Exploring User Experience and Usability of mHealth applications for people with diabetes: An Evaluation Study Using UEQ and HE4EH Checklist

Maria Elanne M. Rodrigues [ Federal University of Ceará | elannemendes@alu.ufc.br ]
Karla Haryanna S. Moura [ Federal University of Ceará | karlaharyanna@alu.ufc.br ]
Karina Castelo Branco [ Federal University of Ceará | karinascb@alu.ufc.br ]
Valéria Lelli [ Federal University of Ceará | valerialelli@ufc.br ]
Windsor Viana [ Federal University of Ceará | windson@virtual.ufc.br ]
Rossana M. C. Andrade [ Federal University of Ceará | rossana@ufc.br ]
Ismayle S. Santos [ State University of Ceará | ismayle.santos@uece.br ]

Abstract Diabetes affects 10% of adults aged 20 to 79 globally, caused by insufficient insulin production. Mobile apps are effective in managing this chronic condition. This study aimed to investigate the experience of using mobile applications for healthcare patients with diabetes in their daily activities. Ten participants diagnosed with diabetes were invited to evaluate five Android mobile applications using a predefined protocol. We used the User Experience Questionnaire (UEQ) to assess the usability and user experience of the apps. In addition, the Heuristic Evaluations for mHealth (HE4EH) checklist was used by three domain experts to assess the apps. The results showed that one of the five apps had issues with the stimulus and novelty aspects of user experience. Among all the apps, the novelty had the lowest Likert scale value. While most apps had creative designs and attractive interfaces, the heuristic assessment revealed many violations and concerns about instruction and suitability due to a lack of crucial information for monitoring user routines.

Keywords: Healthcare mobile applications, User experience (UX), Diabetes

1 Introduction

Mobile applications have collaborated to recognize relevant users’ information through monitoring data, such as usage logs, and users’ context and daily routines (De Nardin et al., 2020). In this perspective, several studies have proposed using mobile apps to prevent, diagnose, and treat diseases (Rodriguez and Wägner, 2019). Mobile apps could also play a crucial role in educating about a particular ailment (e.g., dietary education in diabetes prevention), in capturing physical data from the user that aids prevention (e.g., heart rate and blood glucose), and in helping to maintain a routine for treating a disease (e.g., medication reminder alarms). However, many apps are developed without following a quality or safety recommendation to allow their reliable use in a health context (Molina-Recio et al., 2020). Hence, user experience assessments are conducted to evaluate the quality of these applications and detect problems that may affect their use, such as interaction and interface issues, and the quality of the presented content (Molina-Recio et al., 2020; O’Neill et al., 2022).

Within the set of applications developed to monitor chronic diseases (Rosa et al., 2019), we have the applications that monitor the routine of diabetic patients (Rodriguez and Wägner, 2019). Studies show that Diabetes Mellitus (DM) is one of the most critical health problems today because it is a disease with high morbidity and mortality (Assunção et al., 2001; International Diabetes Federation (2021); Sousa et al., 2019). DM is a chronic and complex metabolic disorder characterized by impaired metabolism of glucose and other energy-producing substances. This disease is associated with various complications in organs essential for maintaining life (Assunção et al., 2001).

Chronic diseases need constant attention and medical care; among several diseases, DM is considered one of the most affecting contemporary man. A survey by the International Diabetes Federation (IDF) estimated in 2021 that 10.5 percent of the world’s population aged 20 to 79 years (537 million people) were living with diabetes. If current trends persist, the number of people with diabetes will be over 783 million by 2045 (International Diabetes Federation, 2021). The routine of the diabetic patient requires strict and efficient control.

The prevalence of smartphones use has increased the impact of mobile technologies on the experience and use of health services. According to the World Health Organization (WHO), mobile health mHealth, is defined as a medical practice supported by mobile devices such as smartphones, tablets, patient monitoring devices, and other wireless devices. Mobile apps in this context should have features that serve the DM patient, such as measurement of blood glucose level, medication control, blood pressure, body mass index, cholesterol, and other functionalities that attend to the reduction of diabetic complications (Menezes et al., 2016).

In view of this scenario, this article presents the evaluation of the usability and User experience (UX) of five mobile applications to assist in the daily lives of diabetic patients. This research aims to investigate the relationship of apps with the behavior of diabetic patients who require daily monitoring. Thus, we have investigated the following research questions:

RQ1. How efficient, transparent, and reliable are mobile apps for people with diabetes?
RQ2. What are the level of stimulation and novelty of the mobile applications for people with diabetes?

RQ3. What is the level of conformance to the HE4EH (Heuristic Evaluation for mHealth) of the mobile applications for people with diabetes?

To investigate these research questions, our study examined the performance of apps in terms of UX attributes and usability aspects. To accomplish this, we utilized a UX questionnaire to evaluate the UX of mobile apps with a sample size of 10 participants who were diabetic patients aged between 18 and 53. Additionally, we conducted a heuristic evaluation of diabetes-specific apps to assess their usability. For this evaluation, we employed the HE4EH instrument proposed by Khowaja and Al-Thani (2020).

