, and G9 - visibility of the system state. The last three categories concentrated the highest recurrences of application accessibility problems.

In redundant representations (G4), the most frequent guideline was 4.1. This guideline describes that applications should not focus only on textual language to give instructions, commands, and present content. The explanation is that autistic people may struggle with verbal or non-verbal communication. Ideally, autistic users have multiple options for receiving the content. Thus, the application, Aprendendo com Biel e seus amigos, may need help to provide alternative means of representing the content for its users.

The main guideline for responses to actions identified (G6) was 6.1, which means that the application does not follow the appropriate recommendations to get the user's attention. If they are distracted, they may make mistakes. In the visibility of the system's state (G9), all accessibility reports were classified in guideline 9.1. It means that the application may need to have oriented correctly about its functionalities and elements or that it did not provide help for solving errors. Figure 8 shows an example of an activity, but the application does not offer clear instructions on interacting with it for its resolution.

An existing accessibility review of this app is cited below:

"Unfortunately, it has some flaws. In the L syllable family, they say ba be bi bo bu. In the word VELA, they say VEJA. To be charged, you must help the child and not confuse their reasoning."

This review belongs to the 4.1 guideline, which defines: *The application should not only focus on texts for content presentation but also provide representations in image, audio, or video and ensure that these representations are close*

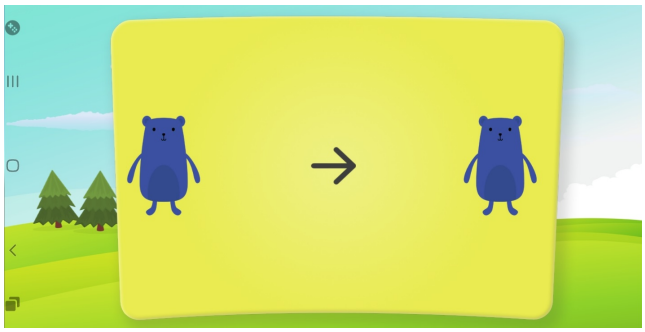

Figure 7. Accessibility report in the application Aprendendo com Biel e seus amigos

to the corresponding text. The picture is far from the corresponding text, as it makes an error when demonstrating word pronunciation.

6.2.2 BBC

The application did not have reviews classified under any BBC Mobile Accessibility Guidelines.

6.3 AutApp Autismo

6.3.1 GAIA

The application presented only one accessibility report. The user review was as follows:

"A tip: the typeface most used for reading by children between 6 and 7 years old is Capital Stick, as they are in the literacy phase. This font favors reading and writing." In Figure 9, we can see that the font is lowercase, and the user reports the desire to leave it in uppercase.

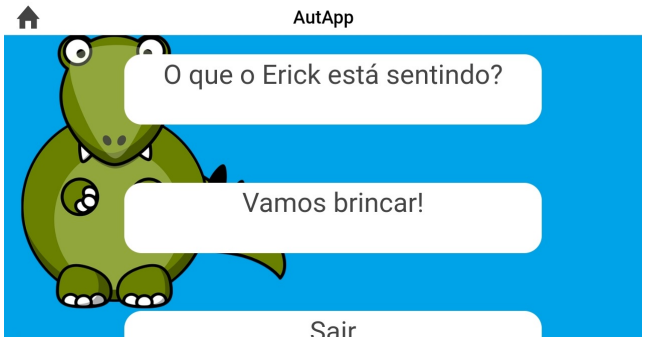

Figure 8. Accessibility report in the application AutApp Autismo

We classified this review in the customization category (G2), guideline 2.1. This guideline establishes that the application allows customizations in colors, text sizes, fonts, and sounds according to the user's preferences.

6.3.2 BBC

The application did not have reviews classified under any of the BBC's mobile accessibility guidelines.

6.4 Autismo projeto integrar

6.4.1 GAIA

As a result of the textual analysis phase, only one review was observed that had an accessibility report. This review was de-

scribed as being from the customization category (G2) and guideline 2.3, which says that the software must offer customization options for the quantity and layout of elements on the screen and customize the functionalities.

The user review says the following:

"Even though it is very simple, I found it really good. Could you create a way for users to develop personalized routines that fit each phase of autistic children, teenagers, or adults? I loved it!"

