

Article

Testing the Quality of the Mobile Application Interface Using Various Methods—A Case Study of the T1DCoach Application

Marek Milosz , Małgorzata Plechawska-Wójcik  and Mariusz Dzieńkowski 

Department of Computer Science, Lublin University of Technology, 36B Nadbystrzycka Str., 20-618 Lublin, Poland; m.plechawska@pollub.pl (M.P.-W.); m.dzienkowski@pollub.pl (M.D.)

* Correspondence: m.milosz@pollub.pl

Abstract: The T1DCoach mobile application is designed to educate patients—children with type 1 diabetes (T1D) and their caregivers and diabetes educators. The idea behind the mobile application is that its users perform actions that the patient needs to perform in real life. These include measuring blood glucose levels, operating the insulin pump, meal calculation, bolus administration, etc. These in-application activities are performed on the patient’s digital twin. To increase user engagement, gamification elements have been implemented in the application. An important element of the T1DCoach mobile application is its interface, which should be adapted to very different groups of users: children, their caregivers and educators. In addition to presenting the T1DCoach application, the paper presents the stage examining the quality of the interface using three research groups: children, their caregivers and educators. The research was conducted using the scenario method, using eye-tracking, recording activities and thinking aloud. After the application testing sessions, surveys were carried out using the System Usability Scale method and focus group interviews were conducted. The research results are presented in the article along with the most important recommendations for improving the application interface.

Keywords: type 1 diabetes education; mobile application; interface quality; eye tracking; thinking aloud; System Usability Scale



Citation: Milosz, M.; Plechawska-Wójcik, M.; Dzieńkowski, M. Testing the Quality of the Mobile Application Interface Using Various Methods—A Case Study of the T1DCoach Application. *Appl. Sci.* **2024**, *14*, 6583. <https://doi.org/10.3390/app14156583>

Academic Editors: Frederico Branco, José Martins and Henrique Mamede

Received: 9 July 2024

Revised: 22 July 2024

Accepted: 25 July 2024

Published: 27 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Continuous technological progress means that many activities and operations previously carried out in a traditional manner are replaced by activities performed using computers and mobile devices, often via the Internet. This usually enables a significant improvement in the efficiency of the activities undertaken, their acceleration, reduction of costs and increase in their effects and benefits.

One of the sectors in which this process is particularly visible is medicine and the medical area of patient support. The use of Internet solutions, including mobile solutions, allows an increase in the level of knowledge and self-awareness of the patient, facilitation of patient control in the periods between medical visits or tests, enabling of regular doctor–patient contact during these periods, and support for regular self-reporting. Applications that support patients in everyday life with a chronic disease, such as diabetes or depression, and are educational, also support the process of faster implementation of the duties and rules related to disease management. Moreover, this type of application is increasingly used as a tool for collecting patient data by medical institutions [1–3].

However, research [4] has shown that there are no applications on the market supporting the education process for patients with type 1 diabetes (T1D) and their caregivers. Therefore, an effort was made to create such an application. T1D disease is genetic and incurable. In short, it means that the patient’s body does not produce insulin, which is necessary to maintain proper blood sugar levels. The only way to live with this disease is to properly administer external insulin. Unfortunately, insulin doses depend on many

factors, including the patient's parameters and the dynamics of his/her disease, as well as lifestyle, including food intake. T1D disease manifests itself among young people and they are faced with the need to control many aspects of their lives.

Medical-related applications are becoming more and more popular, and therefore their success on the market is conditioned by their appropriate quality and usability [5,6]. In such a case, the quality of the Graphical User Interface (GUI) plays a very important role in their evaluation by potential users. These applications should be ergonomic, efficient and it should be easy to use and understand the interface, even by users who have no experience with this type of program. Applications of this type can be classified as broad-use applications, i.e., they are used by very different groups of users. This means that the graphical user interface must be adapted to the needs of a wide audience, including children [7]. Therefore, it is important to be guided by existing criteria and rules for designing user interfaces [8] when designing and creating this type of software, which will ensure a positive reception of the application by a potential user [9]. It is also necessary to introduce into the application development process activities related to testing its GUI with the participation of its potential users and using the research results to improve its quality.

Apart from the Introduction, the paper presents a short review of the literature on software interface quality research, with particular emphasis on research on mobile applications with user participation. Then, the T1DCoach application and its interface as well as the initial research plan are presented. The methods used in the study and the principles of selecting research groups are also indicated. The research implementation process and its results are considered in detail. The final part of the article presents a discussion of the study results along with the main recommendations for improving the T1DCoach application interface. The article ends with conclusions.

2. Literature Review

Application interface quality testing, called Usability Testing, is a technique used to assess the interface quality by a control group consisting of future potential users [10]. The purpose of testing is to check how users manage the system while performing typical (or project-defined) tasks. During these tests, researchers observe (including recording, using various techniques for further research) their behavior and interactions with the app's GUI. During the tests, users from the control groups perform tasks by implementing prepared software use scenarios. In this method, to identify problems related to usability, quantitative data related to participants' performance during the study, qualitative data (cognitive problems), and the level of satisfaction of research participants with the use of the application are examined. The study should involve homogeneous users (the so-called research group) who represent the indicated Personas. As experience shows, the size of a homogeneous research group (i.e., people with characteristics corresponding to those of the Persona in areas important for cooperation with the IT system) should be from 6 to 8 people [11]. Another criterion for selecting the research group is experience in using the IT system. When examining an interface, the main factor is the experience of using it. There are three types of users: novices (without any experience), experienced users and experts [12]. The interface requirements for each user group are different.

Usability research uses various techniques to obtain data on the effectiveness and problems occurring when working with the IT system interface. These techniques are divided into objective and subjective [13]. Objective techniques allow measurement of work parameters such as the speed and quality of individual tasks, the number of clicks, the fact that a task was completed or not within the assumed time, the number of errors made or the number of calls for help. Subjective feelings about the quality of the interface are examined using sociological techniques, such as surveys or interviews with research participants.

The literature contains many publications on the evaluation of user interfaces for mobile applications [14] and methods used during evaluation [15,16], including the evaluation of educational applications [17–19]. Depending on the purpose of the assessment,

the size of the system and the general conditions for conducting the study, researchers use different analysis methods. A combination of several different methods is also very often used [20,21]. Prototyping methods [22,23], heuristic analysis [24,25], cognitive wandering [26,27], expert tests [28], as well as a numerous methods with user participation are used, including scenario testing methods and survey methods, e.g., the SUS (System Usability Scale) survey [29]. More advanced techniques require specific tools. An example is eye tracking [30,31]. This technique, as an objective method, allows for tracking the eye activity of subjects while using, e.g., a website, web or mobile application, and can be used for a multi-aspect analysis of GUI quality. This analysis may take into account not only the indication of weak and strong features and interface points, but also detailed user interaction with the application [32] and user activity in the application while performing specific tasks [33]. The combination of subjective methods, such as surveys and in-depth interviews, along with objective methods is very effective in the process of identifying errors in the GUI, even in a small research sample. This approach allows for identifying and hierarchizing problems related to the usability of the interface and indicating opportunities to remove errors [34,35].

The aim of the paper is to present the quality testing stage of the T1DCoach application interface using three research groups: children, their caregivers and educators.

This type of research most often uses a mixed method, which is why the method of testing the application interface using the scenario method was combined with the registration of activities by means of eye tracking, a SUS survey and a focused group interview. During each research session, users first took part in a test study, then completed the SUS survey assessing the usability of the application, and finally expressed their opinions during a focus group interview.

3. Materials and Methods

3.1. T1DCoach Mobile Application

The T1DCoach application (<https://play.google.com/store>, accessed on 1 July 2024) supports the education of the patient, as well as his/her caregiver and doctor, by shaping correct behavioral habits of patients with type 1 diabetes. The application was created using a simulation model [36] of a patient with T1D and his/her therapy. The simulation model is the progenitor of the digital twin of a T1D patient.

T1DCoach also uses elements of gamification to increase the involvement of adolescent patients in leading a proper lifestyle with type 1 diabetes. The application was created as a result of the project “Supporting education in type 1 diabetes in adolescents—T1DCoach mobile application” as part of the “Things are for people” program financed by the Polish National Centre for Research and Development (Grant No. Rzeczy są dla ludzi/0062/2020-00) and is available free of charge on Google Play (<https://play.google.com/store>, accessed on 1 July 2024) for Android devices under the name: T1DCoach.