Our study is an extended version of the paper “Medication time? A User Experience Evaluation of Mobile Applications targeting People with Diabetes”, published in the proceedings of the XXVIII Brazilian Symposium on Multimedia and the Web (Rodrigues et al., 2022). In this paper, we present a more profound analysis of the mobile apps, including the heuristic evaluation.

The results of both assessments have indicated good usability of the five applications evaluated, emphasizing the efficiency aspect, and presenting values above average, according to the Likert scale adopted. Only the mySugr application presented problems concerning some UX attributes. The study also indicated that functionalities such as interactions between medical professionals and patients are essential for this type of application. These results may contribute to the digital transformation process of the health field that is also accelerating in Brazil (Uslu et al., 2020).

The remainder of this paper is organized as follows: Section 2 presents the theoretical background necessary for understanding the research. Section 3 discuss the related work. The methodology adopted in this research is described in Section 4. Section 5 summarizes the main characteristics of health applications in the context diabetes considered. The evaluation of the diabetes-focused health apps is presented in Section 6. Finally, Section 7 presents the conclusions and future work.

2 Background

2.1 Usability

According to Nielsen (1994), usability is an essential quality attribute that evaluates the ease of use of an interface. In simpler terms, it refers to how easily a user can interact with a tool or system. Nielsen identifies five factors associated with usability:

- **Learnability**: The system should have a simple and intuitive learning curve, allowing users to quickly understand and benefit from it.

- **Efficiency**: The system should be efficient to use, enabling users to be highly productive once they have learned how to operate it.

- **Memorability**: The system should be intuitive and easy to remember, allowing users to continue using it after a period of non-use without having to relearn everything.

- **Errors**: The system should have a low error rate, minimizing user mistakes during its operation, and enabling quick recovery in case errors occur.

- **Satisfaction**: The system should be easy and enjoyable to use, ensuring user satisfaction and ultimately leading to positive experiences.

Similarly, ISO 9126, also known as the International Standard for Software Quality, takes a user-centric approach to address the topic. The standard uses the term “characteristic” to describe the attributes that software should possess. These characteristics include effectiveness (a combination of efficacy and efficiency), productivity (referring to the appropriate allocation of system resources in relation to achieved effectiveness), safety (involving acceptable levels of risk of harm), and user satisfaction. ISO 9126 also provides specific guidelines and criteria for measuring and evaluating each of these characteristics during the software quality evaluation process (Iso, 2001).

2.2 User eXperience - UX

The term “user experience” pertains to a user’s emotions and overall experience when interacting with a product, system, or service. It encompasses a broader scope, including usability, but it goes beyond that. UX considers the user’s perceptions, feelings, attitudes, and behaviors throughout their entire journey with the product or service. It considers aesthetics, emotional impact, brand perception, and how well the user’s goals and needs are met (Hassenzahl et al., 2010; Sousa et al., 2006; and Rogers and Smith-Lovin, 2019).

Although UX is subjective and personal, system can incorporate features that promote good emotions in users and avoid causing unpleasant sensations, always respecting the limitations of users. Some essential aspects for UX to be considered during the (re)design of an interactive system, such as attention, rhythm, fun, interactivity, conscious and unconscious control, and narrative style (Rogers and Smith-Lovin, 2019).

Lopes and Valentim (2019) suggest that mobile applications have specific treatments in their evaluations due to the characteristics and limitations of mobile devices. In addition to usability, the authors argue that the user UX is critical to the success of a mobile application. They suggest that evaluations of mobile applications should address both quality aspects related to user satisfaction and their feelings and emotions when using these applications. However, through a systematic mapping of the literature carried out in the research, the authors noted that few studies address the joint use of UX and usability in the design of mobile applications. The research presented an initial proposal for a technique called Usability and UX Design Technique for Mobile Application (UUDT-MA), which contains recommendations regarding the design of mobile applications, integrating UX and usabil-

---

¹Usability can be viewed as a part of UX focusing on the user’s effectiveness, efficiency, and satisfaction while executing specific tasks with a product/software. UX goes beyond usability and focuses on the quality of interaction between individuals and products (generally technology-related) and other individuals, as well as the emotional and cognitive consequences that arise from this interaction (Hassenzahl et al., 2010).
ity principles to support designers in the early stages of mobile application development.

2.3 Methods for Evaluating UX

UX evaluation methods include identifying and understanding your users, their environment and tasks in context. They collect data on user needs and wants and evaluate attitudes, opinions and user impressions (Laugwitz et al., 2008). Evaluators could inspect these criteria through interviews, questionnaires, focus groups, card classification, and contextual investigation.

Questionnaires, however, are a low-cost and efficient tool (Schrepp et al., 2017). Unlike interviews, questionnaires allow you to collect data from a more significant number of people, even geographically dispersed, composing much larger samples than interviews or focus groups (Barbosa et al., 2021).

An example of a questionnaire for evaluating experience is the User Experience Questionnaire (UEQ²). It consists of 26 items grouped into six categories (Laugwitz et al., 2008). Both classic usability aspects (efficiency, perspicuity, dependability) and UX (stimulation, novelty) are measured. Figure 1 illustrates the categories and items evaluated by the UEQ, the instrument adopted in this research.

