



Improving Usability of a Mobile Application for Children with Autism Spectrum Disorder Using Heuristic Evaluation

Murilo C. Camargo¹✉, Tathia C. P. Carvalho², Rodolfo M. Barros¹,
Vanessa T. O. Barros¹, and Matheus Santana¹

¹ State University of Londrina, Londrina, PR 86057 970, Brazil
murilocrivellaric@gmail.com,
rodolfomdebarros@gmail.com, vanessa@uel.br,
ss.matheus.94@gmail.com

² São Paulo State University, Bauru, SP 17033 360, Brazil
tathiacarvalho@gmail.com

Abstract. Autism Spectrum Disorder (ASD) is a complex clinical condition that includes social, behavioral, and communication deficits. As numbers in ASD prevalence rise significantly, the tools for computer-assisted interventions also increase proportionally, which can be confirmed by the growth in the literature body addressing the issue. The development of autism-specific software is far from being straightforward: it often requires a user-centered approach, with a cross-functional team, and a primary focus on usability and accessibility. One of the most popular methods for finding usability problems is the heuristic evaluation, which is performed by having a group of experts testing the User Interface and providing feedback based on predetermined acceptance criteria. Thus, this paper informs on the assessment of a mobile application for autistic individuals using the heuristic evaluation. The software subjected to evaluation – prototyped in a previous study – addresses organization and behavioral patterns in ASD children. Through the heuristic evaluation, improvements could be performed in the application. Also, lessons learned with the evaluation process include recommendations to help the selection of methods and materials, the conduction of the evaluation, and the definition of the follow-up strategy. By describing the method stepwise and sharing lessons learned, the aim is to provide knowledgeable insights for development teams handling autism-specific software.

Keywords: Heuristic evaluation · Usability · Autism Spectrum Disorder · Mobile application

1 Introduction

Autism Spectrum Disorder (ASD) is a cognitive disturbance that refers to a variety of severity degrees and symptoms – including social, behavioral, and communication deficits [1]. The fifth version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), provided by the American Psychiatric Association [2], expanded

the comprehensiveness on the disturbance, in an attempt to add flexibility to the diagnosis criteria and support health care practices. Especially in the last decade, ASD numbers rose significantly. For instance, in the United States, according to the Autism and Developmental Disabilities Monitoring (ADDM) network, about 1 in 68 children aged 8 has been diagnosed with ASD [3].

Studies report that ASD individuals are especially interested in visual media, due to their frequently high visual processing skills [4, 5]. Based on this preference, many studies have been emerging on the development of computer-assisted interventions for ASD individuals, including communication and social interaction support [6, 7], and applying multiple technologies such as Virtual Reality [8] or Kinect [9]. Computer-assisted interventions provide users with a controlled environment, which can be personalized according to their needs, in order to foster independence and motivation. Though the use of computers represents a novel strategy to improve the welfare of people with ASD, the development of accessible technology is complex: it requires a user-centered approach – preferably with a cross-functional team – and a primary focus on usability, accessibility, and the adaptation of internal procedures and policies [10].

In a previous study, we described the first stages in designing a mobile application for ASD children – named DayByDay [11]. Now, considering the importance of usability throughout the entire development process, we intend to go one step further and improve the usability of our application through a heuristic evaluation. The heuristic evaluation is a well-known method for finding usability problems by having a small group of experts testing the interface and describing design issues based on certain rules (the heuristics). The method is sometimes criticized because it focuses on finding problems rather than evaluating the user performance. Of course, it does not replace user testing. However, this approach is particularly interesting in intermediate development stages: it encourages evaluators to describe specific usability problems and connect them with objective assessment criteria, thus guiding the development team towards a high-quality software product [12].

Multiple sets of heuristics are available in the recent literature [12–15]. In general, they derive from the original set of usability heuristics proposed by Jakob Nielsen in 1994. In this study, we use the set of heuristics described by Islam and Bouwman [15]. The authors propose a semiotic framework for interface evaluation consisting of five major areas: syntactic, pragmatic, social, environmental, and semantic aspects. Those major areas are divided in 16 topics of interest. We believe this approach enables evaluators to cover the entire interface in a comprehensive manner, and at the same time provides an easier vocabulary, which avoids interpretation errors. The full description of the usability heuristics is presented within this paper in a section exclusively dedicated to the method.

