Usability Study of Mobile Applications with Cognitive Load Resulting from Environmental Factors

Beata Karczewska, Elżbieta Kukla, Patient Zihisire Muke, Zbigniew Telec, and Bogdan Trawiński(✉)

Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wrocław, Poland
{elzbieta.kukla,patient.zihisire,zbigniew.telec, bogdan.trawinski}@pwr.edu.pl

Abstract. The main goal of this work was to design and conduct usability tests of mobile applications, with particular emphasis on cognitive load. Not only was the cognitive load due to user interface design errors considered, but also the cognitive load due to external factors such as the real world. The study involved 12 participants, who were young adults, i.e. people aged 20–30 who regularly use mobile devices. The participants were randomly divided into two groups of 6 people. These groups independently carried out the scenario of activities using two different applications, in two different environments: laboratory and field. The collected results were statistically analysed in terms of task completion time, number of actions, number of errors and satisfaction with regard to net promoters score and system usability scale. The analysis of these features allowed to determine the relationship between the quality of the graphical user interface, the impact of the environmental factors and user satisfaction.

Keywords: Usability testing · Cognitive load · Mobile applications · Environmental factors · Net promoter score · System usability scale

1 Introduction

Modern technological development puts more and more emphasis on mobility and adaptation to the constant movement of users. Programmers and application developers constantly forget about the need to adapt their products to people who use them in constant motion. As a result, when designing, you should not only remember about the standard aspects of usability, but also put more and more emphasis on the support of low resolutions, limitations in connectivity and high battery consumption. The critical thing here is to consider the context in which the applications are used.

Although technological and design advances can be seen in terms of the design of application graphical interfaces, their navigation and adaptation to mobility, many companies and application developers still do not take into account the aspect of the

N. T. Nguyen et al. (Eds.): ACIIDS 2021, LNAI 12672, pp. 851–864, 2021.
https://doi.org/10.1007/978-3-030-73280-6_67

inconvenience of the external environment. Overloading a sensory stimulus onto a person is called cognitive overload.

This term comes from psychology and assumes that the human mind, and more precisely memory, is able to process only a certain number of tasks [1]. After overloading memory with external factors, a person will have a difficult action, and his decision-making will also be disturbed. The theory dealing with cognitive load is still criticized in academia for the lack of conceptual clarity and the possibility of a methodological approach to the problem [2].

The main goal of this work is to test the usability of selected mobile applications in terms of cognitive load resulting from environmental factors. For this purpose, adequate user testing was carried out. The research was divided into 4 series, each of which assumed a different combination of tools and environments. So far, the authors of this paper have conducted a number of studies on the usability of mobile applications, including responsive web design [3–5] and data entry design patterns [6, 7].

2 Related Works

In the field of mobile devices, technological advances have led to the widespread use of smartphones to the detriment of desktop computers [8]. However, there are still many limitations and problems with the usability of the above-mentioned devices that affect the cognitive load of users depending on the environment in which they work.

Within the framework of software engineering, the usability of a system plays a key role while defining the perceived quality of the use of its users [9, 10]. In this context, usability is defined as a study of the intersection between users and systems, tasks to be performed as well as expectations in terms of use.

Harrison et al. [11] designed the People At the Center of Mobile Application Development (*PACMAD*) usability model developed to tackle existing usability models' challengers when it comes to mobile devices. The model comprises seven attributes demonstrating the application's usability: effectiveness, efficiency, satisfaction, learnability, memorability, errors and cognitive load. The innovation of the model concerns cognitive load as a new usability metric.

Mobile applications' usability model *PACMAD* includes cognitive load because it can directly affect and can be affected by the usability of the application. Indeed, it seems likely that mobile devices are particularly susceptible to the effects of cognitive overload due to their multiple configurations of tasks to be performed and size limitations [11].

Ejaz et al. [12] have argued that cognitive load is not only a supplementary attribute of usability, but one of main attributes when it comes to emergency scenarios. As a result, the authors made experimental comments and conclusions during the usability testing demonstrating the importance of cognitive load.

In context of mobile devices, cognitive load is related to the mental effort needed by a user to carry out tasks while utilizing a mobile device. Although it not a new concept, and does not rank first in usability research, but it is now gaining more popularity in the usability field due to the fact that the attention of users is usually divided between tasks which are being performed simultaneously [8].

Furthermore, Parente Da Costa et al. [10] provide another additional contribution by proposing a set of usability heuristics related to the software applications used in smartphones, taking into consideration the user, task and the context as usability factors and cognitive load as a crucial attribute of usability.