The questionnaire for this research included UEQ aspects of attractiveness, pragmatic quality and hedonic quality (Valentim et al., 2015). For the attractiveness dimension, the questionnaire quantifies the global value of the application based on the perception of quality. For hedonic quality, it checks how the application can support the needs to develop and how the application can advance in terms of originality, interest, and stimulation. For pragmatic quality, the questionnaire identifies the quality of an application and the degree of success that users achieve their usage goals.

The positive emotional involvement of users during interaction can lead to acceptance of an interactive system and changes in user behaviour. It is up to the interaction designer to decide which subjective aspects should be promoted during the interaction and articulate this with the other quality of use criteria (Barbosa et al., 2021).

2.4 Heuristic Evaluation of Health Apps

Heuristic evaluation is an method of usability analysis in which several experts are presented with an interface design and report on observed considerations (Nielsen and Molich, 1990). To support this activity, a set of appropriate list of heuristics may be used as a checklist to guide the evaluation. In addition, depending of the application type, different sets of heuristics can be used.

In this context, Khowaja et al. (2020) propose a set of heuristics to support the evaluation of mHealth applications. This set is an specialization of traditional existing heuristics. In our research, we selected ten heuristics proposed by the instrument, focusing on the assessment of: system status visibility; user control and freedom; the combination between the system and the real world, consistency and pattern; error prevention; helping users recognize, diagnose, and overcome errors; flexibility and efficiency of use; aesthetic and minimalist design; general help and documentation; and finally, self-monitoring.

We adopted the usability inspection technique to apply this instrument. In it, experts verified the adequacy of checklist items for each heuristic and identified its severity level when such an item is violated.

3 Related Work

Evaluating the usability and quality of mobile health apps is a recurring topic (Zapata et al., 2015). For example, Barbosa et al. (2018) presented the Emagreça@Saudável app that had its ergonomics evaluated by eight experts. Regarding diabetic patients, we also found research focusing on usability evaluation, such as that presented by (Vêscovi et al., 2017; Marques et al., 2020; Vêscovi et al., 2017).

Vêscovi et al. (2017), for example, describe the process of developing and validating a mobile device application on foot risk assessment and classification for people with DM. For this, the authors conducted a methodological study in four stages: Definition of requirements and elaboration of the conceptual model; Generation of implementation alternatives and prototyping; Testing and Implementation. The authors evaluated usability following Nielsen’s heuristics and through validation with nurses checking functionality, reliability, usability, and efficiency. As a result, the authors developed a mobile application which presents elements for risk assessment and classification integrated with the clinical findings with recommendations for each type of risk.

Marques et al. (2020) evaluated the end-user usability of a prototype application for diabetic foot self-care. The study presents a heuristic evaluation of the usability of a hybrid app. The assessment had 15 users of an ambulatory service of attention to the person with diabetes. The study was conducted in the Brazilian Northeast’s capital city in April 2018. The lowest score obtained was 77, and the highest was 112, with average overall usability of 96.1 points. These results attest a good usability according to the scale used.

Several other works seek to identify the state of the mobile applications’ usability built for health professionals. Among them, authors da Silva et al. (2021) deal with an integrative literature review answering the question: “how is the usability of mobile applications built for health professionals analyzed?” The search was done in indexed databases seeking Portuguese, English, and Spanish papers. The results identified eight articles from South America (n=2), Asia (n=2), North America (n=2), and Africa (n=2). The mobile applications found in the documents focused on nurses and physicians mainly. The authors evaluated them according to ease of use, functionality, innovative character, reliability, efficiency, and appropriateness of the app content.

To understand how interactions with systems and or products occur and evoke emotion and experiences, Forlizzi and Battarbee (2004) sought to understand the various connotations of experience in the use of interactive systems, researching the experience resulting from interactions between people and products. Paiva et al. (2020), in turn, identified state

²https://www.ueq-online.org/
of the art in the literature on developing mobile devices in applications for elderly health. A systematic mapping performed by the researchers was conducted to provide an interdisciplinary investigation of articles addressing applications for the health of the elderly. The data were collected under the following classification: destination of the application, the profile of the elderly, spatiotemporal distribution, techniques for empirical validation, and type of software engineering research.

In summary, much of the existing literature (Barbosa et al. (2018), Vêscovi et al. (2017), and Marques et al. (2020)) focuses on proposing mobile health applications and subsequently conducting usability evaluations of the developed systems. In contrast, our paper evaluates a more extensive range of mobile applications aimed at supporting diabetes management during patients’ daily routines.

Furthermore, we encountered some secondary studies related to m-health usability evaluations (da Silva et al. (2021), Paiva et al. (2020)). However, our approach goes beyond by employing mixed methods: (i) an evaluation of end-user experience and (ii) a heuristic evaluation of usability by HCI specialists.

## 4 Methodology

Figure 2 illustrates the methodology used to conduct this research. The methodology was applied to evaluate the selected mobile applications for people with diabetes. The goal was to evaluate the level of efficiency, transparency, reliability, the level of originality and stimulation of such applications. The methodology consists of three steps: (1) identification of mobile applications; (2) UX evaluation, and (3) heuristic evaluation. Each step is described in the following subsections.

### 4.1 Identification of mobile applications

In order to identify mobile applications developed for people with diabetes, we started to search mobile apps in the Google Play Store. We filtered our search results by selecting diabetic tracking and monitoring applications in which patients can manage features such as their glucose levels; regulating their diet; receiving alerts for measurements or injections; and tips from professionals. Furthermore, we analyzed apps identified in academic studies extracted from Google Scholar³, which were available for download in Android Playstore.