This work is structured as follows: Sect. 2 provides a brief context on the current status of the mobile application (DayByDay); Sect. 3 describes the evaluation method applied in terms of the usability heuristics set, planning, participants, materials, and procedures; Sect. 4 presents the results of the evaluation process; Sect. 5 presents the improvements performed on the application, and a comparison between the former version of the interface and how it changed after the heuristic evaluation; Sect. 6 discusses the lessons learned with the process; and Sect. 7 provides conclusions and future work directions. By presenting the heuristic evaluation process and discussing

lessons learned targeted specifically to the ASD context, this paper can positively assist development teams handling accessibility.

2 Previous Work: DayByDay

The first stage in the development process of our mobile application was described in a previous study [11]. The application – named DayByDay – aims at helping ASD children aged 8 to 12 organize their routine and perform daily activities, such as “have breakfast”, “brush the teeth”, or “make your bed”. Children must select an activity from a list displayed on the screen and follow the steps in order to complete it. For instance, the activity “have breakfast” may include the steps “pick a bowl”, “add cereal”, “pour some milk”, and “eat the meal”. By completing activities, children earn points (represented by stars, in this case). After collecting a number of stars, they are rewarded with a real-life token of their choice (a game, a leisure activity, some candy, etc.).

Parents, educators, and therapists are also encouraged to participate. In the “care-giver management system” within the app, parents and educators are asked to tailor content-specific features, such as the activities to be performed, the sub-steps in each activity, and the list of rewards available to the ASD children. In the first version of the application, there are options to upload images, use the built-in camera to take pictures, or record voice messages. In addition, they can monitor the children’s progress through push notifications and a control center. One of the main concerns of the earliest version is to provide feedback to the ASD users. This is achieved by two different strategies: one, by earning stars each time they successfully complete an activity (visual feedback); and two, by receiving messages (either written or recorded) and real-life rewards from their caregivers after their performance.

Personalization is also a topic of major importance. Autism deals with a variety of severity degrees, which means there are considerable differences in interests and skills. In this context, parents and educators are entitled to select the most suitable output channels from the multimedia options provided – upload images, take pictures, record voice messages, write memos, organize the scheduled activities, or a combination of those. Besides that, the reward system intends to foster motivation and reinforce desired behavioral patterns when the ASD children successfully complete an activity. This first version of the app was prototyped and presented to parents of ASD children in a focus group session. In addition to that, a questionnaire was also applied – results show a positive acceptance overall, with only a few minor improvement suggestions. A comprehensive explanation on the first version is available in [11].

Figure 1 shows some of the screens designed for the first version. The screen to the left is the list of activities presented to the child. Activities marked with a “check” are completed, whilst the ones with a “padlock” are not available at the moment. In the center, there is an example of the activity “change clothes”. In this case, pictures of the pieces of clothes were uploaded to help the child. Finally, the screen to the right is showed after the completion of an activity. It includes the stars earned so far, a message from the parents, and action buttons. The cloudy background was designed to add playfulness and draw the children’s attention. Also, the background color would

change according to the time of the day: light blue for mornings, sunset orange for afternoons, and a purple sky at night.

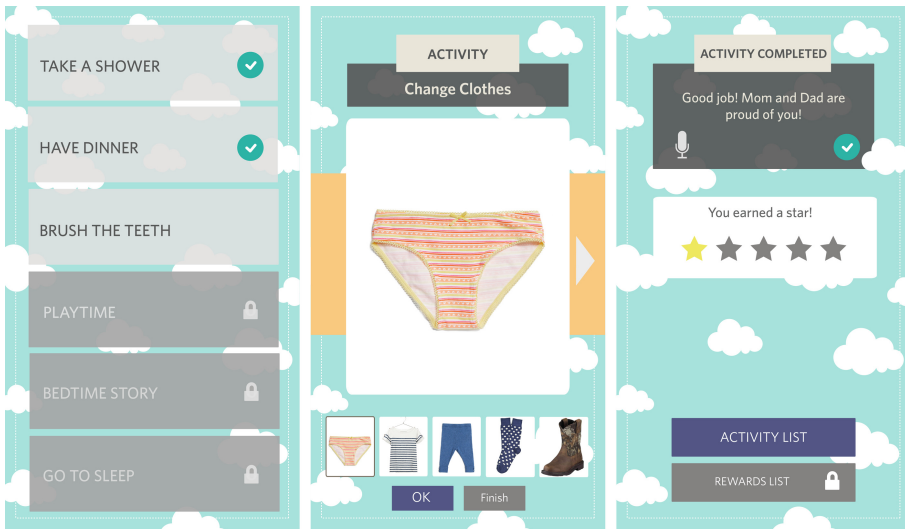


Fig. 1. DayByDay screens, first version. List of activities (left), activity steps (center), and activity completed screen (right).