As mentioned before, the *PACMAD* usability model was developed to tackle the challengers of existing usability models such as Nielsen, Mobile usability heuristic and the ISO 9241-11 standard when it comes to mobile devices as shown in Table 1 [13].

Table 1. Models and usability metrics applied for mobile applications. Source [13]

Nielsen	Mobile usability heuristic	ISO 9241-11	PACMAD
Learnability	Findability and visibility of system status	Effectiveness	Effectiveness
Efficiency of use	Real word and match between system	Efficiency	Efficiency
Error frequency	Minimalist design and good ergonomics		Learnability
Memorability	Mapping and Consistency	Satisfaction	Satisfaction
Satisfaction	Screen readability, ease in input		Error
	Personalization and flexibility, efficient of use		Cognitive load
	Social conventions and aesthetic, privacy		
	Realistic error management		

In addition, the concept cognitive load is influenced by the characteristics of the subject, the characteristics of the task and the interactions between the two [14]. The characteristics of the task may include the time pressure, difficulty level of the task, the newness of the task, remuneration after performing the task, and the environment in which the task is being performed. In the other hand, the characteristics of the subject refer to stable factors that are unlikely to change with the task or environment, like the subjects' previous knowledge, preferences, cognitive capabilities and cognitive style and [15].

In that regard, it should be also highlighted that, the level of cognitive load of a participant who is taking part in a cognitive load experiment is primarily caused by the task complexity. But in reality, cognitive load is influenced by several factors like environmental disturbances, time pressure and task complexity [16]. In the same context, Ferreira et al. [17] argued that the actual magnitude of cognitive load that a person experiences is influenced by the environmental and social factors, individual differences and the tasks being performed.

3 Setup of Usability Experiments

3.1 Selection of Mobile Applications

In order to study the most important aspects of the usability of mobile applications with regard to cognitive load, the key is to select the appropriate applications. The aim of the study was to examine the influence of external factors on the efficiency of using the application and the positive reception of the application as well as the cognitive load caused by deficiencies in the design of the graphical user interface. It was important to study various types of actions ranging from simple and obvious activities and paths to follow to complicated activities such as entering data, preparing and making decisions, performing calculations.

For this reason, mobile home budget management applications seem to be particularly useful for the purposes of our research. In this type of apps, the key issue is the good design of human-computer interaction, due to the specificity of activities. There are many different actions available in the home budget applications, including categorizing, planning, entering and reporting expenses. This makes it necessary to simplify the action for the user. Moreover, there are many applications on the market that have taken up the topic of interface simplification too eagerly, resulting in unclear messages and the arrangement of action buttons.

Two home budget management applications available in the *Google Store* were selected for the research described in this article, differing from each other in terms of market maturity, popularity among users and the perceived usefulness of the graphical interface. These were the following mobile applications: "*Fast Budget - Expense & Money Manager*" [18] and "*Family Budget*" [19]. The former had over 1 million downloads, about 92 thousand ratings with an average rating of 4.6 on a 5-point scale. We will denote it as *App1* in the further part of the paper. In contrast, the latter had fewer than 1,000 downloads, few ratings and comments, and a smaller range of functionality and lower perceived ease of use. This application in the rest of the paper will be marked as *App2*. Sample screen shots illustrating the graphical user interfaces of *App1* and *App2* are shown in Figs. 1 and 2, respectively.

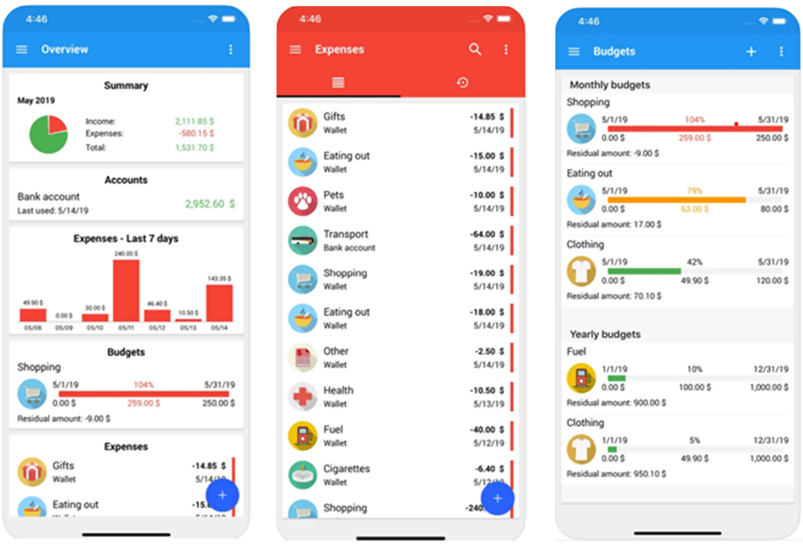

Fig. 1. Sample screenshots of the *Fast Budget* application (*App1*). Source [13].