We selected six Android mobile applications since it is the leading mobile operating system ⁴.

The selected applications were:

1. mySugr - Diabetes Tracker Log
2. Blood Sugar Tracker
3. Blood Sugar: Diabetes Tracker
4. Glucose Tracker
5. Glic
6. Diabetes - Sugar in the Blood

The applications used in our evaluation are presented in Section 5.

### 4.2 UX evaluation

In this section, we present the selection of the target audience and the procedures used to conduct the UX evaluation.

³https://scholar.google.com.br/
⁴In Brazil there are about 84.84% users of the Android platform. Available on: https://gs.statcounter.com/os-market-share/mobile/brazil
4.2.1 Selection and characterization of the target audience

In our research, we selected the participants based on the Convenience Sampling method, which refers to choosing people that are convenient to the study, in our case, we have used known contacts (Stratton, 2021).

We contacted the people diagnosed with diabetes initially. Then, a group was created on a synchronous chat (i.e., WhatsApp) to guide the participants during the research since the evaluation was conducted remotely due to the COVID-19 pandemic. We then presented our research proposal and objectives to the participants.

4.2.2 Materials and Methods

The ISO 9241 defines, evaluating the initial experience of using a product, system, or service involves examining a person’s responses and perceptions resulting from its usage. To observe participants’ responses and perceptions regarding our study subjects, we administered a questionnaire. Hence, our research aimed to evaluate the user perspective for each application individually, based on the criteria of the UEQ Schrepp et al., 2014.

4.2.3 Procedure

To evaluate the participants’ first experience using the selected mHealth applications, the researchers demonstrated their functionalities and provided a roadmap of actions the users needed to follow. Each participant was allowed to choose two out of the six applications presented. The roadmap was given orally and served as a guide for participants to perform tasks such as starting the application, setting the blood sugar level, and closing the application. The applications were installed and used on the participants’ cell phones. The evaluation was conducted remotely and lasted for an average duration of 15 to 30 minutes.

4.2.4 Ethical Aspects

We ensured that ethical aspects regarding the users were respected throughout this research. Since these evaluations did not entail continuous use of the applications but consisted of single-use and evaluation sessions, we did not identify any significant risks to the participants. First, users were required to fill out an Informed Consent Form (ICF) indicating their agreement to participate in the research, and they had the opportunity to withdraw from the study at any time if they wished to do so. Additionally, user data was anonymized, and no records of images or videos from the evaluation sessions of the applications were kept.

4.2.5 Data analysis

The data analysis was performed using three questionnaires. Initially, the participants were invited to answer the Characterization of the Target Audience questionnaire and, subsequently, the UX Evaluation questionnaire. The data collected in the UX evaluation of the applications were analyzed quantitatively using the UEQ analysis tool (spreadsheet that calculates the attractiveness, transparency, efficiency, reliability, stimulus and originality indices with the information provided by the user regarding use).

Each participant answered the second questionnaire twice since they evaluated two applications. Once the evaluation was completed, the third questionnaire called Experience report after using the applications was applied. The goal was to collect data that can reveal the UX, the frequency of apps usage after our evaluation and the difficulties faced by them. All questionnaires are available on the repository.⁵

⁵https://github.com/elannemendes/webmedia2022
4.3 Heuristic Evaluation

In this section, we present the procedures used to conduct the heuristic evaluation.

4.3.1 Materials and Methods

The evaluation was based on heuristics for mHealth apps (HE4EH) proposed by Khowaja and Al-Thani (2020). The HE4EH method contains an extensive list of items to be evaluated and categorized into 25 heuristics taken from 8 articles in the literature. It has a total of 436 checklist items. For instance, they cover Nielsen’s usability aspects, mobile interaction issues, accessibility features, e-health topics, and monitoring heuristics.

We chose the nine heuristics that appeared the most in the literature selected and one (H25, self-monitoring) related to the type of application we were evaluating. The nine most frequent heuristics also have a strong relationship with UX and usability aspects that were the focus of the previous stages of the research. Those heuristics are described in details in the studies proposed by Group et al. (2010) and Lacerda et al. (2015).

We used these heuristics to evaluate the five applications selected by the participants of the UX evaluation (see Section 5). Table 1 presents those ten heuristics selected by the authors and the number of their checklist items to be evaluated.

We created a spreadsheet in Google Forms to report the problems identified in the evaluation. For each identified problem, the evaluator notified the problem description; a possible solution; the heuristic violations; the nonconformity checklist items, and the severity classification of the non-observance. The severity classification ranged from 0 to 4, respectively, “unimportant”, “cosmetic”, “simple”, “serious”, and “catastrophic”.