3 Method

Even though questionnaires and focus groups are important sources of feedback, these approaches have significant limitations. On the one hand, respondents tend to answer favorably to the object of interest (in this case, the interface) when they are confronted face to face. On the other hand, in case interviewees are not able to express their knowledge in words, what they say is not necessarily what they mean [16]. In a heuristic evaluation, evaluators tend to be more comfortable to give their opinions and are more likely to provide “negative” feedback, that is, list usability problems. This method also makes the evaluation process easier, because evaluators are not expected to provide complex explanations, but rather connect the usability problems they find with a standardized list of principles. Furthermore, the first part of the assessment (described in [11]) was conducted only with parents of ASD children. Involving experts and combining multiple evaluation techniques is a good approach to have a multidisciplinary overview from different fields of knowledge [10].

Generally, there are two evaluation methods applying usability heuristics: user testing (testing involves end-users), and usability inspection (also called “expert evaluation”, does not involve end-users) [12]. In this study, we conduct a usability inspection because we want to improve usability and prevent errors prior to having end-users testing the interface. The usability inspection involves a small set of

evaluators, who assess the interface and judge its compliance with recognized usability principles, namely the heuristics.

3.1 Usability Heuristics Set

Heuristics are general rules that describe common properties of usable interfaces. This study uses the Semiotic Interface sign Design and Evaluation (SIDE) framework, proposed by [15]. SIDE consists of five major areas (syntactic, pragmatic, social, environmental, and semantic aspects), divided in 16 minor topics. The framework properties are summarized below:

Syntactic Evaluation. This area involves the evaluation of primary visual/graphic aspects, namely: interactivity, color, clarity and readability, presentation, context, and consistency.

1. Interactivity: Related to the level of interactivity in the interface and how it connects to the user previous knowledge in digital environments;
2. Color: The colors used in the interface, how light the colors are, and whether there is enough color contrast between elements;
3. Clarity and readability: Elements in the interface must be clear, concise, and easy to understand. Text must be informational and easy to read;
4. Presentation: Related to the overall appearance of the interface and the structure of the elements (layout, arrangement, font size, etc.);
5. Context: Measures the level of accuracy in the interpretation of the web domain and/or the application name, and how much sense users make of it;
6. Consistency: Related to the design patterns used throughout the interface that can help users in the interpretation process.

Pragmatic Evaluation. Also related to visual/graphic interface aspects, this area evaluates how the interface elements influence the user perception of it. Topics of interest include the position, amplification, relations, and coherence.

1. Position: Elements must be placed in common positions, with respect to the user habits in digital environments, in order to favor the user localization;
2. Amplification: This attribute is concerned with elements that can be amplified – such as icons, thumbnails, small images, short text – and whether users understand their “abbreviated” meaning;
3. Relations: Relationships between elements (action buttons, hyperlinks, text/image) must allow a “cause-and-effect” interpretation;
4. Coherence: Digital elements must match real-world conventions, so that users can infer relationships in a logical manner.

Social Evaluation. This area assesses the digital environment in relation to the social context and how well the interface is organized in terms of cultural representations. It includes: cultural marker, matching, organization, and mapping.

1. Cultural marker: Refers to the use of elements (colors, language, iconic representations, etc.) within a specific cultural context;

2. **Matching:** Proper use of digital metaphors to express reality, conventions, and real-world objects;
3. **Organization:** Related to how the content is organized in categories within the interface and whether that organization favors an easy and safe navigation;
4. **Mapping:** In case the interface deals with complex concepts or activities, it must apply a “generalization” (simple-to-complex) approach.

Environmental Evaluation. The environmental part of the evaluation is concerned with the ontologies used to universalize the interface.

1. **Ontology:** The interface must include universalization traits, so that users are able to interpret the referential meaning of objects.

Semantic Evaluation. Finally, the semantic evaluation includes the interpretation accuracy, which is measured in five levels.

1. **Interpretation accuracy:** Measures the accuracy of the user interpretation in relation to the interface design. The five possible levels are – accurate, moderate, conflicting, erroneous, or incapable.

3.2 Participants and Materials

The application was prototyped using Marvel [17]. It is an easy-to-use online software that allows static screens to become fairly interactive without codification. Of course, Marvel does not integrate back-end components or server/database solutions. However, it does provide a full front-end experience with all functional and interactive capabilities, thus allowing the evaluator to inspect the application in a smartphone. Working with a mobile prototype enables evaluators to perform a more accurate inspection whereas the end-user will be running the application in mobile devices.