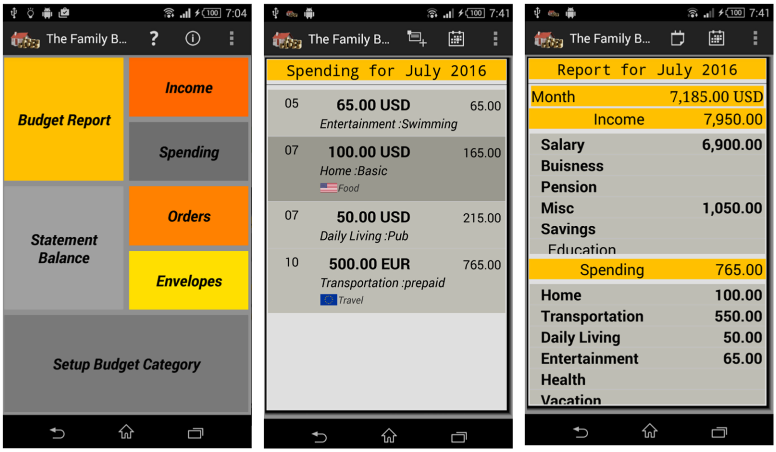

Fig. 2. Sample screenshots of the *Family Budget* application (*App2*). Source [14].

3.2 Environments and Groups of Participants

To adequately investigate the effect of cognitive load on the usability of selected mobile applications, the study was conducted in two different environments. The first was a laboratory environment, i.e. tests were conducted in a room where external stimuli were reduced [20]. The participants were able to use the apps without problems related to weather, noise, and other external stimuli.

The second environment was field-based, i.e. tests were carried out in the real, outdoor world. This maximized the cognitive load affecting the user and more accurately reflect the use of the app in natural settings [21]. This allowed us to maximize the cognitive load affecting the user and more accurately portray the use of the app in a real-world setting [21]. During this research, participants were not only exposed to naturally occurring distractions, but were also forced to participate in interactions with the moderator and answer additional questions that caused diversions.

A very important decision was the size of the study groups. Because of the need to perform the study free of charge, to prepare the subjects, and to systematize the external conditions in the field environment, the groups could not to be too large.

According to Nielsen [22], a sufficient number of testers is five. Following this logic, due to the small gain in the number of deficiencies found relative to the number of people taking the test, it is uneconomical to use more than five users [23]. However, as Faulkner [24] later demonstrated, the above statement is not accurate. There can be situations where a group of five people find 95% of the bugs in an application, but it may turn out that only 55% of errors will be found.

Twelve young people aged 22 to 31 were invited to the research, including 5 women and 7 men, as volunteers. All persons declared that they are advanced users of mobile devices. In order to minimize the influence of the learning factor on the research results, the participants were randomly divided into two groups A and B of 6 people. Table 2 shows the assignment of environments and applications to individual groups A and B.

Table 2. Arrangement of group of participants using applications in laboratory and field conditions

	Laboratory	Field
App1	Group A	Group B
App2	Group B	Group A

This arrangement allowed, with only two groups, to test the performance of both applications under different conditions with the elimination of the negative effect of participants learning the activities. As a result, we had 4 independent groups of participants. Subsequent groups by environment and application will be further labeled: *Lab_App1_A, Fld_App2_A, Lab_App2_B, Fld_App1_B.*

3.3 Task Scenario and Metrics Collected

In both applications, the number of tasks and their content were identical, the conditions under which the tests were performed were different. The research was conducted on the same new model of the Samsung Galaxy S10 smartphone with Android, in order to remove possible differences resulting from screen sizes of different brands and different operating systems. Tasks to be completed during the study are presented in Table 3.