Table 1. Heuristics selected by the authors

<table>
<thead>
<tr>
<th>#</th>
<th>Heuristic</th>
<th>Checklist items (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Visibility of system status</td>
<td>35</td>
</tr>
<tr>
<td>H2</td>
<td>User control and freedom</td>
<td>23</td>
</tr>
<tr>
<td>H3</td>
<td>Match between system and real world</td>
<td>26</td>
</tr>
<tr>
<td>H4</td>
<td>Consistency and standards</td>
<td>40</td>
</tr>
<tr>
<td>H5</td>
<td>Error prevention</td>
<td>16</td>
</tr>
<tr>
<td>H6</td>
<td>Help users recognize, diagnose, and recover from errors</td>
<td>04</td>
</tr>
<tr>
<td>H8</td>
<td>Flexibility and efficiency of use</td>
<td>23</td>
</tr>
<tr>
<td>H9</td>
<td>Aesthetic and minimalist design</td>
<td>36</td>
</tr>
<tr>
<td>H10</td>
<td>Help and documentation</td>
<td>22</td>
</tr>
<tr>
<td>H25</td>
<td>Self-monitoring</td>
<td>67</td>
</tr>
</tbody>
</table>

4.3.2 Procedure

Three evaluators conducted the heuristic evaluation (all are authors of this paper). Each expert evaluated individually the five applications (i.e. mySugr - Diabetes Tracker Log, Blood Sugar Tracker, Blood Sugar: Diabetes Tracker, Glucose Tracker and Glic), and filled their usability problems reports. The experts reported the results using Google Forms and then stored them on a spreadsheet to be later organized for the data analysis step. Once the evaluations were completed, the evaluators consolidated the reports by discussing the identified usability problems and their severity. The consolidation was conducted in a remote meeting and took about 5 hours.

4.3.3 Data analysis

The experts evaluated the five applications based on the checklist items of each heuristic. The evaluations generated much information regarding the usability and functionality problems identified for the applications. Initially, the experts organized the information for each application and then manually performed the data analysis. Next, the number of checklist items violated by the heuristics and the severity level was accounted for.

5 Mobile Health Apps Selected

5.1 mySugr - Diabetes Tracker Log

The mySugr app ⁶ is a free mobile diary, which helps users to keep their diabetes data under control. The application consists of some functionalities, such as:

- Customized dashboard with information about diet, medications, carbohydrate intake, and blood glucose levels;
- Simple graphs of blood glucose levels;
- HbA1c estimation, which measures the amount of blood sugar (glucose) attached to hemoglobin;
- Daily, weekly and monthly reports that can be shared with doctors;
- Backup data following regulatory standards.

Figure 3 shows screenshots of two features of mySugr (i.e., blood glucose logging and diabetes management) extracted from the Android Playstore.

![Figure 3. mySugr — Screenshots.](https://play.google.com/store/apps/details?id=com.mysugr.android.companion&hl=en&gl=US)
5.2 Blood Sugar Tracker

The Blood Sugar Tracker is a mobile app which aims to help people with diabetes to record their blood sugar levels and perform analyses on the recorded data.

The application is free and allows functions such as:

- filtering blood sugar readings by event type (before a meal, after a meal and fasting),
- getting blood sugar levels automatically calculated,
- editing your blood sugar range based on your condition, and
- writing notes in your blood sugar records.

Figure 4 shows two examples of these features.

Figure 4. Examples of Blood Sugar Tracker features.

5.3 Blood Sugar: Diabetes Tracker

This Smart Health Apps Inc app allows the user to convert blood sugar values between various measurements. It is a free tool for tracking blood sugar levels and converting them to units from mmol/L to mg/dL and vice versa.

The app also includes helpful resources like recipes and videos, providing blood sugar testing with an understanding of the use and storage of dietary sugars. Users have access to seven charts allowing them to analyze the input data, show trends, changes in glucose, hemoglobin, well-being and etc.

Figure 5 shows examples of application functionality.

Figure 5. Screenshots of the Blood Sugar: Diabetes Tracker.

With the help of regular sugar recording, the app makes it possible to observe blood sugar levels. It suggests how the situation can be treated from this information and provides the user’s physician with all the required information to treat and control diabetes effectively (see Figure 6).

Figure 6. Screenshots of the Glucose Tracker

5.4 Glucose Tracker

Glucose Tracker is a free app with some features paid. It focuses on people with gestational diabetes, types 1 and 2. The glycemic diary will allow tracking of regular hemoglobin, sugar levels, pressure, insulin reminders, medications, condition and weight for each record. The app has an insulin tracker that helps set the daily value and monitor the medications taken by the patient.

5.5 Glic

The app Glic - Diabetes e Glicemia helps people with diabetes, connecting patient and medical staff to facilitate carbohydrate counting and insulin dose calculation. It was developed to serve patients with type 1 diabetes, type 2 diabetes, LADA, MODY, and gestational diabetes (see Figure 7).

This free application allows data sharing between professionals involved in the treatment. The user can program alerts for hypoglycemia, carbohydrate counting, insulin dose calculation, generate blood glucose graphs, and automatically calculate doses based on their medical prescription.
5.6 Diabetes

The application Diabetes\textsuperscript{11} assists in monitoring and analyzing data, such as Diabetes, blood glucose level. This free app works as a mobile assistant for people with diabetes. Some features of the app are:

- Annotation of blood glucose level - (mg/dl) or (mmol/l);
- Graph generation;
- Data exchanging between patient and doctor;
- Exportation of collected data to CSV and XML formats.

Figure 8 shows examples of application functionality and graphics.