In order to gather feedback from different fields of knowledge, DayByDay evaluation was performed with a cross-functional team, involving design, special education, and computer science experts. Six evaluators (E) formed the evaluation team: (E1) a computer science professional specialized in mobile development, (E2) a project manager dealing with accessible software development, (E3) a user experience designer, (E4) a graphic designer with experience in accessible applications, (E5) a teacher working with accessibility in Basic Education, and (E6) a university professor who researches accessibility for autism.

3.3 Procedures

During the evaluation session, the evaluator is expected to go through the interface several times, inspecting and comparing the elements with a list of recognized usability principles (the heuristics). Prior to the beginning of the evaluation, a hard copy of the sixteen heuristics was provided for each evaluator. Then, the researcher presented directions for the inspection. In this heuristic evaluation, participants are asked to perform activities in two given scenarios.

In the first scenario, evaluators should complete an activity designed for the ASD audience, namely “have breakfast”. In order to start, the evaluator should choose the activity. It consists of four steps: “pick a bowl”, “add cereal”, “pour milk”, and “eat the meal”. Participants should progress through all the steps in order to complete the activity. In the second scenario, evaluators pretend they are caregivers of ASD children, and are requested to create and upload a new activity to be displayed in the child’s application. This task is more complex, since it involves more options and functionalities. First, evaluators must sign in on the application. This action requires filling a form with personal information and choosing between default personalization options. Then, they are asked to create a new activity, which consists of four steps: select a date to insert the activity; write the name of the activity; upload, from the phone gallery, pictures to be used in the step-by-step; and confirm the actions.

In both scenarios, despite the fact that evaluators had specific tasks to complete, they were also encouraged to explore other sections within the interface. They could, for instance, access the “settings” and navigate through the various personalization options available. It was suggested that evaluators performed the tasks in both scenarios at least twice. In the first time, they would have an overview of the application, learn the mechanics and try to complete the task. In the second time, they should inspect every component individually, and provide detailed feedback according to the list of heuristics. However, evaluators could play with the interface as many times as they wished. There was no time limit either. They are allowed to test the prototype freely as long as they still have reasons to do so.

Prior to the beginning of the session, the researcher provided a brief explanation on the application and about the tasks the evaluators were expected to accomplish in each scenario. After that, the evaluation itself proceeded without further assistance during the evaluator’s performance. Evaluators inspected the interface individually – that is, one at a time in the evaluation room – and listed the usability problems they found based on the heuristics set. Furthermore, each participant was asked to rate the problems in a scale from 1 to 5, as described below:

- Rate 1: Not a usability problem;
- Rate 2: Cosmetic problem. Should only be fixed if there is extra time available;
- Rate 3: Minor usability problem with low priority to fixing;
- Rate 4: Major usability problem with high priority and important to fix;
- Rate 5: Usability catastrophe that needs urgent fixing.

Feedback collected with each evaluation session was compiled in individual evaluation reports, which were then analyzed by the researchers. Together, they provide meaningful insights on usability problems that may assist the development team in fixing bugs and preventing errors. Results of the evaluation are presented in the next section.

4 Results

By choosing experts from various fields of knowledge, the heuristic evaluation collected high-quality feedback on multiple aspects of the interface, such as graphic elements, implementation issues, non-functional attributes, teaching-learning adaptation, and so on. After the evaluation sessions, results were compiled in evaluation reports and analyzed. Tables 1, 2, 3 and 4 summarize the findings from the heuristic evaluation. In addition to analyzing each evaluator's results individually (E1, E2, E3, E4, E5, E6), we also consider the overall number of problems (simply indicated as E).

Table 1. Number of usability problems divided by evaluator.

Numbers refer to:	E	E1	E2	E3	E4	E5	E6
Child's interface	8	2	0	3	0	3	0
Caregiver system	49	11	7	9	12	5	5
Overall interface	57	13	7	12	12	8	5

Table 1 presents the number of usability problems found by all six evaluators. A total of 57 problems were detected. The caregiver system (used to create and manage activities) had the majority of detected issues (84.2%), whilst the child's interface (used by the ASD children to perform activities) had only 8 problems. E2, E4 and E6 did not find problems in the child's interface. Overall, the problems are fairly distributed among all the evaluators, but are more prominent in the caregiver system, suggesting that this portion of the interface needs priority.