This classification is because the user feels the need to create personalized routines. As seen in Figure 10, the app has existing tasks, and the user can only choose one of them and customize the time and frequency of that task.

Figure 9. Accessibility report on the application Autismo projeto integrar.

6.4.2 BBC

In the classification according to the BBC Mobile Accessibility Guidelines, the review was identified as being in the Design category, more explicitly belonging to the Adjustability guideline of this category. The guideline is justified because it establishes that interactive media should be adjustable to the user's preferences. However, this only happens when the user tries to personalize a task according to their interests, and the routine cannot be adjusted.

6.5 Jade Autismo

6.5.1 GAIA

Analyzing the GAIA guidelines that appeared most frequently in Jade Autismo reviews, we can observe that users' main difficulties with the application's accessibility were related to the customization (G2) and response to actions (G6) guidelines. The first category defines that users should have the option to adjust the interface according to their preferences. Problems were also identified in four other GAIA categories, namely: redundant representations (G4), interactions on touch screens (G10), system state visibility (G9), and affordance (G7).

One of the problems identified by a user in the Redundant Representations category (G4) is described in a review and illustrated in Figure 11:

"The game should say the names of the animals and the colors and not make the same sound for everything."

The classification of this review in the Redundant Representations category (G4) occurs because category 4.1 states that content presentation should adopt through multiple means (text, image, and sound). In this case, only an image is represented, and the sound does not represent the figure.

Figure 10. Accessibility report on the application Jade Autismo

The second most common category was Response to Actions (G6), which mentions the need to provide appropriate feedback for interactions, considering the characteristics of the application's target user. An example of a user review classified in this category is:

"Excellent application, but suddenly it stops working."

In this review, the user praises the application but mentions that the app has stopped responding and needs to give feedback on what is interrupting the proper functioning of the application.

In the Customization category (G2), the leading guideline that appeared was 2.3, which may indicate that the application is not allowing users to customize the layout and quantity of elements in the interface. In the Response to Actions category (G6), the most general guideline was 6.1, which means that, for its users, the application is not providing adequate responses for correct actions or alerts for possible errors (wrong actions by the user).

6.5.2 BBC

It can be observed that there was a greater distribution of reviews among the categories, where those that had accessibility reports were: Scripts and dynamic content, focus, and audio and video. In one of the user reviews, they report about the application:*"Very good app, I liked it, and I want to congratulate you!!! Just constructive criticism, the letters should be in shape for better visualization of the autistic, and the order of placing from top to bottom was not necessary. The important thing is the pairing. Bjss"*, this review has been characterized as being from three categories of the BBC Mobile Accessibility Guidelines, Design and Scripts and Dynamic Content and Focus.

It was classified as being from the Design category guideline, as the guideline says that interfaces should provide multiple ways to interact with content, which according to the review, still needs to be done. In the Scripts and Dynamic Content classification, it is defined in the Input Control directive, as the font could not be changed. In the Focus category, it was defined as the order of focus, as the user needed help understanding the logic in the sequence in which cards should be paired.

6.6 Matraquinha

6.6.1 GAIA

This application had the highest number of accessibility reports. Reviews mentioning accessibility issues were present

in 38% of all extracted reviews. We found accessibility reports that violate eight of the GAIA guidelines. The most frequent reports were in the categories of customization (G2), response to actions (G6), and system state visibility (G9).

The customization guideline (G2) that appeared most frequently was 2.2, which recommends giving different ways to users view information, such as through audio, text, and images, so that they can choose how they want to view it. The application had several reports in this category, so the application may not have allowed users to view the information in different options.

One report that relates to the customization category (G2) is:

"I think the game is very efficient, but I want to leave a tip for emotions: it would be nice to have the option to put our photos in place of the stick figures."

This report is classified as accessibility because, in guideline 2.3, there is a recommendation that the user has the option to customize the functionalities, and the application does not offer this possibility. In Figure 13, we can see the activity that the user describes and reports that he would like the option to customize it.

The responses to actions category had all reports classified in guideline 6.1, meaning that the application does not provide adequate feedback to the user's actions. The main guideline found in the system state visibility category was 9.1, representing that the application may not have provided instructions on interacting with elements and functionalities or did not help in the error resolution process.