5.7 Apps comparison

Figure 9 illustrates a comparison of the six applications previously presented. They have three features in common: glycemic recording, medication reminders, and report generation. These three resources are significant for the life of diabetic users. They add agility to their routine. The reports are a history of their disease status and can be used for glycemic control and communication with the users’ physicians.

Only one app has a food table feature. The food table is an informative resource with some types of food that the user may or may not adopt for their meals.

We emphasize that among the features of the applications, there is the option to share historical blood glucose and insulin data. However, only three applications have this feature.

6 Assessment of mHealth apps focusing on diabetes

6.1 Participants Profile

The questionnaire for characterization of the target audience was answered by 10 voluntary participants, six of them female and four male. As for the participants’ level of education, five of them had complete higher education, four had incomplete higher education, and one participant had complete elementary school education. All of them had a medical diagnosis of diabetes, where: six (60%) of them were diagnosed with type 2 diabetes (i.e., a type in which it is considered a chronic disease and the body does not produce insulin or creates insulin resistance); and the other four (40%) participants were diagnosed with type 1 diabetes, which is considered a chronic disease in which the pancreas produces little or no insulin.

As illustrated in Figure 10, it was possible to characterize the main symptoms that negatively affect the routine of the participants. It is worth noting that they could mark more than one option. Based on the answers, we observed that for the study subjects, the diabetes affects mainly the high pressure (10) and loss of energy and constant fatigue (7).

About the respondents’ perception regarding the use of software technologies by people with diabetes, 7 out of 10 participants (70%) answered that they strongly agree that the use of technologies supports in daily life. Also, three of them strongly agree that technologies can encourage glucose control and monitoring. However, five disagree that there are technologies suitable for people with diabetes.

6.2 UX Evaluation

As we described in the methodology Section (see Section 4), a presentation of the 6 selected applications was provided for the participants. The choice of application was up to each one of them that should choose two applications. After choosing and using them, each participant answered the questionnaire about his perception of the applications.

Table 2 presents the mHealth apps selected by the participants. The most chosen were: Glucose Tracker (8 participants), Glíc - Diabetes and Glicemia (5 participants). The other apps chosen by the participants, were: mySugr - Diabetes Tracker Log (3), Blood Sugar: Diabetes Tracker (2) and Diabetes (2). The Blood Sugar Tracker app was not chosen by any of the participants.

The UX evaluation questionnaire was applied separately for each application and online. An average was made between the answers of the participants to obtain results of the characteristics defined in the questionnaire in relation to the use of the chosen application, such as: attractiveness,
transparency, efficiency, reliability, stimulus and originality. Figure 11 shows the main results of the questionnaire, with the evaluation of all apps, with indices for attractiveness, pragmatic quality and hedonic quality, with mean and variance values and graphs with benchmark samples, respectively. Overall, the results ranged from 1.92 to 2.75 regarding the attractiveness of the apps. In pragmatic quality, they ranged between 1.00 and 2.21 and in hedonic quality the results ranged between 0.08 and 2.09.

According to the UEQ Schrepp analysis tool Schrepp (2019), values between -0.8 and 0.8 represent a neutral evaluation of the corresponding scale, values greater than 0.8 represent a positive evaluation and values less than -0.8 represent a negative evaluation.

As depicted in Figure 11, all applications showed positive results for attractiveness and efficiency. Four apps presented positive evaluations for attractiveness, pragmatic, and hedonic qualities. The app Glucose tracker outperformed the others. Only the mySugr - Diabetes Tracker Log had a neutral evaluation for the hedonic quality, having problems with novelty and stimulation aspects.

6.3 User perception after using the apps

After the UX evaluation, the participants reported their perception of the use of the apps. One of the relevant questions for the survey was to understand if any of the participants already used any apps to support their illness on a daily basis. From this question, the result was that none of the 10 participants did use.

We also investigated what the participants missed most in the evaluated apps. Some of them presented similar reports indicating the lack of resources in technology in support of interaction among users with diabetes. This is the case of participants P6 and P9’s answers.

“An application that allows users to interact, such as giving tips, talking about the disease and symptoms” (P6)

“To be able to interact with other users” (P9)

The participants were asked an open question to identify potential restrictions for using the mHealth applications, such as a lack of internet or space on the cell phone, knowledge/dissemination, interest, and confidence. Figure 12 presents the identified impeditive factors for using the apps.

The result pointed out that 10 (100%) of the participants marked lack of trust as a strong impeditive, and 6 participants (60%) mentioned lack of space on mobile and lack of...
knowledge/dissemination as impeding factors. Lack of Internet was the least mentioned (2 participants) as an impeding factor. In our survey, a question was also directed to identify how many participants continued using the app after the study. According to the results of the questionnaire, it was possible to identify that 7 of the 10 participants adopted the use of health apps, among the Glucose Tracker, Diabetes and Glic, mentioned by the participants themselves, to support their daily routines.

### 6.4 Heuristic Evaluation

The evaluation was conducted by three evaluators who are the first three authors of this paper. One of them is PhD student and the others are master’s students. All of them have experience in mHealth applications and heuristic evaluation. However, only one expert (PhD student) was familiar with diabetes apps.