Problems were ranked according to their severity in a scale from 1 to 5. The ranking is demonstrated in Table 2. Most problems were ranked 4 (24.5%) and 1 (22.8%), but ranks 2, 3 and 5 also appeared significantly. Evaluators E1, E3, E4, and E6 indicated most problems were medium to high priority (ranks 3, 4, and 5), in contrast with the feedback received from evaluators E2 and E5, who ranked problems mainly as 1 or 2 (lowest to low priority).

Finally, Tables 3 and 4 present the usability problems ranked and divided in the two parts of the interface. Table 3 summarizes the data related to the child's interface, while Table 4 presents the ranking for the caregiver system. E1 reported to have found only minor problems in the child's interface, but considerably severe problems in the caregiver system (ranks 4 and 5). E2 did not find any problems in the child's interface, and neither did E4 and E6. A few problems were reported in the child's interface by three evaluators (E1, E3, and E5). In the caregiver system, in turn, the number of problems was higher and they were considered more critical (ranks 3, 4, and 5). If we take into account the area of expertise, the majority of problems are related to implementation issues (25 problems – 43.8% – reported by science computer professionals E1 and E4). In second place, reports suggest problems in the educational domain (17 problems, or 29.8%, reported by E3 and E6). Finally, the least amount of problems is from the interface design (15 problems, or 26.3%, reported by E2 and E5).

Table 2. Number of usability problems divided by severity level.

Severity level	E	E1	E2	E3	E4	E5	E6
1 (lowest)	13	2	3	1	2	4	1
2 (low)	10	2	2	1	2	2	1
3 (average)	11	2	0	3	5	1	0
4 (high)	14	4	1	4	2	1	2
5 (highest)	9	3	1	3	1	0	1

Table 3. Usability problems (child's interface).

Severity level	E1	E2	E3	E4	E5	E6
1 (lowest)	1	0	0	0	2	0
2 (low)	0	0	1	0	1	0
3 (average)	0	0	2	0	0	0
4 (high)	1	0	0	0	0	0
5 (highest)	0	0	0	0	0	0

Table 4. Usability problems (caregiver system).

Severity level	E1	E2	E3	E4	E5	E6
1 (lowest)	1	3	1	2	2	1
2 (low)	2	2	0	2	1	1
3 (average)	2	0	1	5	1	0
4 (high)	3	1	4	2	1	2
5 (highest)	3	1	3	1	0	1

In order to be fixed, problems should be prioritized accordingly. Table 5 lists the main problems – divided by severity levels – reported by the evaluators. According to the scale used in this study, rank 1 is the lowest priority level and 5 is the highest. In some cases, the same problem was reported by more than one participant. Furthermore, very similar problems were found in several screens (for instance, a button that does not follow design patterns and appears multiple times throughout the interface). For this reason, similar problems appear only once. In addition, each problem is related to the corresponding principle from the heuristics set.

5 Application Improvements

After the heuristic evaluation, developers analyzed carefully the feedback collected and planned improvements in the application. Unfortunately, some of the expert recommendations could not be performed due to technology and/or human resources limitations. However, the development team did try other solutions when experts' suggestions were not possible to implement. One of the major problems detected

Table 5. Usability problems description.

Severity level	Problem description	Problem refers to:
1 (lowest)	“The interface should include a video tutorial”	Interactivity
	Add options to sign in via Facebook or Google	Interactivity
2 (low)	Some of the icons used in the interface may be confusing	Matching
	“The sign in form is too long and I do not know whether it refers to the caregiver or the child”	Organization
	“The interface could include a gallery with default pictures”	Interactivity
	Font size is too small	Clarity and readability
3 (average)	Review buttons “save” and “share” position	Position
	Some buttons do not follow web conventions	Consistency
	Add filters and organization options to the search tool	Organization
	Replace the button “age” for “date of birth”	Interpretation accuracy
4 (high)	“The colors used in the interface may not be the best to fit every user’s preferences”	Color
	“There is too much text information on the interface. Many ASD children cannot read”	Clarity and readability
	Section titles are too long	Clarity and readability
	Include labels to all input fields	Presentation
	Signing in is a long and exhausting task	Mapping
	It is difficult to find the buttons to go to the next step	Organization
5 (highest)	“The rewards system must be revised. It takes too long for the child to get the reward, and they may not be able to associate it with success”	Relations
	Remove “delete” option for images. It may cause an accidental exclusion	Presentation
	“No fill” should not be a color option	Color
	Headings and sub-headings need to be clearer. It is difficult to understand what they imply	Organization

during the heuristic evaluation is the use of text. In fact, autism covers a broad spectrum of symptoms, and ASD individuals are often not able to read. This is why a picture-oriented approach is more effective (Fig. 2).