6.6.2 BBC

The category with more accessibility problems was Design, with 15 reports. The following review is classified in this category: *"I thought it was very cool, but the speech induces a mechanics of the child who will imitate. It could be a more natural speech since it also stimulates orality in addition to stimulating communication!"*, this report is found in the guideline of adjustability of the project category, as the user reports the desire to change the character's speech in the application, which was impossible.

6.7 OTO (Olhar tocar ouvir)

6.7.1 GAIA

This app only had one review reporting accessibility. It was:

"The APP does not sound on my son's tablet."

Despite being only one, it has been classified under GAIA guidelines 6.1 and 9.1 as it violates guideline 6.1 when it fails to emit sound feedback and violates guideline 9.1 because the application is not providing a resolution for the issue of no sound.

6.7.2 BBC

The application did not have reviews classified under any of the BBC's mobile accessibility guidelines.

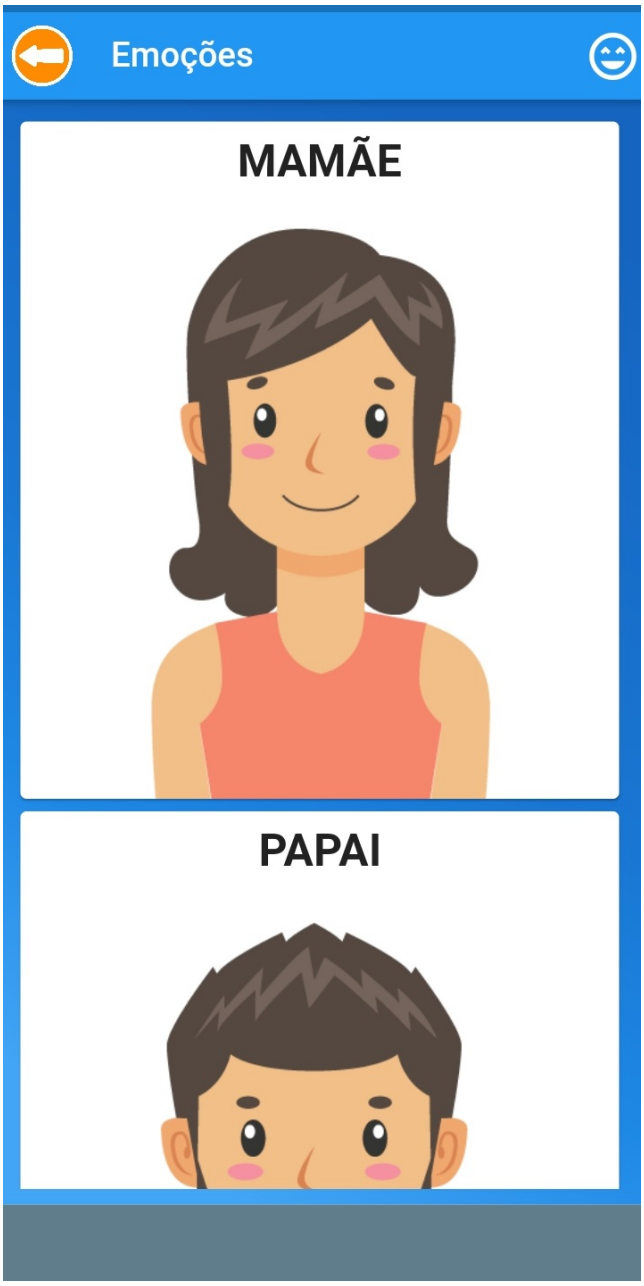

Figure 11. Accessibility report on the application Matraquinha

7 Discussion

Through textual analysis and accessibility classification of user reviews, we observed that most accessibility reports were related to two applications: ABC Autismo and Matraquinha. In addition, the accessibility reports focused on some specific problems of the two applications, and in general, the main complaints were of the bug type. This result is similar to what was also found in the study by Alqahtani and Orji (2019), who used user reviews of mental health apps to find usability issues.

The most frequent problem of ABC Autismo was related to Responses to system actions (according to GAIA) and Notifications (according to BBC). It happened due to the many reports that applications did not open on users' devices, stopped working, or crashed. The Matraquinha application was the one that had the most reviews of accessibility problems. The main guidelines affected, according to GAIA,

were visibility of the system state, responses to actions, and customization. In contrast, according to BBC, the main ones were design, audio and video, and notifications. The application had many reports that the sound was low or even absent, which explains the large number of reports related to the visibility of the system state, response to actions, audio and video, and notifications. At the same time, the problems in customization and design were related to frequent reports of users expressing the desire to customize application cards.