The basic functionalities of the T1DCoach application include modelling the results of the required actions of a type 1 diabetes patient on his digital twin called GABI. The digital twin is visualized using an avatar. The virtual avatar must eat, take insulin, control the level of glucose in the blood (glycemia), operate the insulin pump, calculate boluses (additional doses of insulin), etc. The graphically animated reactions (and there are 24 of them) of the digital twin are adequate to the real ones, thanks to the developed and tuned simulation model of a patient with type 1 diabetes [36].

T1DCoach is a mobile application dedicated to Android devices. It is educational, not medical. Its task is to enable the care of a virtual patient, during which the user performs the following activities:

- checks the current glycemia level and intervenes if it is inappropriate;
- composes meals, specifying carbohydrate substitutes, and serves these meals with additional doses of insulin (so-called boluses);

- observes the course of therapy over a longer period of time and, if the results are unsatisfactory, reprograms the insulin pump (changes therapy recommendations) and/or changes the nutritional plan.

From the main application window, which contains the basic screen element, i.e., the animated avatar GABI, it is possible to access the four main functionalities offered by the tested application (tiles at the bottom of the main screen—Figure 1):

1. MEALS—allows the management of meals: composing and serving meals, calculating carbohydrate replacements and calculating the size of additional insulin doses (i.e., boluses);
2. INSULIN PUMP—enables the operation and management of the insulin pump (check and top up the insulin level in the tank, replace the pump's batteries, calculate insulin doses, configure base settings, bolus settings and pump sensitivity settings);
3. GLUCOMETER—used to immediately measure the glucose level and display it in the form of a notification;
4. CGM—(Continuous Glucose Monitoring) enables continuous, daily monitoring of glycemia levels.

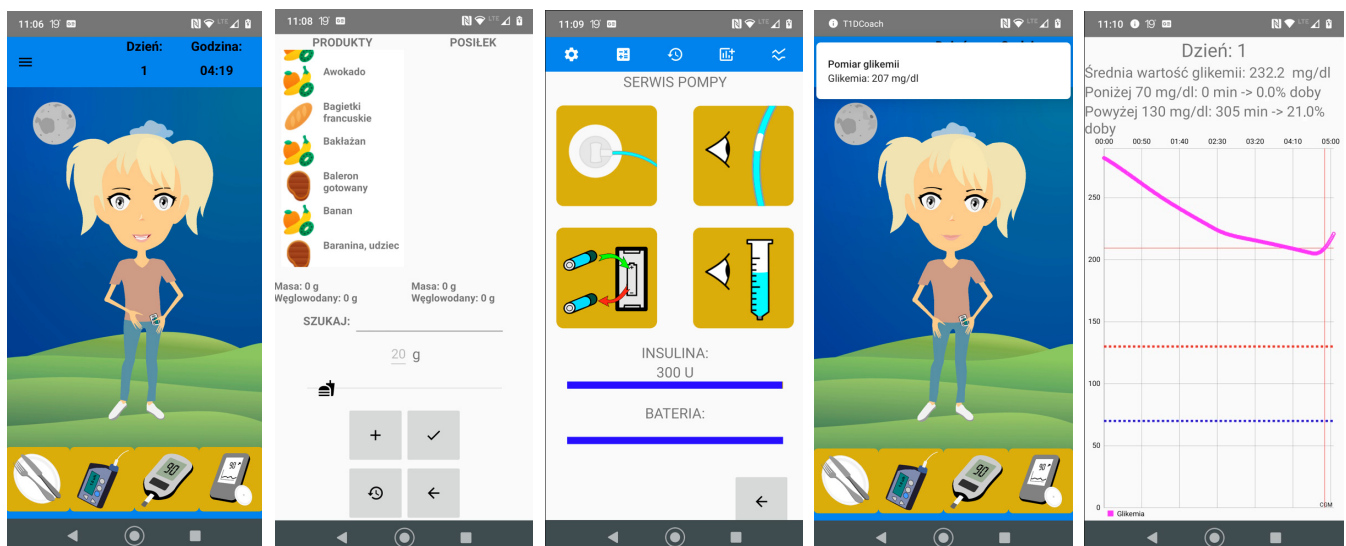


Figure 1. Selected T1DCoach application screens (in order from the left)—main, meal composition, insulin pump operation, glycemia measurement and CGM result (two lines indicate the normal glycemia levels).

3.2. Study Workflow

Research on the T1DCoach application interface began with planning. Figure 2 shows the planned workflow for testing the T1DCoach application interface. It consists of three stages: strategic planning, research implementation, postprocessing and report preparation. Each stage consists of sequential or parallel actions that provide the specific results shown in Figure 2.

3.3. Methods Used for Interface Quality Testing of T1DCoach Application

At the Strategic Planning stage (Figure 2), the user testing method was selected to test the quality of the T1DCoach application interface. Three groups of users have been defined, who are the natural recipients of the T1DCoach application: the Diabetic Patients, their Caregivers and the T1D Educators.

Based on the literature review, own experience with similar research [21] and available hardware and software, the following research techniques were selected (Figure 2):

- scenario with the participation of users, combined with eye-tracking techniques and the think-aloud technique;

- survey technique using the SUS questionnaire;
- focused group interview technique.

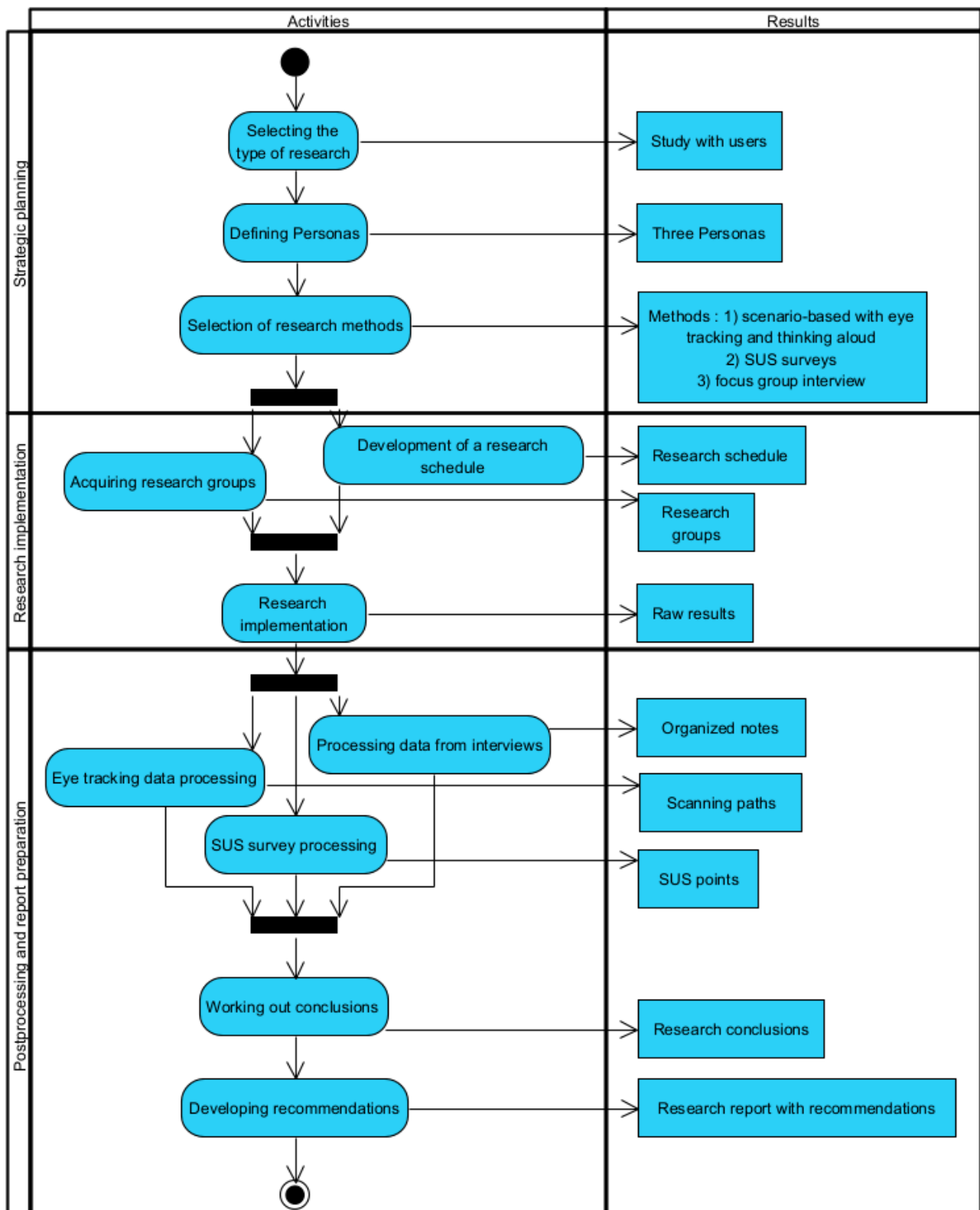


Figure 2. Study workflow.