Research for *Fld_App2_A* and *Fld_App1_B* groups, in order to reflect the standard use of the application as much as possible, was carried out in the outdoor world, i.e. in the middle of the day downtown. This allowed for a natural accumulation of environmental cognitive load. Additionally, participants were asked to walk continuously, and additional distracting questions were asked as they performed the tasks. Distracting questions to be answered during the study are shown in Table 4.

Table 3. Tasks to be completed during the study

No.	Task
1	Add an expense of PLN 20 to the Food category on January 10
2	Add an income of PLN 100 in the Salary category today
3	Add a spending limit of PLN 1000 for the month of January
4	Add a new expense category called “Birthday Party”
5	Read your January income/spending balance in the Report function

The study followed the ISO model of usability which is composed of three basic attributes: efficiency, effectiveness, and satisfaction. In total four metrics of efficiency and effectiveness were collected during completion of task scenarios. Moreover, two satisfaction questionnaires were administered after the whole scenario was accomplished. The metrics collected during the study are listed in Table 5.

Table 4. Distracting questions to be answered during the study

No.	Question	Comment
1	What did you eat for dinner yesterday?	Referring to the recent past requires a focus on the answer
2	What are your plans for the coming weekend?	The answer requires a longer elaboration, along with the use of the user's creativity
3	What color is my coat? (the moderator's coat)	The person will be forced to look away from the application, focus their attention on another object, and then return to the interrupted thread of activity

Table 5. Metrics collected during usability tests

Attribute	Metrics	Description	Unit
Efficiency	Time	Scenario completion time by a user excluding time for answering the distractive questions	[s]
	Actions	Number of clicks, scrolls, taps, swipes, etc. to complete the whole scenario	[n]
Effectiveness	Errors	Number of incorrect actions during completion of the whole task scenario	[n]
	Requests for help	Number of requests for assistance during the completion of the whole scenario	[n]
Satisfaction	NPS	Net Promoter Score - score of the single question survey administered after completion of the whole scenario	[−100,…,100]
	SUS	System Usability Scale - score of the 10 question survey administered after completion of the whole scenario	[0,…,100]

Time of Task Completion. Each of the respondents had time to perform their tasks measured. Time of scenario completion was measured for individual participant with an electronic stopwatch on a separate device, the examination was started with a voice announcement by the moderator. In the case of field tests, measuring was suspended while the respondent answered the distracting questions. In order to be able to make sure that the measured times were correct, the actions of each respondent were recorded using the *Screen Recorder* app provided by Google [25]. Due to the purpose of the research, which is to examine the impact of cognitive load on the usability of the application, the time of completing the entire task scenario was measured without considering the times of individual tasks.

Number of Actions Performed. This measure was obtained by analyzing the operation on the application using the *Screen Recorder* app. The action was touching the screen to tap on a given element, moving (scrolling), zooming in or out the screen. Later in the paper, the use of the term number of actions will refer directly to the term number of moves made and tapped elements.

Number of Errors. For each application, the optimal path with the fewest actions was determined. To obtain a measure of incorrect actions, the number of actions performed at a predetermined optimal path was subtracted from the number of actions performed by the user.

Number of Requests for Help. Despite the assumed failure to help the respondents, each of them had the opportunity to ask a question regarding the required activity. Each of these requests was recorded, including rhetorical questions without having to answer them. Due to the fact that the application should be designed so that the user can navigate in it without outside help, any, even rhetorical, requests for help were treated as negative reactions.

Net Promoter Score (NPS). The *Net Promoter Score* was adopted to measure participant satisfaction with the application. *NPS* is commonly used in business as a standard metrics of customer experience and approach to predict customer loyalty [26, 27]. We administered a question: "How likely is it that you would recommend this application to a friend or colleague?". The responses for this answer were scored on an 11 point scale from Very Unlikely (0) to Extremely Likely (10). According to the standard approach, the participants who responded with 9 or 10 were labelled Promoters; those who answered from 0 to 6 were termed Detractors. Responses of 7 and 8 were considered Neutrals and ignored. *NPS* was calculated by subtracting the percentage of Detractors from the percentage of Promoters according to Formula (1).

$$NPS = \left(\frac{No.\,of\ Promoters - No.\,of\ Detractors}{Total\ no.\,of\ Respondents}\right) * 100\% \tag{1}$$

NPS values vary between −100 and +100, with −100 occurring when each respondent critically evaluates the application and does not recommend it to other people, and +100 corresponds to a situation where each user recommends the application to friends. Positive *NPS* values are considered good results and values above 50 are considered excellent.