To further explore how the GAIA and BBC guidelines are related to the reviews analyzed, we developed the Sankey chart to illustrate the relationship between the guidelines as reflected by the ranking results (Figure 12). The graph shows how reviews ranked in each GAIA category are related to the BBC categories. Wider arrows indicate a higher occurrence of user reviews classified with GAIA and BBC categories. You can see which categories from each set of guidelines are most closely related according to the results of our analysis.

It was already expected that, by following the GAIA guidelines, more accessibility problems would be identified since it was developed taking into account the main difficulties that an autistic person may have while using the app, such as maintaining focus and the need for adaptations in the font used, for example.

Taking the GAIA categories as a starting point, it can be seen that reviews classified as Customization (G2), were also widely classified with BBC Design and Audio and Video. Reviews classified as Responses to actions (G6) and System State Visibility (G9) were also classified as Notifications. The BBC Design category has a relationship with several GAIA categories, as it involves aspects of color, contrast and customization features, focus, and choice, which are considered by different GAIA categories. Although BBC is not specific to autism, this analysis indicates that it still allows for identifying accessibility issues affecting this audience.

We can consider that BBC is more generic and cover some accessibility aspects that also affect users with ASD. On the other hand, GAIA has guidelines about accessibility aspects related to ASD characteristics. Thus it has a more significant potential to support identifying accessibility issues that impact this specific public. Our results confirm that hypothesis, as we identified more user reviews related to GAIA guidelines than BBC ones.

The research by Branco et al. (2020) also assessed the accessibility of educational apps for autistic children using the inspection method. Through the textual analysis conducted in this research, we evaluated apps not considered in previous studies, such as Matraquinha, Autismo projeto integrar, AutApp, and Teacch.me. The reviews report experiences from different users, which enriches the evidence about accessibility problems faced. It was also possible to identify problems related to the category of touchscreen interaction, for which there was no evidence of accessibility problems in previous research.

In this way, it is possible to highlight that user reviews are a potential source of information about autistic people's usage experiences. This methodology pointed out accessibility problems identified in usability inspections conducted in previous research. An example is the ABC Autismo, where the Visibility of the System State guideline appeared more frequently in Branco et al. (2020) and in this research. We observed similar results for the OTO application.

As a limitation of the textual analysis in the context of evaluating technologies for autistic users, it is impossible to relate the problems identified to the different degrees of autism of the users, as this information is not available on the Play Store and other app stores.

It is essential to point out that applications frequently undergo maintenance, which can be adaptive, corrective, or evolutionary, since this is part of the life cycle of software. The accessibility problems mentioned in the article may have already been solved in the application. In addition, new obstacles related to accessibility may have arisen. When writing this article, some of the apps used in the study were no longer available on the Play Store. They are: ABC Autismo, AutApp, Aprendendo com Biel e seus amigos and Teacch.me. However, some apps are still available for download on other platforms and can be installed via APK files. More current reviews may bring new results not identified in this research, which considered reviews up to 2020.

However, the objective of the research was to explore whether it was possible to find accessibility problems using the proposed methodology. Although the research is not intended to provide a conclusive diagnosis of the accessibility of applications, we can consider that user reviews, often informal, can reveal the negative impact of a poorly designed interface design. As seen in the article, these users often make claims for improvements in the application due to the difficulties they encounter. It is important to note that not all users have in-depth knowledge of accessibility guidelines. As a result, it may be difficult for them to identify what accessibility is. For example, when users request dark screen mode to be included, they may not know that this functionality reduces eye strain and improves accessibility.

8 Conclusion and Future Work

This work aimed to conduct a textual analysis of the reviews of Play Store users who use educational applications for autistic children to understand how users addressed the issue of accessibility in their reviews. We adopted two guidelines to classify the accessibility problems identified: GAIA and BBC.