3.3.1. Interface Testing Using Scenario Techniques

The scenario technique involves the implementation of specific actions in the system, and the scenario is the foundation of the usability test. It enables verification of important aspects of the quality of the tested application and consists of tasks that are presented to users participating in the research and that are performed by them. Proper definition of the scenario and developing the tasks according to which the tests are carried out is an important issue that is of great importance for the success of the study [37]. While scenarios represent higher-level descriptions [38], tasks define what end users should try to achieve with the application under review. Each scenario contains at least one task, but most often it includes several tasks [39]. Properly designed tasks should show the final goal and cannot be step-by-step instructions. Research participants should complete the tasks one after another. The description of the context of the task is also important because it can help users become more involved in the experiment.

3.3.2. Eye-Tracking Research

Eye-tracking research allows determination of which elements attract users' attention, for how long and in what order. These tests are non-invasive and are carried out in conditions similar to the normal use of a computer or smartphone. Research conducted using this technique uses eye trackers, i.e., devices for recording, measuring and analyzing eye movements, attention, blinks and changes in pupil diameter over time. The main measures of attention are fixations and saccades. Fixations are moments of stopping (focusing) the eyes on a specific point lasting approximately 100–300 ms. During fixation, visual information is downloaded to the brain. In turn, saccades are rapid eye movements lasting from 20 to 80 ms, associated with shifting attention to another place. Three to five saccades occur in one second. When the eyes move, no visual information is transmitted to the brain.

Fixation sequences (and their durations) create scan paths. Scan paths are visualized in the form of circles of various sizes connected by lines representing the route followed by a given user's attention while performing actions in the application interface. It has been observed that the course of scanning paths may be influenced by the tasks performed by the experiment participant and various individual factors, such as gender, knowledge or user experience [40,41]. This type of visualization is sometimes used to check the effectiveness of user searches—longer scan paths may be interpreted as less effective information search [42].

Due to the fact that the tasks performed are goal-oriented, various quantitative metrics are used in eye-tracking research. These are:

- execution time of individual scenario tasks and the entire scenario,
- the average number and average duration of fixations and the average number of blinks while performing individual scenario tasks (indicating cognitive load);
- success rate of completing individual tasks of the scenario and the entire scenario (as %).

3.3.3. Thinking-Aloud

When executing scenarios, in order to learn not only what the user is looking at, but also what he or she is thinking, the thinking-aloud technique is used [43]. This involves research participants expressing their thoughts aloud. These out-loud thoughts are recorded and then analyzed.

3.3.4. Surveys and SUS

Testing the quality of the interface is complemented by surveys that allow obtaining of users' opinions and feelings about activities with the interface. The most popular and widely used by practitioners is the SUS method.

This method is characterized by the simplicity and reliability of the measurements performed. It takes the form of a questionnaire consisting of 10 statements. It is used to quickly measure how people perceive the usefulness of the computer systems they

use [44]. The word “system” in each question/statement may be replaced by “website”, “mobile application”, “software” or “digital product”, because SUS is largely technology neutral [45]. The SUS scale is usually used immediately after using the system is assessed, without much thought and before any summary or discussion [46].

Each of the 10 statements in the SUS method is rated on a five-point Likert scale from “Strongly disagree” to “Strongly agree”, which corresponds to a numerical range from 1 to 5 [45].

As a result of processing SUS evaluation surveys, an assessment of the subjective level of satisfaction is determined for each study participant after working with the application. This rating is determined for each participant and for the entire group of respondents. The subjective level of satisfaction with working with the system for each user (j) is determined according to the formula [45]:

$$SUS_j = \left(\sum_{i=1,3,5,7,9} (S_{ij} - 1) + \sum_{i=2,4,6,8,10} (5 - S_{ij}) \right) * 2.5 \quad (1)$$

where:

i —question number from the SUS questionnaire (Q1, . . . , Q10);

S_{ij} —rating of the j -th participant of the i -th question, where the rating “definitely not” is mapped to the value 1 and “definitely yes” to the value 5 (Likert’s value).

Although SUS results (1) range from 0 to 100, they are not percentages. The overall interface rating is the arithmetical average of individual users’ ratings. After testing 500 different systems, the average SUS value (so-called industrial average) was 68 [45]. Therefore, an SUS value above 68 is above the industrial average, and below 68 is below it [45]. An SUS score of 68 is classified as acceptable.

3.3.5. Focus Group Interview

Immediately after the survey, a focus group interview is conducted with the participants. This technique belongs to moderated qualitative research and involves a directed, focused conversation with research participants about specific problems. This interview is used to better understand users’ behavior and learn about their feelings. The questions asked and discussion may revolve around positive and negative elements of the interface, as well as proposals for its modification. A group conversation on a specific topic usually involves three to twelve participants [47]. Small groups are preferred, especially when the topic requires deeper exploration and when participants have long and relevant experiences that they want to share with other group participants [48]. The course of the conversation is controlled by the moderator, whose role is to ask questions to the group and direct the discussion to desired topics and problems. A general set of questions to encourage discussion among focus group participants is prepared in advance, but questions can also be added or modified in real time as needed. The information obtained from a group interview is synergistic because participants, by presenting their feelings and observations, simultaneously learn the opinions of other people, which may influence their initial point of view.

3.4. Details of the Implementation of the Methods and Techniques Used

Based on the analysis of the functionality of the T1DCoach app, presented in its project on the Use Case Diagram, seven research scenarios were defined:

- S1.** *Measuring blood glucose using the glucometer function and reading it*
- S2.** *Measuring blood glucose using the CGM function and reading it*
- S3.** *The administration of the meal is preceded by the administration of a bolus*
- S4.** *Checking the history of meals and boluses*
- S5.** *Checking the condition and maintenance of the insulin pump*

S6. Changing the insulin pump settings**S7. Review achievements and scoreboard**

Scenarios S1–S7 were tested by experts before the actual study on a group of users (the so-called pilot). In each of the research scenarios, homogeneous tasks were separated, the implementation of which was tracked and assessed separately.

An example of scenario S3 is shown in Figure 3.

1. Check your current blood glucose reading from your glucometer.
2. Deliver a bolus using the Bolus Wizard. Specify 3.5 as the Planned Carbohydrate replacement. Enter your current blood glucose, calculate the bolus and deliver it.
3. 15 minutes after the bolus, introduce a meal consisting of: 70 grams of bread, 10 grams of Camembert cheese and 20 grams of tomato.

Figure 3. Content of scenario S3—Serving a meal preceded by a bolus.

Additionally, training scenarios were developed for the experiment and were used to pre-train users before the actual test.

For the purposes of the research, the standard SUS survey was expanded with the following questions and assessments to be indicated (but not taken into account in determining the subjective level of satisfaction with working with the system—SUS points):

1. I think the user friendliness of the application is: terrible, poor, average, good, very good.
2. Rate the difficulty/ease of using the application: very difficult, difficult, fairly easy, easy, very easy.
3. Are the functions offered by the application satisfactory? (fully satisfactory, satisfactory, difficult to determine, problematic, unsatisfactory).

For the group interview in the presented experiment, three questions were developed to initiate the discussion, which the moderator asked three groups of users of the T1DCoach application, i.e., Patients, Caregivers and Educators:

1. Is the application visually attractive?
2. Can the application be useful in educating young people with type 1 diabetes?
3. Can the intuitiveness and ergonomics of the application interface be improved? How?—Specify interface elements that should be changed/added.