In the first version of the application, the list of activities was presented mainly with the help of text. In the improved version, however, pictures replaced text, also including icons to help the visualization and the understanding. The interface is as simple as possible, to avoid overexposure to unnecessary information. Pictures were used to replace text throughout the application, keeping written information to a bare minimum. In the activity steps, for instance, images take over almost the entire screen

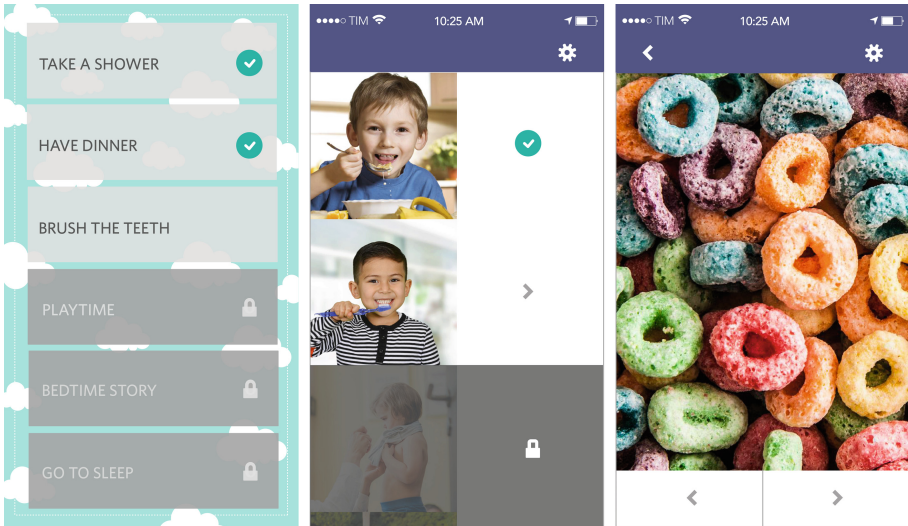


Fig. 2. DayByDay screens. List of activities, first version (left); list of activities, improved version (center); activity step, improved version (right).

(Fig. 2, picture to the right). This helps ASD children focus on the activity and decreases the need for assistance at the same time.

The colors used in the interface also needed a revision. The first version of the interface used vibrant colors by default, as well as a cloudy background, with hopes to make the interface more attractive. However, experts mentioned that this overload of visual information could potentially be unpleasant to ASD individuals. They might be attracted to vibrant colors, or they could feel anxious or irritating about it. In the second version of the application, the development team tried to overcome this problem by providing a color selection option. In this new functionality, caregivers (or parents, or teachers) choose the colors that best suit the child's preferences (Fig. 3).

In relation to the rewards, experts pointed out that it might not be as effective if they are handed long after the completion of an activity. In our first version, children collected stars as they completed tasks, and only after completing all the activities they would be rewarded. This hinders the association between a successful performance and the pleasure of receiving a reward. In order to solve this problem, the improved version brings an instant rewarding system. Besides a visual feedback displayed on the screen, the child chooses a reward right after the activity. We encourage parents to have at hand some real-life options, but we also provide some digital rewards (mini-games, stickers, etc.). In the new version, rewards are simpler, but more frequent and immediate.

Finally, the long form required to sign in – which, in the first version, included extensive personal information and was split into two screens – was synthesized in only two text input fields in the improved version. This time, users are simply asked to type their e-mail account and choose a password (Fig. 4).

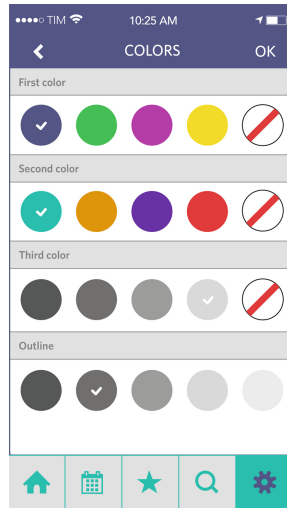


Fig. 3. DayByDay screen, improved version. Color selection.