As described in Section 4.3 of the methodology, we performed the heuristic evaluation based on 10 heuristics selected from HE4EH. Table 3 shows the number of violations per heuristic found in the five evaluated applications.

The results show that one or more usability violations were found for most heuristics in the applications Glucose Tracker (105 violations), mySugr (89 violations) and Blood Sugar: Diabetes Tracker (109 violations). This latter with largest number of violations overall. The applications Diabetes and Glic did not present violations for the Combination between system and real world (H3) heuristic. Also, Glic did not present any violations also for H5, H6 and H9 heuristics. The main heuristics frequently violated were Self-monitoring (H25) (320 violations), followed by Visibility of system status (H1) (36 violations), Help and documentation (H10) (28 violations), Flexibility and efficiency of use (H8)(27 violations), and Error prevention (H5) (21 violations).

Regarding the violations per heuristic, Table 3 shows the
The highest number of violations (dark green color background) identified in the applications for each heuristic, which were:

- H1 with 9 violations in the Glucose Tracker, Blood Sugar: Diabetes Tracker and Diabetes apps.
- H2 with 6 violations in Gluc app.
- H3 with 2 violations in Blood Sugar: Diabetes Tracker app.
- H4 with 5 violations in mySugr app.
- H5 with 7 violations in Glucose Tracker app.
- H6 with 3 violations in Glucose Tracker, mySugr, Blood Sugar: Diabetes Tracker and Diabetes apps.
- H8 with 7 violations in Glucose Tracker app.
- H9 with 7 violations in Diabetes app.
- H10 with 11 violations in Diabetes app.

Based on the results, we observed that Blood Sugar: Diabetes Tracker and Diabetes apps had more violations in the Self-monitoring heuristic (H25), an important characteristic for apps aimed at people with diabetes. Furthermore, this heuristic had a high number of violations in all apps (60 or more violations). mySugr app is the one with lower number of violations for H25.

Table 4 shows the severity level of the problems identified using HE4EH in evaluating the five applications. Usability problems were rated on a 5-point scale (“0 - unimportant”, “1 - cosmetic”, “2 - simple”, “3 - serious”, and “4 - catastrophic”). The total values indicated that most problems found were simple usability problems with low-resolution priority, 396 in total. The application with the highest number of identified problems (112) was Diabetes with a variation between 1 and 4 in severity level, with a high rate of simple and serious problems (with high priority for correction). All applications presented problems in the serious severity level, 68 in total. However, only Blood Sugar: Diabetes Tracker, Glucose Tracker, and Diabetes applications presented problems regarding the catastrophic severity level, which prevent the user from completing the task and must be urgently corrected, Blood Sugar: Diabetes Tracker presented 9 problems, followed by Glucose Tracker with 6 problems, and Diabetes with 2 problems.

6.5 Discussion

The first results that were surprising in this research relate to the low uptake of apps by the ten participants. Although, they showed awareness that the use of technologies in everyday life could be positive, yet none of the participants made use of apps prior to this study.

6.5.1 RQ1. How efficient, transparent, and reliable are mobile apps for people with diabetes?

The results showed that the app Glucose Tracker has a higher pragmatic quality score, followed by the app Gluc. To determine the pragmatic quality, the values of efficiency, transparency and reliability are considered. Regarding efficiency, all apps showed results rated as either Above average, Good or Excellent. The reliability aspect was the one with the lowest rating among the three aspects considered, ranging from Poor to Excellent. It is possible to observe, that 3 of the 5 applications excelled in classic usability aspects such as attractiveness, which is the overall impression of the product, and transparency, which is the familiarization. These results indicate that, in general, all apps present a positive evaluation of the pragmatic quality, with values above 0.8 as determined by the scale presented in the previous section. Therefore, we claim that all applications, according to the participants, present a positive degree of usability.

6.5.2 RQ2. What are the level of stimulation and novelty of the mobile applications for people with diabetes?

The research presented results for question RQ2, considering hedonic quality, with samples with relevant values for the app Glucose Tracker, which considers originality and stimulation. Overall, the participants’ answers about the UX were positive, regarding the application Glucose Tracker app, presenting higher mean and variance values than all other apps, as shown in Figure 11. The mySugr - Diabetes Tracker Log app showed values on the scale below 0 for novelty, ranking on the lower bound scale of the benchmark.

These results indicate that the mySugr - Control diabetes! app presents a neutral rating according to the corresponding scale, while the rest of the apps present a positive rating of hedonic quality, with values above 0.8. With this, it is pos-
We found it concerning that more than 40% of the self-monitoring heuristic items were not fulfilled in the applications, as shown in Table 4. This indicates the need for further improvements for applications accessible to all users, including long-term patients, their caregivers, and other stakeholders. Often, these users have limited experience interacting with apps due to various challenges they may face in their daily lives and work (Wang, 2017).

Surprisingly, the apps we selected performed well in the HE4EH analysis. We assessed ten heuristics from HE4EH, encompassing a total of 292 checklist items (Table 1). Table 5 presents the conformance percentage for each application heuristic. All mobile applications achieved high levels of compliance for most heuristic items, except heuristic H25, which pertains to self-monitoring.

We observed that the checklist items defined the importance of using and evaluating diabetes apps using HE4EH. The evaluated heuristics provided valuable insights for continuously improving these mobile applications.