The moderator wrote down the answers to these questions on a paper form, which took into account four elements: attractiveness, usefulness, intuitiveness and other comments.

3.5. Research Groups

To test the usability of the T1DCoach application interface, the activities of the most important actors in the system were planned: Patient, Caregiver and T1D Educator. Consequently, three research groups were defined for the study: Patients (G1), Caregivers (G2) and Educators (G3).

Homogeneous research groups were differentiated in terms of the following factors: role (patients vs. healthy people around them), age (adolescents vs. adults), level of

knowledge about T1D (none vs. high), education (primary vs. higher), ability to use a smartphone (high vs. low), playing computer games (yes vs. no), etc.

The size of each research group was set at 10 people according to [49]. The research participants had to deal with the tested interface of the application for the first time (apart from reading the instructions, which not all of them did). They were therefore novices to the interface being studied. However, they had substantive knowledge of T1D and related concepts.

In order to become familiar with the application and the interface, users from all three research groups were trained to perform two practice activities in the interface before the actual study. This action allows the user to become familiar with the interface and work with it. Thanks to this, the actual study simulates the work of users who have some, but not too much, experience in working with the application.

The research group was provided by the Circle of Aid to Children and Youth with Diabetes of the Lublin Regional Branch of the Society of Friends of Children.

3.6. Research Implementation

3.6.1. Preparatory Procedure

Before starting the actual research with users, the following preparatory activities were carried out:

- a detailed research schedule was defined.
- participants were recruited for the study and given instructions on how to use the application, asking them to familiarize themselves with it.

During the preparatory work for the actual study, the following tasks were carried out:

1. Research scenarios were selected and defined.
2. A research environment was prepared.
3. The tested T1DCoach software was installed on the laboratory equipment.
4. The correctness of the scenarios was verified and the deadline for their implementation was determined.

3.6.2. Research Session Procedure

The following activities were performed during each research session:

1. Familiarizing participants with the general description of the system, emphasizing their roles in the process of learning how to deal with T1D, handling a virtual patient, assessing and monitoring his health, and using the dictionary of terms.
2. Informing participants about the principles of conducting research involving humans and obtaining their consent to participate in the research.
3. Completion by the participants of a preliminary survey to determine the parameters (metrics) of the participant (gender, age, school/class, profession/number of years worked, experience in working with applications/IT systems, level of substantive knowledge regarding the management of type 1 diabetes, time of daily work with a computer/tablet/smartphone).
4. Calibration of the eye tracker (a process necessary for proper operation of the device).
5. Independent execution of scenarios by research participants (within the time limit set by experts), under the supervision of a moderator.
6. Completion of the SUS evaluation survey by research participants, allowing for the assessment of the subjective level of satisfaction of research participants with working with the application.
7. After completing the tasks by a given group of participants (i.e., after all research sessions of a given group), the moderator conducted a Focus Group Interview with the participants. The questions asked and the discussion focused on positive and negative elements of the interface, as well as proposals for its modification.

3.6.3. Place of Testing

The research was carried out in the Laboratory of Interface Ergonomics belonging to the Department of Computer Science of the Lublin University of Technology, Lublin, Poland. They took place in a specialized room, lit without shadows, intended for testing interfaces. Thanks to this, all research sessions were carried out in the same, standard conditions. During the eye-tracking test, each participant sat on a comfortable, adjustable chair, ensuring the correct position of the participant during the research—Figure 4. Two moderators took part in each of the research sessions. The first one explained the details of the study to the participants, constantly monitored the proper course of the experiment and prepared records on the computer regarding the correctness and manner of performing subsequent tasks by the research participants. The second moderator was responsible for the efficiency of the devices used for recording and acquiring data, replaced the medical lenses in eye-tracking glasses if necessary, and made sure that the participants adopted the correct posture during the tests.

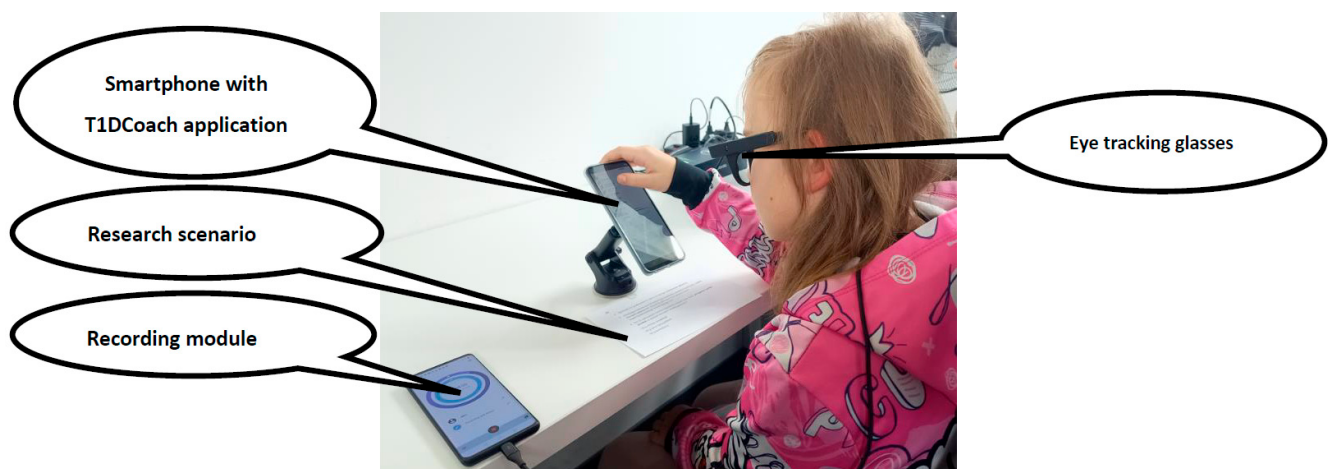


Figure 4. Test stand.

3.6.4. Equipment Used

The following devices were used in the research:

- Moto g73 5G smartphone (Motorola, Inc., Chicago, IL, USA) running Android 13 with a 6.5-inch IPS display with a resolution of 2400×1080 px. The smartphone has 8 GB of RAM and a 256 GB disk, and is equipped with a MediaTek Dimensity 930 processor, which has 8 cores. The T1DCoach application for testing purposes was installed on this device.
- Pupil Invisible mobile eye tracker (Pupil Labs, Berlin, Germany), in the form of glasses with a built-in video camera (30 Hz; resolution 1088×1080 px; visual range $82 \times 82^\circ$) for recording the scene in front of the research participant. A camera (200 Hz; 192×192 px) with an infrared illuminator is used to observe the eyes. The glasses are connected via a cable to a module that records the image of the scene and eye activity. The basic parameters of the eye tracker are sampling frequency of 200 Hz and accuracy (in ideal conditions) of 0.6° (binocular).
- Recording module—OnePlus 8 smartphone running Android with dedicated Pupil Invisible Companion software (v.3.5) installed, which, in addition to recording, also automatically sends files to the Pupil Cloud, where they can be analyzed, downloaded and subjected to further advanced analysis.
- Pupil Cloud online platform for storing, visualizing and analyzing eye-tracking data, which uses the computing power of the cloud.

3.6.5. Application Testing Session

The test studies were conducted by moderators—experts in interface research (Figure 5). Test sessions were conducted individually. A single session lasted approximately 30 min. All stages of the study were carried out during one day with participants from different research groups, because, for example, for legal reasons, patients (children/adolescents) came to the tests with their caregivers. While examining the T1DCoach application interface, participants carried out the tasks presented to them, constituting research scenarios. Each group of participants, Patients (G1), Caregivers (G2) and Educators (G3), performed the same research scenarios. The tasks were carried out in such a way that the subject sat at a desk on which, on a dedicated holder, there was a smartphone with the T1DCoach application already running. The subjects were presented with individual tasks printed on small cards. Each task was printed on a separate sheet of paper and changed on an ongoing basis by the moderator. The initial state of the application as a whole was identical for each research session (one member of the research group).



Figure 5. The application testing session with data acquisition using an eye tracker.

3.6.6. Surveys and Focus Group Interviews

The SUS survey extended with additional questions (Section 3.4) was conducted immediately after testing the application individually by each participant in the pencil-and-paper mode. Children who had trouble reading and understanding the survey were helped by their caregivers.