System Usability Scale (SUS). In our experiments, the user satisfaction was measured using the System Usability Scale [28–31]. *SUS* is a satisfaction questionnaire that has become the industry standard for over 30 years of use. *SUS* consists of 10 questions that users answer using a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Questions marked with odd numbers refer to the positive aspects of using the application, whereas even-numbered questions refer to negative aspects. The *SUS* score ranges from 0 to 100. The mean *SUS* score of the 500 studies surveyed by Sauro [28] is 68. Thus, an SUS score above 68 can be considered above average, and anything below 68 is below average. Research to date shows that *SUS* is a reliable and important measure of perceived utility. The advantage of *SUS* is that it is relatively fast, easy and inexpensive, and at the same time provides a reliable means of measuring usability.

3.4 Statistical Analysis Approach

The values of metrics collected for individual participants of each group: *Lab_App1_A, Fld_App2_A, Lab_App2_B, Fld_App1_B* were used to examine differences in performance between individual groups. The parametric and non-parametric statistical tests for independent groups were employed as shown in Table 6. To determine what type of statistical significance tests should be used, the normality and equality of variance of the individual metrics are examined. In the event that if there is no evidence for rejecting null hypotheses in the Shapiro-Wilk test and Fisher-Snedecor, the parametric tests Student's t-test for independent groups should be employed. On the other hand, if the null hypotheses in the Shapiro-Wilk test is rejected, then the nonparametric Levene's test and Mann-Whitney U test should be applied. The level of significance was set to 0.05 in each test. All tests were performed using the *PQStat* software for statistical data analysis [29].

Table 6. Summary of statistical tests for independent groups in the *PQStat*

Feature examined	Test	
Normality of distribution	Shapiro–Wilk test	
Feature examined	**Normal distribution**	**Nonnormal distribution**
Equality of variances	Fisher-Snedecor test	Levene's test
Feature examined	**Parametric tests**	**Nonparametric tests**
Means/Medians	Student's t-test for independent groups	Mann-Whitney U test

4 Results of Usability Testing

The results of usability testing for individual groups of participants are presented in Figs. 3, 4, 5 and 6. It is clearly seen that average and median time of task completion, number of actions performed, number of errors committed, number of requests for help are lower for groups working in laboratory conditions as well as for groups utilizing *App1*. Statistical tests for independent groups revealed that the differences between mean values of following performance measures task completion time and error numbers were statistically significant in each case, but not for the number of actions. In all tests the distributions of collected data were normal and variances were equal. In consequence the parametric Student's t-tests for independent groups were carried out.

The results of the satisfaction questionnaires are presented in Figs. 7 and 8. In the case of both *NPS* and *SUS* Score measures, the subjective ratings of both applications were identical regardless of the environment in which they were used. Moreover, *App1* was positively assessed and *App2* was rejected by the research participants.

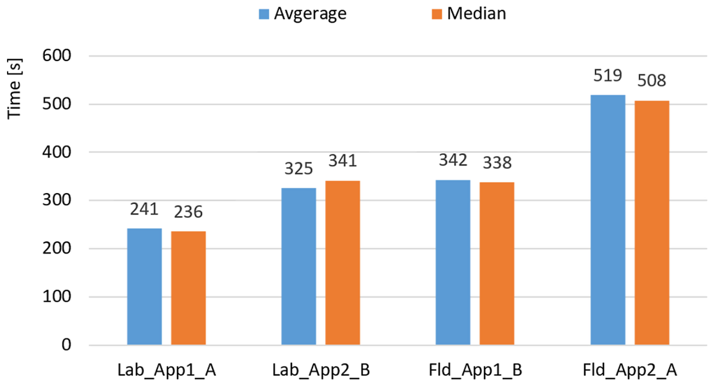

Fig. 3. Average and median time of task completion by individual groups of participants

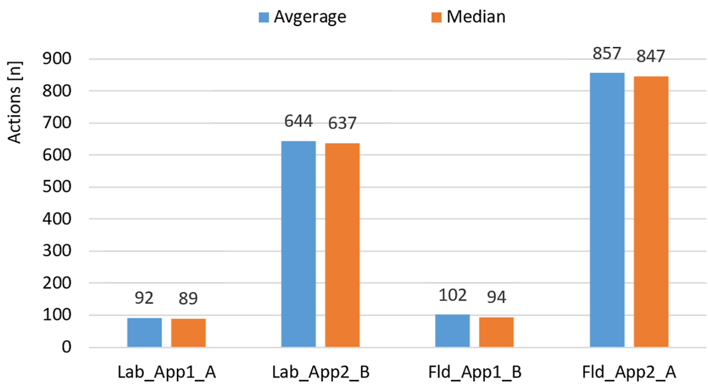

Fig. 4. Average and median number of actions performed by individual groups of participants