During the classification stages, it was possible to identify two behaviors of the users. The first one is that the main authors of the reviews of the educational applications were not the children, but their guardians, who closely followed the use by the autistic child. Second, it was clear that those responsible looked favorably on educational applications, and it was not simply another application that their children used. They were confident in the benefits the applications intended to bring and used the application as an auxiliary tool in the educational process, such as literacy, teaching phrases and emotions, creating routines, etc.

Reviews made by users have proven to be an excellent way to understand the applications since they represent the feedback of those who use the apps daily to fulfill their tasks. In this way, the evaluations allow a more complete and in-depth view of the quality and effectiveness of the applications. The

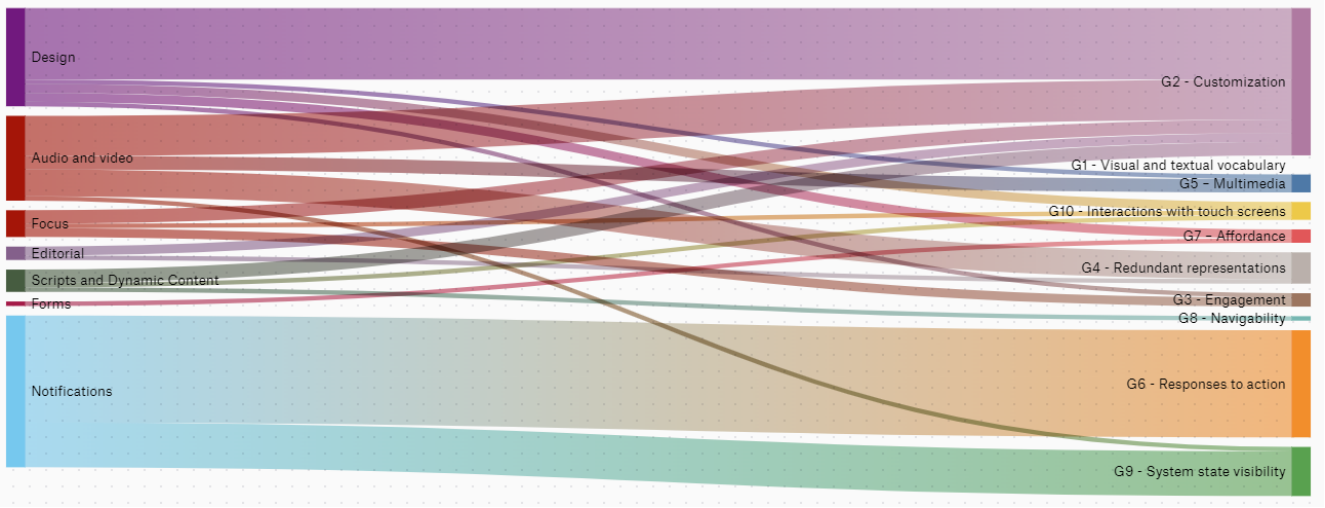

Figure 12. User reviews classifications - the relationship between GAIA and BBC guidelines

reviews can be used precisely in the evolution and improvement of applications. Understanding ways to improve the accessibility of an application is a fundamental task to promote its improvement. In this way, it is possible to ensure that the app is accessible and valuable for all users, regardless of their specific limitations or conditions.

The classification step of user reviews indicated that seven out of eight of the applications had accessibility problems according to their users' reviews. Accessibility issues were found for all categories of the GAIA guidelines, but not for all the categories of BBC. It may indicate that the applications follow general accessibility guidelines related to images, links, and structure, but do not follow specific accessibility guidelines for autistic users' needs.

During the textual analysis of user reviews, two applications stood out for concentrating the highest number of accessibility reports made by their users. They were: ABC Autismo and Matraquinha. The fact that they are more popular applications may have influenced this result, as they have more user reviews.

The adoption of BBC guidelines has proven to be useful in comprehensively analyzing accessibility, as GAIA is not specific for mobile applications. However, we did not identify a user review related to BBC and not related to GAIA. We could classify all reviews about accessibility into some GAIA categories. By analyzing the relationship between GAIA and BBC guidelines, we noted that problems related to design, audio and video, and focus (according to BBC) are commonly also related to customization (according to GAIA). At the same time, problems related to notifications (according to BBC) are also related to responses to action and system state visibility (according to GAIA). It can support future work to automatically classify user reviews.