The focus group interview was conducted immediately after completing the survey. The interview was conducted in small groups of three to five people. This allowed participants to express their opinions and comments immediately. The groups in the interview consisted of representatives of different research groups: from G1 to G3. This was due to organizational conditions because the test studies were completed synchronously by people from different research groups (especially Patients and Caregivers, but not only).

3.6.7. Parameters of Research Implementation

Table 1 presents the workload data from the T1DCoach application quality testing process.

Table 1. Research workload.

No.	Parameter	Unit of Measurement	Value
1	Number of research sessions	pcs.	3 planned + 1 additional
2	Number of research participants	pcs.	30 (10 for each research group)
3	Research implementation time (data acquisition)	h	38
4	Labor consumption of data acquisition	man-hour	152
5	Labor consumption of post-processing raw data	man-hour	280
6	Labor consumption in developing recommendations and reports	man-hour	128
7	Report size	pages	25 (without attachments)

Interface quality testing is quite expensive. The presented study cost 560 man-hours of highly qualified researchers equipped with expensive hardware and software.

4. Results

Research group G1 (Patients) consisted of four girls and six boys. The average age was almost 10 years, weight 38 kg, height 145 cm and nearly 3 years' duration of illness.

Research group G2 (Caregivers) consisted of five women and five men. The average age in this group was just over 42 years, education: secondary, and experience in T1D—almost 3 years (which obviously correlates with the patients' data).

Research group G3 (Educators) consisted of nine women and one man (this corresponds to the demographics of the nursing profession). The average age in this group was nearly 40 years, education: higher, and experience in T1D—over 11 years.

4.1. Application Interface Testing Results Obtained Using Eye Tracking

The analysis of eye tracker data obtained during testing of the application interface allowed to determine the implementation times of individual scenarios by individual research groups—Table 2.

Table 2. Average times in seconds of scenario implementation by individual research groups.

	S1	S2	S3	S4	S5	S6	S7	Total
G1	8.3	64.0	344.4	71.1	27.0	353.4	110.3	978.4
G2	4.1	42.8	292.5	73.1	16.5	314.1	54.6	797.7
G3	2.9	54.0	308.5	52.9	26.4	217.8	51.4	714.0

Similarly, the processing of raw data allowed for obtaining of data on the average number of fixations during the implementation of scenarios by individual research groups—Table 3.

Table 3. The average number of fixations during the implementation of scenarios by individual research groups.

	S1	S2	S3	S4	S5	S6	S7	Total
G1	19.4	178.1	914.0	209.2	87.3	1007.9	320.0	2735.9
G2	10.8	121.2	803.9	234.0	43.4	896.8	166.4	2276.4
G3	8.4	166.6	1008.3	150.1	84.6	830.2	146.1	2394.3

The values of the next metric—average fixation duration—are presented in Table 4.

Table 4. Average fixation time when implementing scenarios by individual research groups.

	S1	S2	S3	S4	S5	S6	S7	Average
G1	350.6	318.6	302.2	292.7	236.4	319.5	268.2	298.3
G2	365.4	301.7	306.3	260.0	285.4	287.7	281.3	298.2
G3	224.8	196.2	187.8	198.1	181.3	214.9	186.7	198.5

The values of the next metric—the average number of blinks—are presented in Table 5.

Table 5. The average number of blinks while implementing the scenarios by individual research groups.

	S1	S2	S3	S4	S5	S6	S7	Total
G1	2.6	7.2	54.5	12.3	4.3	63.9	13.4	158.2
G2	0.5	5.5	40.4	13.0	1.9	50.6	7.8	119.7
G3	1.0	15.9	85.7	18.3	6.6	62.4	12.8	202.7

The research showed different effectiveness of the implementation of individual scenarios by individual research groups—Table 6.

Table 6. The average success rate in % of scenario implementation by individual research groups.

	S1	S2	S3	S4	S5	S6	S7	Average
G1	100.0	50.0	56.7	60.0	85.0	27.5	25.0	57.7
G2	100.0	100.0	66.7	80.0	95.0	70.0	55.0	81.0
G3	100.0	90.0	80.0	75.0	100.0	85.0	60.0	84.3

During the research, nearly 200 videos were obtained showing various interface scanning paths by users during application tests. A small part of the records (approx. 5%) were lost due to problems with saving them in the cloud. They were used for qualitative analysis of eye-tracking data. Their examples are presented in Figure 6.

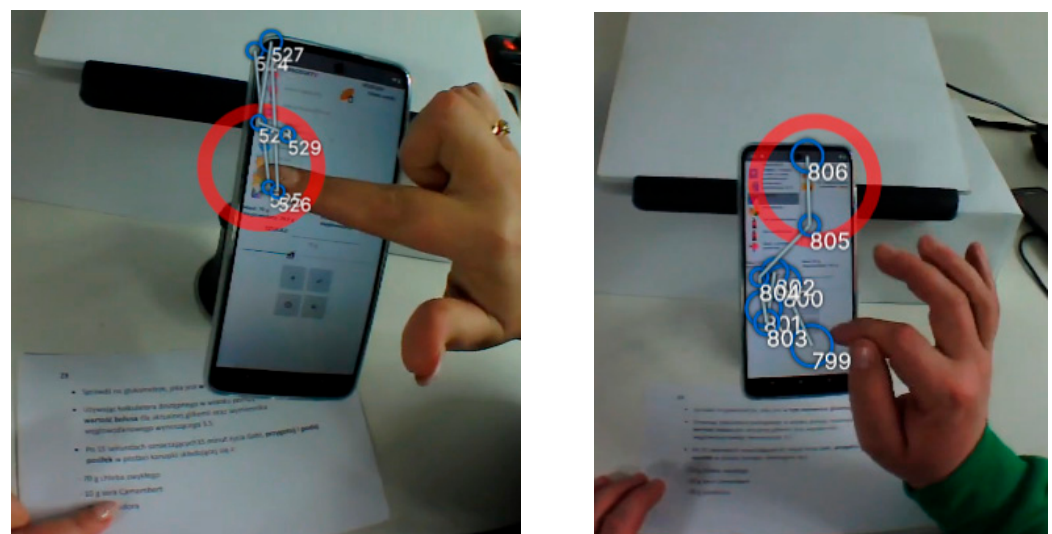


Figure 6. Examples of scan paths during scenario implementation (descriptions in the text).

The first example (Figure 6 on the left) shows the user's difficulty in finding a product from a long product list. This problem is the result of not using the available search engine, which makes searching for ingredients cumbersome and time-consuming.

The second example (Figure 6 on the right) shows a common problem when composing meals. After selecting the next product and determining its weight, the user does not add it to the meal list and, as a result, the served meal does not contain the last product.

4.2. Results of Survey Research

The results of research using the SUS survey are presented in Tables 7–9 (red colored numbers will be discussed in the Discussion section). These tables have the following designations: Q1, . . . , Q10—numbers of questions in the SUS survey; P1, . . . , P10—numbers of participants in the research group. Results below the industry average (i.e., 68) are highlighted in red.

Statistical results of answers to additional questions (attached to the SUS survey), presented in Section 3.4 are presented in Table 10.

4.3. Focus Group Interviews Results

Participants in the focus group interview repeatedly emphasized that:

- The application interface is “clear”, “simple”, “minimalist”, “understandable”, “intuitive” (in a positive sense).

- The avatar used, its animation, colors and animation of its behavior (GABI) are positively associated with type 1 diabetes and its symptoms for the patient.
- Due to the psychology of adolescent boys, it was proposed to implement a male avatar in the application: e.g., MIKI. The avatar should be selected when initially selecting the parameters of the digital twin for simulation (gender, body weight, age, etc.).

Table 7. Results of the SUS survey for patients (G1).

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS
P1	4	1	5	2	3	1	4	1	5	2	85.0
P2	5	3	4	5	4	4	4	3	3	3	55.0
P3	5	3	4	3	4	3	4	2	4	4	65.0
P4	4	3	3	2	4	1	4	1	4	2	75.0
P5	5	3	5	1	4	1	4	1	5	1	90.0
P6	2	4	3	2	5	4	4	5	2	1	50.0
P7	2	3	4	2	4	3	5	2	3	3	62.5
P8	2	1	5	2	3	2	5	1	4	2	77.5
P9	4	1	2	4	5	1	2	1	5	3	70.0
P10	4	2	4	2	4	2	5	2	2	1	75.0
										max	90.0
										min	50.0
										avg	70.5
										deviaton	12.6

Note: red numbers will discuss in Section 5.2.