Despite the positive results in most of the heuristics, we found it concerning that more than 40% of the self-monitoring heuristic items were not fulfilled in the applications. This indicates the need for further improvements for the studied applications, as shown in Table 5.

### 6.5.4 Research limitations

Like any scientific research, this work presents some limitations. Regarding the evaluation of the UX: (i) the number of participants of the evaluation (10) may not be a representative number for people with diabetes in the Brazilian context, besides, the recruitment occurred by convenience; (ii) The evaluated apps (5) may not represent all the apps that help in their daily routine of people with diabetes; these results are exploratory and cannot be generalized; (iii) The sequence of app use in the evaluations followed the order of preference of the participants. No pattern was observed in this order. However, previous experience with the first application may have influenced the evaluation of the second.

Regarding the heuristic evaluation phase, the authors of this work may have limited the identification of application compliance with other heuristics by selecting only some HE4EH heuristics for evaluation. However, we stand out that 292 items were checked. Another restriction was the evaluation involving three evaluators, the minimum amount recommended in heuristic evaluation studies. More evaluators would enrich the collection of opinions and discussion of results. The self-monitoring heuristic verification items also depended on patient monitoring equipment to perform connectivity and usability tests.

### 7 Concluding remarks and future work

This work had as its primary objective to evaluate the UX of mHealth applications by people diagnosed with the chronic disease of diabetes. We also perform a heuristic evaluation by domain experts using HE4EH. The elaboration of the work started with a study to identify the existing applications both on Google Play and academic works. We conducted UX assessments of five apps with ten users, identifying their knowledge about mHealth apps and discovering the positive and negative effects of using technology to help their daily routines.

We also characterize the main aspects of diabetes that negatively affect the participants’ routines. The initial study was conducted remotely with the participation of 10 users diagnosed with type 1 or type 2 diabetes. The use of the applications represented a new possibility to contribute to the participants’ daily life. The tested applications positively benefit the users and can support the traditional glucose level notation and analysis methods. The heuristic assessment performed with HE4EH points out how to improve a mHealth application and how its checklist items specify criteria indispensable to cover any gaps the application presents.

In future work, we intend to conduct new evaluation studies with other existing applications, increase the number of participants, and extend the age groups not included in the current research. In addition, the results should guide the creation of an application that uses new aspects, such as gamification and sensor data not found in the tested apps.

### Acknowledgements

The present work was supported by the Federal University of Ceara (UFC), Brazil. This research was partially funded by CNPQ through project (No. 314425/2021-7), the institution also benefited the first author and thank CNPQ for the Productivity Scholarship of Rossana Maria de Castro Andrade 1D (No. 306362/2021-0). We thank CAPES for the assistance with grant support that helped the second author.

Maria Elanne M. Rodrigues - Conceptualization, Methodology,
Table 5. Percentage of conformance by heuristic in each application

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Glucose Tracker</th>
<th>Glic</th>
<th>mySugr</th>
<th>Blood Sugar: Diabetes Tracker</th>
<th>Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>96.85%</td>
<td>98.6%</td>
<td>98.25%</td>
<td>96.85%</td>
<td>96.85%</td>
</tr>
<tr>
<td>H2</td>
<td>99.08%</td>
<td>98.62%</td>
<td>99.08%</td>
<td>99.31%</td>
<td>99.31%</td>
</tr>
<tr>
<td>H3</td>
<td>99.74%</td>
<td>100%</td>
<td>99.74%</td>
<td>99.48%</td>
<td>100%</td>
</tr>
<tr>
<td>H4</td>
<td>99.20%</td>
<td>99.60%</td>
<td>98.00%</td>
<td>98.80%</td>
<td>99.20%</td>
</tr>
<tr>
<td>H5</td>
<td>98.88%</td>
<td>100%</td>
<td>99.68%</td>
<td>99.04%</td>
<td>99.04%</td>
</tr>
<tr>
<td>H6</td>
<td>99.88%</td>
<td>100%</td>
<td>99.88%</td>
<td>99.88%</td>
<td>99.88%</td>
</tr>
<tr>
<td>H8</td>
<td>98.39%</td>
<td>98.85%</td>
<td>98.62%</td>
<td>99.08%</td>
<td>98.85%</td>
</tr>
<tr>
<td>H9</td>
<td>98.92%</td>
<td>100%</td>
<td>99.64%</td>
<td>98.2%</td>
<td>97.48%</td>
</tr>
<tr>
<td>H10</td>
<td>98.9%</td>
<td>99.56%</td>
<td>99.56%</td>
<td>98.24%</td>
<td>97.58%</td>
</tr>
<tr>
<td>H25</td>
<td>57.12%</td>
<td>57.12%</td>
<td>59.8%</td>
<td>55.78%</td>
<td>55.78%</td>
</tr>
<tr>
<td><strong>Média</strong></td>
<td><strong>94.70%</strong></td>
<td><strong>95.24%</strong></td>
<td><strong>95.23%</strong></td>
<td><strong>94.47%</strong></td>
<td><strong>94.40%</strong></td>
</tr>
</tbody>
</table>

References


Nielsen, J. and Molich, R. (1990). Heuristic evaluation of


Valentim, N. M. C., Silva, W., and Conte, T. (2015).