In future work, we propose to optimize the process of extracting and analyzing user reviews using an automatic tool. The UUX-Post tool is an excellent alternative. It is a tool that has already been used in other studies on user reviews and can add to the continuity of the work. The use of keywords is a methodology with potential for improvement. Morphological variations and synonyms can enrich the set of words and make automatic classification more effective. Considering other applications and more current reviews can enable the discovery of new results, enriching the current results.

References

Abdul Aziz, N. S., Ahmad, W. F. W., and Zulkifli, N. J. b. (2015). User experience on numerical application between children with down syndrome and autism. In *Proceedings of the International HCI and UX Conference in Indonesia*, pages 26–31.

Acosta-Vargas, P., Salvador-Acosta, B., Salvador-Ullauri, L., Villegas-Ch., W., and Gonzalez, M. (2021). Accessibility in native mobile applications for users with disabilities: A scoping review. *Applied Sciences*, 11(12).

Aljedaani, W., Mkaouer, M. W., Ludi, S., and Javed, Y. (2022). Automatic classification of accessibility user reviews in android apps. In *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*, pages 133–138.

Alqahtani, F. and Orji, R. (2019). Usability issues in mental health applications. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, pages 343–348.

Association, A. P. et al. (2014). *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. Artmed Editora.

Branco, K. d. S. C., Pinheiro, V. M. d. S., Damian, A. L., and Marques, A. B. (2020). Investigating the first user experience and accessibility of educational applications for autistic children. In *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*, pages 1–10.

Britto, T. C. P. and Pizzolato, E. B. (2016). Gaia: uma proposta de um guia de recomendações de acessibilidade de interfaces web com foco em aspectos do autismo. *Revista Brasileira de Informática na Educação*.

Da Silva, T. H. O., Freitas, L. M., Mendes, M. S., and Furtado, E. S. (2019). Textual evaluation vs. user testing: a comparative analysis. In *Anais do X Workshop sobre Aspectos da Interação Humano-Computador para a Web Social*, pages 21–30. SBC.

Diniz, L. d. N., de Souza Filho, J. C., and Carvalho, R. M. (2022). Can user reviews indicate usability heuristic issues? In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA. Association for Computing Machinery.

Eler, M. M., Orlandin, L., and Oliveira, A. D. A. (2019). Do android app users care about accessibility? an analysis of user reviews on the google play store. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, pages 1–11.

Freitas, L. M., da Silva, T. H. O., and Mendes, M. S. (2016). Evaluation of spotify: an evaluation textual experience using the maltu methodology. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, pages 1–4.

Genc-Nayebi, N. and Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, 125:207–219.

Koyani, S. J., Bailey, R. W., Nall, J. R., Allison, S., Mulligan, C., Bailey, K., and Tolson, M. (2004). *Research-based web design & usability guidelines*. Computer Psychology Washington, DC.

Magaton, H. C. and Bim, S. A. (2017). The use of educational applications by children with autistic spectrum disorder: A case of study. In *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, pages 1–10.

Marques, A. B. and da Silva Monte, L. (2021). How are software technologies being evaluated with autistic users? a systematic mapping. *Universal Access in the Information Society*, pages 1–11.

Marques, A. B., Melo, G. A., Pereira, V. G. R., and de Vasconcelos, P. F. C. (2021). Criação e avaliação de personas para sistemas com foco em pessoas com transtorno do espectro autista: Um relato de experiência sobre o uso de netnografia. *iSys-Brazilian Journal of Information Systems*, 14(3):68–97.

Melo, A., Santos, J., Rivero, L., and Barreto, R. (2017). Searching for preferences of autistic children to support the design of user interfaces. In *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, pages 1–10.

Santiago, M. T. and Marques, A. B. (2022). Are user reviews useful for identifying accessibility issues that autistic users face? an exploratory study. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems*, IHC '22, New York, NY, USA. Association for Computing Machinery.

Vu, P. M., Nguyen, T. T., Pham, H. V., and Nguyen, T. T. (2015). Mining user opinions in mobile app reviews: A keyword-based approach. *arXiv preprint arXiv:1505.04657*.

Wang, C., Liu, T., Liang, P., Daneva, M., and van Sinderen, M. (2021). The role of user reviews in app updates: A preliminary investigation on app release notes. In *2021 28th Asia-Pacific Software Engineering Conference (APSEC)*, pages 520–525.