Table 8. Results of the SUS survey for Caregivers (G2).

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS
P1	4	1	4	1	4	2	5	1	4	2	85.0
P2	5	2	4	2	4	1	5	1	4	1	87.5
P3	5	4	4	2	4	2	4	1	4	2	75.0
P4	5	1	3	3	4	3	4	3	3	3	65.0
P5	5	2	4	1	4	2	5	1	4	1	87.5
P6	4	2	4	1	5	2	5	1	4	2	85.0
P7	3	2	2	3	1	4	2	4	2	2	37.5
P8	3	2	4	1	3	3	1	2	2	2	57.5
P9	3	4	2	5	3	2	3	2	2	5	37.5
P10	4	2	4	1	4	2	4	2	4	1	80.0
										max	87.5
										min	37.5
										avg	69.8
										deviaton	19.7

Note: red numbers will discuss in Section 5.2.

Table 9. Results of the SUS survey for Educators (G3).

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS
P1	5	2	4	2	4	3	5	2	4	3	75.0
P2	5	1	4	1	5	2	5	1	4	2	90.0
P3	5	1	4	1	5	1	5	1	4	1	95.0
P4	4	2	4	1	4	2	5	1	2	2	77.5
P5	4	2	3	2	4	3	4	2	3	2	67.5
P6	3	2	4	1	3	2	4	3	4	1	72.5
P7	5	1	4	1	5	1	5	1	4	1	95.0
P8	5	1	4	1	4	2	5	1	4	2	87.5
P9	5	1	4	3	4	4	4	2	3	2	70.0
P10	4	1	4	1	4	2	4	1	5	2	85.0
										max	95.0
										min	67.5
										avg	81.5
										deviaton	10.3

Note: red numbers will discuss in Section 5.2.

Table 10. Score values for additional questions for groups G1, G2 and G3 (*).

Values	G1—Patients			G2—Caregivers			G3—Educators		
	Friendliness	Ease	Satisfaction	Friendliness	Ease	Satisfaction	Friendliness	Ease	Satisfaction
Max	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Min	3.0	3.0	1.0	3.0	2.0	3.0	3.0	3.0	3.0
Avg	4.2	3.4	3.3	4.3	3.5	4.0	4.3	3.5	4.2
Deviation	0.7	0.7	1.1	0.7	1.1	0.8	0.7	0.7	0.6

(*) on a scale: 1—very small, ..., 5—very big.

Interview participants highly assessed the usefulness of the T1DCoach application in the practice of educators, caregivers and patients, emphasizing:

- friendliness and accessibility of the interface;
- a very interesting and useful idea;
- the application is very useful for teaching medical students.

Interview participants also highly appreciated the intuitiveness and ergonomics of the T1DCoach application in the practice of type 1 diabetes education, pointing out the following:

- easy to use without training—immediate use;
- use of familiar, common controls;
- correct and specialized vocabulary of the application;
- no runtime errors.

5. Discussion

5.1. Application Testing

The application testing results, presented in Table 2, indicate that the total average implementation time of all scenarios in the Patient group (G1) was the longest and amounted to 16 min and 18 s. Caregivers (G2) completed the scenarios in an average time of 13 min and 18 s, while Educators (G3) took 11 min and 54 s. This shows that the most efficient users of the T1DCoach application were Educators.

The results presented in Tables 3–5 concern metrics that determine the cognitive effort put in by research participants when implementing the scenarios. In the case of the number of fixations and fixation duration, higher levels of these indicators indicate greater cognitive effort. The number of blinks is interpreted differently—the lower the number of blinks, the greater the cognitive effort.

The results regarding the average number of fixations (Table 3) for all scenarios show that the most fixations occurred in the group of patients (G1)—2735.9. In the Educators group (G3), there were an average of 2394.3 of these. The fewest fixations occurred in the Caregiver group (G2)—an average of 2276.4.

The average fixation duration (Table 4) during the implementation of all scenarios was almost the same in the Patient (G1) and Caregiver (G2) groups and amounted to 298.3 ms and 292.2 ms, respectively. Much shorter fixations occurred in Educators (G3) and lasted on average 198.5 ms.

In turn, the average number of blinks (Table 5) during the implementation of the scenarios was 158.2 in the group of patients (G1). There were 119.7 of them in the Caregiver group (G2). The highest number of blinks, i.e., 202.7, occurred in the Educators group (G3).

The average success rate of scenario implementation (Table 6) was clearly lower in the group of patients (G1) than in the groups of caregivers (G2) and educators (G3) and amounted to 57.7%, 81.0% and 84.3%, respectively, for the individual groups. All participants had the greatest problems with completing the tasks in scenario S6 (changing insulin pump settings) and in scenario S7 (reviewing achievements and the scoreboard). In the Patients group (G1), the average success rate of these scenarios was 27.5% and 25.0% for S6 and S7, respectively. This result was much lower than in the groups of Caregivers (G2) and Educators (G3). This result is consistent with the practice of T1D treatment—children do not change the insulin pump settings themselves due to the high complexity and responsibility of this process [50].

The results of thinking aloud allowed us to identify three groups of problems: critical, important and unimportant.

Critical problems include, among others: difficulties in detecting the side menu, difficulties in finding how to set the weight of products, unclear distinction between the “Add” (+) and “Confirm” (v) buttons, problems with the visibility of hidden products in the “Food History” and lack of awareness of having a username, difficulties with identifying user’s badges and the unintuitive way of presenting them in the “Champions Board”.

Significant problems included the use of identical icons for various functions, too high sensitivity of sliders and their location too close to the edge of the smartphone screen, inability to distinguish between an insulin base and a temporary base, and an invisible top menu and its options.

In turn, minor problems included problems with finding the latest meal/bolus entries located at the end of the list, failure to move to the previous/next day in CGM and difficulties in finding the list of players (use of the name “Champions’ Board”).

As a result of the analysis of the think-aloud records, the following conclusions and recommendations were formulated:

- adding hints and prompts for confirming important operations;
- changing the appearance and location of the slider;
- enabling weight entry from the keyboard;
- improving the visibility and intuitiveness of buttons;
- changing the order in the history of boluses/meals so that the latest entries are at the beginning (top) of the list;
- adding clear separation of temporary base settings from insulin base settings;
- rebuilding the “Champions Board” so that the user can clearly see himself on the list and his badges will be displayed automatically.

5.2. Survey Research

The analysis of the SUS surveys indicates that all research groups showed a subjective level of satisfaction higher than the industrial average, i.e., 68. The quality of the T1DCoach application interface was best assessed by Educators (G3)—81.5 (Table 9), followed by Patients (G1)—70.5 (Table 7), and the lowest among Caregivers (G2)—69.8 (Table 8). The average SUS value for all study groups was 73.9, which is significantly above the benchmark: 68.

Additional ratings (apart from questions from the SUS methodology) were also higher than average (on a Likert scale of 1–5, the average is 2.5) and amounted to over 3.3 for Friendliness, Ease of use, and Satisfaction (Table 10). These ratings ranged from 3.3 to 4.3.

It should be emphasized that, despite the overall very positive ratings of the interface quality in SUS surveys with proprietary extensions, some ratings do not exceed the industry standard, i.e., 68 (Tables 7–9—red colors). A detailed analysis suggests the following reasons for such assessments: poor knowledge of information technology, problems with using mobile applications, and insufficient knowledge of diabetes problems and insulin pump applications, as well as lack of sufficient reading comprehension skills (children).

5.3. Focus Group Interviews

Users' statements during interviews confirmed the need for development of the application, its usefulness and positive feelings towards its interface. They also allowed to obtain a number of comments and recommendations, the most important of which was the need to implement a male avatar.