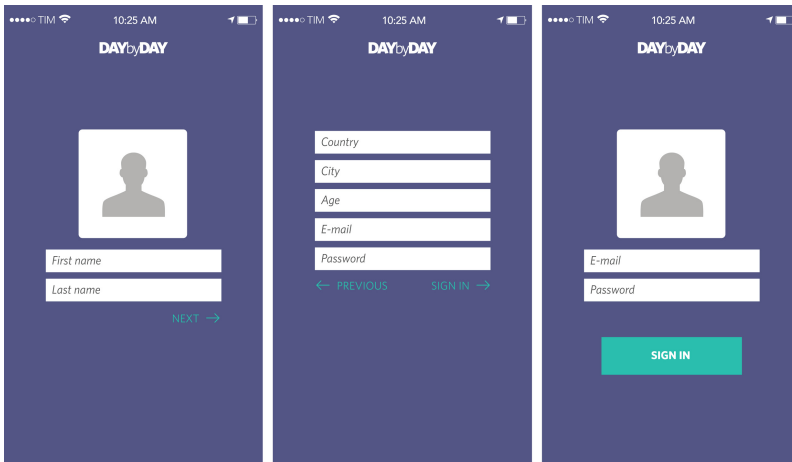


Fig. 4. DayByDay screens. Sign in form, first version (left); sign in form (continued), first version (center); sign in form, improved version (right).

In addition to all the improvements described so far, the development team also performed the adjustments ranked with lower priority levels, including font size, icons design, headings and sub-headings description, layout and elements position arrangements, action buttons patterns, among others. By the end of the development, all of the experts' recommendations had been successfully accomplished, even when the team had to explore alternative solutions due to project limitations.

6 Lessons Learned

Results achieved with the heuristic evaluation show prominent improvements in the User Interface, especially in terms of layout arrangement (elements position, design patterns), visual presentation strategies (use of colors, textual information), and customization options (rewards, color scheme, information management). Although the heuristic evaluation was undeniably helpful in providing high-quality feedback for the development team, dealing with autism added complexity to the process. This is why lessons learned are important: they share the knowledge and experience acquired throughout the process, and might support the development team in future decision-making.

In this study, lessons learned are divided in three main topics: (1) selection of the method and materials, (2) conduction of the evaluation process, and (3) definition of the follow-up strategy. First of all, regarding the selection of the method and materials, it is assumed that developers choose the usability inspection, because this is the object under appreciation in this study. On the one hand, it is necessary to think about the team composition and, on the other, the materials used to perform the evaluation. During this study, it became clear the importance of having a cross-functional team inspecting the interface. Experts from various fields of knowledge bring together their own background and unique points of view that provide high-quality feedback on usability aspects. Thus, developers must consider including experts that make sense to the project scope (in this case, computer science, design, and education).

Furthermore, the materials selected must provide an evaluation scenario that is as close to reality as possible. The technology used in this study (the Marvel prototyping software) allowed experts to test the interface in a mobile device and with most of the functionalities and options available, which increases the evaluation accuracy. However, it does not support cross-platform interaction. This requires a new round of evaluation, since testing the interface in different operational systems could potentially result in different usability problems. Also, the heuristics set must be selected according to the development context and the nature of the software. Many heuristics are available both in literature and across the Internet, so choosing the right one is a matter of research.

In addition to that, clear instructions must be provided for the evaluators. Of course, researchers do not intend to influence the experts' judgment, but they do need to feel confident about the interface and know what is expected from their evaluation. In this sense, a focus group could have been performed prior to the beginning of the evaluation sessions. It allows team members to connect with one another in the first place. Furthermore, researchers could have had the opportunity to provide more detailed explanation on the application (learning objectives, autism specifications, technology limitations, etc.). Development experts, for instance, may have a lot of experience with programming languages and usability issues, but we cannot assume they understand the ASD specific clinical conditions. After the focus group (or another similar method) the evaluation session should proceed individually, so that experts are not influenced by external judgments. Once the evaluation is completed, data is extracted, categorized, and analyzed, which leads to the definition of the follow-up strategy.

The quantitative analysis relating the number of problems to a specific part of the software (as in Tables 1, 2, 3 and 4) allows a systemic view of the usability problems and the identification of deficits in a particular area. For instance, most of the problems were detected in the caregiver system, suggesting that this part of the interface needed more attention. Also, many problems were reported by the computer science and education experts, which suggests that, in terms of design, the software reached a positive acceptance overall. In turn, the qualitative analysis (Table 5) provides a detailed description of each problem individually. In addition, connecting problems with the heuristic principles and ranking them according to a severity scale helps the prioritization process and supports further improvements. After all, the development team is responsible for deciding what is possible based on the financial, technological, and human resources available.

7 Conclusions and Future Work

In order to detect and perform improvements in the interface using heuristic evaluation, this study reported on the evaluation (planning, conduction, and results) of DayByDay, an ongoing mobile application for children with Autism Spectrum Disorder (ASD). With the participation of a cross-functional team of experts, it is concluded that this method, when applied consistently and according to the development context peculiarities, may provide accurate data and insights on usability issues. Feedback received from the experts helped the development team improve the former version of the software. Furthermore, lessons learned during the process are important to raise awareness on the results achieved and future challenges to overcome.