6. Conclusions

The results of eye-tracking studies showed how the target users of the T1DCoach application, divided into three groups, dealt with using the T1DCoach application: Patients (G1), Caregivers (G2) and Educators (G3). Even though, in order to increase the effectiveness of learning how to use the application, dedicated instructions were prepared for each user group that show the application's capabilities, people participating in usability tests obtained different results. It turned out that the best achievements were achieved by Educators, who performed the tasks included in the research scenarios the fastest, with the least cognitive effort and with the fewest errors. In turn, Patients (G1) achieved the poorest results and therefore had the greatest problems with implementing the scenarios. It can therefore be assumed that the good results for Educators were related to their extensive knowledge and experience of diabetes and high digital competences. The poorer results for young Patients (G1) in using the application were most likely caused by lower skills in coping with the disease. It is worth noting that computer skills facilitating the use of the application also varied depending on age. They were higher in the older group of patients.

The basic conclusion from the SUS examination of the interface quality of the T1DCoach mobile application is that the T1DCoach application interface exceeds the quality of the industrial standard by over 8%. The high quality of the interface was therefore confirmed in the study. Similar conclusions can be drawn from focus group interviews.

In addition to cosmetic changes to the interface, the research put forward the following recommendations:

- implement a male avatar (done in the form of MIKI—Figure 7);
- do not limit the values entered into the application during validation (towards positive values) (applies to bolus, food, etc.)—this is to teach responsibility in everyday life (that requirement was not provided by the application developers);
- highlight the existence of a side menu;
- improve the visibility of meals in the Food History;
- lower the sensitivity of the sliders and add windows for entering values;
- change the order of displaying historical data so that the newest are at the top of the list.

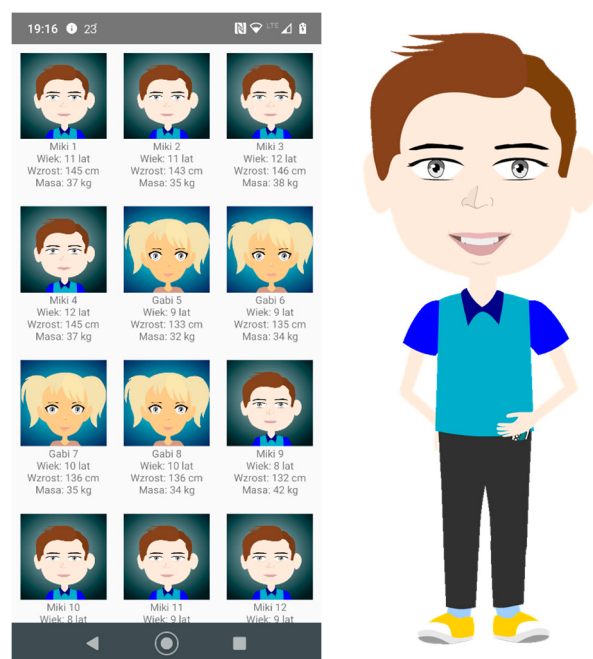


Figure 7. Avatar selection screen with an implemented male avatar of a T1D patient (on the left) and the full form of MIKI (on the right)—implementation of one of the recommendations.

The recommendations were introduced into the application, which improved the quality of its interface. This was confirmed in subsequent acceptance tests of the application in real conditions.

Author Contributions: Conceptualization, M.M.; methodology, M.M., M.P.-W. and M.D.; software, M.D.; validation, M.M., M.P.-W. and M.D.; formal analysis, M.M.; investigation, M.M., M.P.-W. and M.D.; resources, M.M.; data curation, M.D.; writing—original draft preparation, M.M., M.P.-W. and M.D.; writing—review and editing, M.M., M.P.-W. and M.D.; visualization, M.D. and M.M.; supervision, M.M.; project administration, M.M.; funding acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Polish National Center for Research and Development under the *Things are for people* program (Grant No. Rzecz są dla ludzi/0062/2020-00).

Institutional Review Board Statement: The research procedure and documentation received a positive opinion from the Scientific Research Ethics Committee of the Lublin University of Technology No. 1/2022 of 13 April 2022.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Haze, K.A.; Lynaugh, J. Building patient relationships: A smartphone application supporting communication between teenagers with asthma and the RN care coordinator. *CIN: Computers, Informatics. Nursing* **2013**, *31*, 266–271. [\[CrossRef\]](#) [\[PubMed\]](#)
- Herrmann, S.; Power, B.; Rashidi, A.; Cypher, M.; Mastaglia, F.; Grace, A.; McKinnon, E.; Sarrot, P.; Michau, C.; Skinner, M.; et al. Supporting patient-clinician interaction in chronic HIV care: Design and development of a patient-reported outcomes software application. *J. Med. Internet Res.* **2021**, *23*, e27861. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lambrech, A.; Vuillerme, N.; Raab, C.; Simon, D.; Messner, E.M.; Hagen, M.; Bayat, S.; Kleyer, A.; Aubourg, T.; Schett, G.; et al. Quality of a supporting mobile app for rheumatic patients: Patient-based assessment using the user version of the mobile application scale (uMARS). *Front. Med.* **2021**, *8*, 715345. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nowicki, T.; Nowicki, G.; Ślusarska, B. Advancement Level of Mobile Applications Intended for Type 1 Diabetes Therapy Supporting. *J. Educ. Health Sport* **2020**, *10*, 704–713. [\[CrossRef\]](#)