The development of accessible software is complex and requires an iterative process, so that the development team can accomplish the best possible results prior to releasing the software. Therefore, results in this paper are not the final milestone, but rather represent the second phase of an ongoing development process. This means that future upgrades and improvements needs to be performed in the application, which will then be subjected to testing procedures involving end-users. Above all, the sharing of experiences in autism – a complex topic still emerging in literature – helps tailor best practices within the community of developers and shape the way we see autism accessibility in the future.

References

1. Wojciechowski, A., Al-Musawi, R.: Assistive technology application for enhancing social and language skills of young children with autism. *Multimedia Tools Appl.* **76**, 5419–5439 (2017)
2. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders, DSM-5*. Artmed, 5th edn (2013)
3. Wingate, M., Kirby, R.S., Pettygrove, S., Schulz, E., Gosh, T.: Prevalence of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report*, Center for Disease Control and Prevention, pp. 1–21 (2014)

4. Kamaruzaman, M.F., Ranic, N.M., Norb, H.M., Azaharia, M.H.H.: Developing user interface design application for children with autism. *Soc. Behav. Sci.* **217**(1), 887–894 (2016)
5. Mazurek, M.O., Shattuck, P.T., Wagner, M., Cooper, B.P.: Prevalence and correlates of screen-based media use among young youths with autism spectrum disorders. *J. Autism Dev. Disord.* **42**(1), 1757–1767 (2012)
6. Alvarado, C., Munoz, R., Villarroel, R., Acuña, O., Barcelos, T.S., Becerra, C.: ValpoDijo: developing a software that supports the teaching of Chilean idioms to children with autism spectrum disorders. In: 12th Latin American Conference on Learning Technologies (LACLO) (2017)
7. Zhang, L., Fu, Q., Swanson, A., Weitlauf, A., Warren, Z., Sarkar, N.: Design and evaluation of a collaborative virtual environment (CoMove) for autism spectrum disorder intervention. *Trans. Access. Comput.* **11**(2), 11 (2018)
8. Winoto, P., Xu, C.N., Zhu, A.A.: “Look to remove”: a virtual reality application on word learning for Chinese children with autism. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2016. LNCS, vol. 9739, pp. 257–264. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40238-3_25
9. Cai, Y., Chia, N.K.H., Thalmann, D., Kee, N.K.N., Zheng, J., Thalmann, N.M.: Design and development of a virtual dolphinarium for children with autism. *Trans. Neural Syst. Rehabil. Eng.* **21**(2), 208–217 (2013)
10. Panchanathan, S., McDaniel, T., Balasubramanian, V.N.: An interdisciplinary approach to the design, development and deployment of person-centered accessible technologies. In: 2013 International Conference on Recent Trends in Information Technology (ICRTIT), pp. 750–757. IEEE (2013)
11. de Oliveira Barros, V.T., de Almeida Zerbetto, C.A., Meserlian, K.T., Barros, R., Crivellari Camargo, M., Cristina Passos de Carvalho, T.: DayByDay: interactive and customizable use of mobile technology in the cognitive development process of Children with autistic spectrum disorder. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014. LNCS, vol. 8514, pp. 443–453. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07440-5_41
12. Pribeanu, C.: A revised set of usability heuristics for the evaluation of interactive systems. *Informatica Economică* **21**(3), 31–38 (2017)
13. Quiñones, D., Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C.A.: Developing usability heuristics: a formal or informal process? *IEEE Latin Am. Trans.* **14**(7), 3400–3409 (2016)
14. Sim, G., Read, J.C.: Using computer-assisted assessment heuristics for usability evaluations. *Br. J. Edu. Technol.* **47**(4), 694–709 (2016)
15. Islam, M.N., Bouwman, H.: Towards user-intuitive web interface sign design and evaluation: a semiotic framework. *Int. J. Hum. Comput. Stud.* **86**(1), 121–137 (2016)
16. Ardito, C., Buono, P., Caivano, D., Costabile, M.F., Lanzilotti, R.: Investigating and promoting UX practice in industry: an experimental study. *Int. J. Hum. Comput. Stud.* **72**(6), 542–551 (2014)
17. Marvel Homepage, Free Mobile & Web Prototyping. <https://marvelapp.com>. Accessed 15 Jan 2019