5. Feingold-Polak, R.; Barzel, O.; Levy-Tzedek, S. A robot goes to rehab: A novel gamified system for long-term stroke rehabilitation using a socially assistive robot—Methodology and usability testing. *J. NeuroEng. Rehabil.* **2021**, *18*, 122. [\[CrossRef\]](#)
6. Villamañe, M.; Alvarez, A. Facilitating and automating usability testing of educational technologies. *Comput. Appl. Eng. Educ.* **2024**, *32*, e22725. [\[CrossRef\]](#)
7. Als, B.S.; Jensen, J.J.; Skov, M.B. Comparison of think-aloud and constructive interaction in usability testing with children. In Proceedings of the 2005 Conference on Interaction Design and Children, Boulder, CO, USA, 8–10 June 2005; Association for Computing Machinery: New York, NY, USA, 2005; pp. 9–16. [\[CrossRef\]](#)
8. Mazumder, F.K.; Das, U.K. Usability guidelines for usable user interface. *Int. J. Res. Eng. Technol.* **2014**, *3*, 79–82. [\[CrossRef\]](#)
9. Park, K.S.; Lim, C.H. A structured methodology for comparative evaluation of user interface designs using usability criteria and measures. *Int. J. Ind. Ergon.* **1999**, *23*, 379–389. [\[CrossRef\]](#)
10. Riihiahio, S. Usability testing. In *The Wiley Handbook of Human Computer Interaction*; Norman, K.L., Kirakowski, J., Eds.; John Wiley & Sons Ltd.: New York, NY, USA, 2018; Volume 1, pp. 255–275. [\[CrossRef\]](#)
11. Eysenbach, G.; Köhler, C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ* **2002**, *324*, 573–577. [\[CrossRef\]](#)
12. Nasir, M.; Ikram, N.; Jalil, Z. Usability inspection: Novice crowd inspectors versus expert. *J. Syst. Softw.* **2022**, *183*, 111122. [\[CrossRef\]](#)
13. Trukenbrod, A.K.; Backhaus, N.; Thomaschke, R. Measuring subjectively experienced time in usability and user experience testing scenarios. *Int. J. Hum. Comput. Stud.* **2020**, *138*, 102399. [\[CrossRef\]](#)
14. Weichbroth, P. Usability of mobile applications: A systematic literature study. *IEEE Access* **2020**, *8*, 55563–55577. [\[CrossRef\]](#)
15. Zardari, B.A.; Hussain, Z.; Arain, A.A.; Rizvi, W.H.; Vighio, M.S. QUEST e-learning portal: Applying heuristic evaluation, usability testing and eye tracking. *Univers. Access Inf. Soc.* **2021**, *20*, 531–543. [\[CrossRef\]](#)
16. Pratama, I.W.; Sudarsana, E.S.; CahayavWidiyanto, A.A. Exploring Two Methods of Usability Testing: System Usability Scale And Retrospective Think-Aloud. *J. Akad. Vokasi* **2023**, *2*, 33–43. [\[CrossRef\]](#)
17. Kamińska, D.; Zwoliński, G.; Laska-Leśniewicz, A. Usability testing of virtual reality applications—The pilot study. *Sensors* **2022**, *22*, 1342. [\[CrossRef\]](#)
18. Kumar, B.A.; Chand, S.S.; Goundar, M.S. Usability testing of mobile learning applications: A systematic mapping study. *Int. J. Inf. Learn. Technol.* **2024**, *41*, 113–129. [\[CrossRef\]](#)
19. Prokopia, V.; Tselios, N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *J. Res. Technol. Educ.* **2022**, *54*, 392–409. [\[CrossRef\]](#)
20. Wronikowska, M.W.; Malycha, J.; Morgan, L.J.; Westgate, V.; Petrinic, T.; Young, J.D.; Watkinson, P.J. Systematic review of applied usability metrics within usability evaluation methods for hospital electronic healthcare record systems: Metrics and Evaluation Methods for eHealth Systems. *J. Eval. Clin. Pract.* **2021**, *27*, 1403–1416. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Borys, M.; Milosz, M. Mobile application usability testing in quasi-real conditions—The synergy of using different methods. 11th International Conference on Human System Interaction, HSI2018, Gdansk, Poland, 4–6 July 2018; pp. 362–368. [\[CrossRef\]](#)
22. Akmal Muhamat, N.; Hasan, R.; Saddki, N.; Mohd Arshad, M.R.; Ahmad, M. Development and usability testing of mobile application on diet and oral health. *PLoS ONE* **2021**, *16*, e0257035. [\[CrossRef\]](#)
23. Setiyawati, N.; Purnomo, H.D.; Mailoa, E. User Experience Design on Visualization of Mobile-Based Land Monitoring System Using a User-Centered Design Approach. *Int. J. Interact. Mob. Technol.* **2022**, *16*, 47–65. [\[CrossRef\]](#)
24. Nugroho, A.; Santosa, P.I.; Hartanto, R. Usability evaluation methods of mobile applications: A systematic literature review. In Proceedings of the 2022 International Symposium on Information Technology and Digital Innovation (ISITDI), Padang, Indonesia, 27–28 July 2022; pp. 92–95. [\[CrossRef\]](#)
25. Khajouei, R.; Zahiri Esfahani, M.; Jahani, Y. Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 55–60. [\[CrossRef\]](#)
26. Farzandipour, M.; Nabovati, E.; Tadayon, H.; Jabali, M.S. Usability evaluation of a nursing information system by applying cognitive walkthrough method. *Int. J. Med. Inform.* **2021**, *152*, 104459. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Farzandipour, M.; Nabovati, E.; Sadeqi Jabali, M. Comparison of usability evaluation methods for a health information system: Heuristic evaluation versus cognitive walkthrough method. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 157. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Milosz, M.; Plechawska-Wójcik, M.; Borys, M.; Laskowski, M. Quality improvement of ERP system GUI using expert method: A case study. In Proceedings of the 6th International Conference on Human System Interaction, HSI2013, Sopot, Poland, 6–8 June 2013; pp. 145–152. [\[CrossRef\]](#)
29. Hyzy, M.; Bond, R.; Mulvenna, M.; Bai, L.; Dix, A.; Leigh, S.; Hunt, S. System usability scale benchmarking for digital health apps: Meta-analysis. *JMIR Mhealth Uhealth* **2022**, *10*, e37290. [\[CrossRef\]](#)
30. Novák, J.Š.; Masner, J.; Benda, P.; Šimek, P.; Merunka, V. Eye tracking, usability, and user experience: A systematic review. *Int. J. Hum. Comput. Interact.* **2023**, 1–17. [\[CrossRef\]](#)
31. Julian, I.; Murad, D.F.; Riva'i, R.Y. Combining UEQ and Eye-Tracking Method as Usability Evaluation for Mobile Apps. In Proceedings of the 2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS), Makassar, Indonesia, 25–26 October 2021; pp. 1–6. [\[CrossRef\]](#)
32. Țichindelean, M.; Țichindelean, M.T.; Cetină, I.; Orzan, G. A comparative eye tracking study of usability—Towards sustainable web design. *Sustainability* **2021**, *13*, 10415. [\[CrossRef\]](#)

33. Oyekunle, R.; Bello, O.; Jubril, Q.; Sikiru, I.; Balogun, A. Usability evaluation using eye-tracking on E-commerce and education domains. *J. Inf. Technol. Comput.* **2020**, *1*, 1–13. [\[CrossRef\]](#)
34. Aiyegbusi, O.L. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. *Qual. Life Res.* **2020**, *29*, 325–333. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Milosz, M.; Chmielewska, M. Usability Testing of e-Government Online Services Using Different Methods—A Case Study. In Proceedings of the 13th International Conference on Human System Interaction, HSI2020, Tokyo, Japan, 6–8 June 2020; pp. 1–5. [\[CrossRef\]](#)
36. Nowicki, T. Virtual therapy using Type 1 Diabetes Direct Simulator. *J. Phys. Conf. Ser.* **2021**, *1736*, 012031. [\[CrossRef\]](#)
37. Russ, A.L.; Saleem, J.J. Ten factors to consider when developing usability scenarios and tasks for health information technology. *J. Biomed. Inform.* **2018**, *78*, 123–133. [\[CrossRef\]](#)
38. Lowry, S.Z.; Quinn, M.T.; Ramaiah, M.; Schumacher, R.M.; Patterson, E.S.; North, R.; Zhang, J.; Gibbons, M.C.; Abbott, P. (NISTIR 7804) Technical Evaluation National Institute of Standards and Technology, Testing and Validation of the Usability of Electronic Health Records. In *NIST Interagency/Internal Report (NISTIR) 7804*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2012. [\[CrossRef\]](#)
39. Vincent, C.J.; Blandford, A. Usability standards meet scenario-based design: Challenges and opportunities. *J. Biomed. Inform.* **2015**, *53*, 243–250. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Eraslan, S.; Yesilada, Y. Patterns in Eyetracking Scanpaths and the Affecting Factors. *J. Web Eng.* **2015**, *14*, 363–385.
41. Underwood, G.; Humphrey, K.; Foulsham, T. Knowledge-Based Patterns of Remembering: Eye Movement Scanpaths Reflect Domain Experience. In *HCI and Usability for Education and Work*; Holzinger, A., Ed.; Springer Berlin: Heidelberg/Berlin, Germany, 2008; Volume 5298, pp. 125–144. [\[CrossRef\]](#)
42. Ehmke, C.; Wilson, S. Identifying Web Usability Problems from Eye-Tracking Data. In Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It, Swinton, UK, 3–7 September 2007; British Computer Society: Swinton, UK, 2007; pp. 119–128. [\[CrossRef\]](#)
43. Lewis, J.R. Usability Testing. In *Handbook of Human Factors and Ergonomics*, 3rd ed.; Salvendy, G., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006; p. 1282. [\[CrossRef\]](#)
44. Tullis, T.S.; Stetson, J.N. A comparison of questionnaires for assessing website usability. In Proceedings of the UPA: Usability Professionals' Association Conference, Minneapolis, MN, USA, 7–11 June 2004; pp. 1–12.
45. Brooke, J. SUS: A Retrospective. *J. Usability Stud.* **2013**, *8*, 29–40.
46. Brooke, J. SUS: A “quick and dirty” usability scale. In *Usability Evaluation in Industry*; Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L., Eds.; Taylor and Francis: London, UK, 1996. [\[CrossRef\]](#)
47. Morgan, D.L. *Focus Groups as Qualitative Research*, 2nd ed.; Sage: Thousand Oaks, CA, USA, 1997. [\[CrossRef\]](#)
48. Langford, J.; McDonagh, D. *Focus Groups. Supporting Effective Product Development*; Taylor and Francis: London, UK, 2003. [\[CrossRef\]](#)
49. Macefield, R. How to specify the participant group size for usability studies: A practitioner's guide. *J. Usability Stud.* **2009**, *5*, 34–45.
50. Nimri, R.; Oron, T.; Muller, I.; Kraljevic, I.; Alonso, M.M.; Keskinen, P.; Milicic, T.; Oren, A.; Christoforidis, A.; Den Brinker, M.; et al. Adjustment of Insulin Pump Settings in Type 1 Diabetes Management: Advisor Pro Device Compared to Physicians' Recommendations. *J. Diabetes Sci. Technol.* **2022**, *16*, 364–372. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.