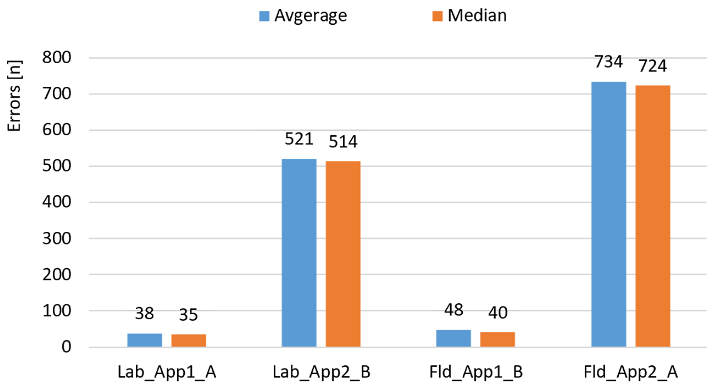

Fig. 5. Average and median number of errors committed by individual groups of participants

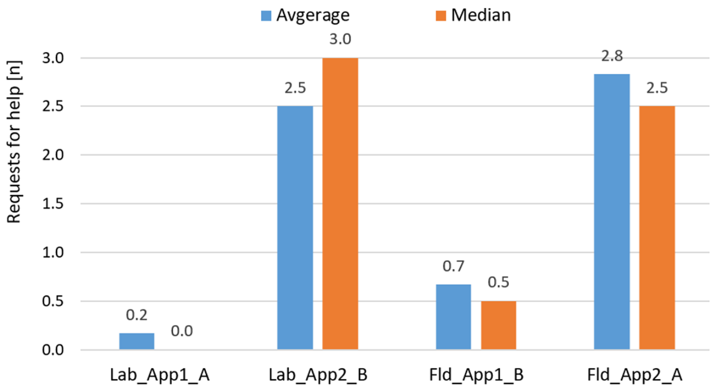

Fig. 6. Average and median number of requests for help by individual groups of participants

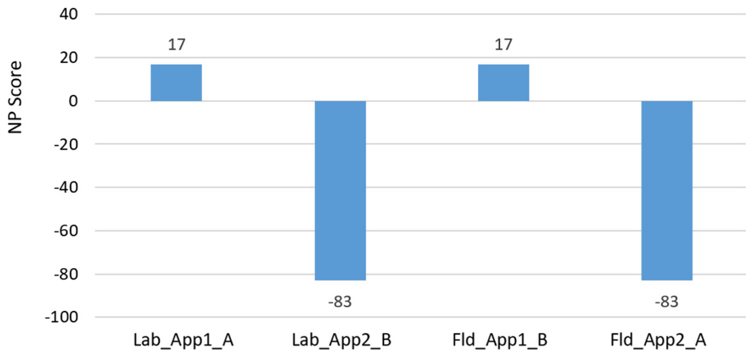

Fig. 7. *Net Promoter Score* for individual groups of participants

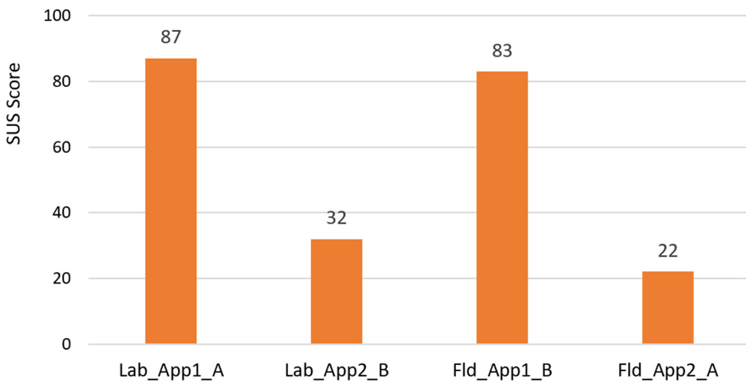

Fig. 8. *SUS* Score for individual groups of participants

5 Conclusions

In this study, the usability of two mobile applications was examined in terms of cognitive load resulting from environmental factors, with the participation of people aged 20–30. During the research, two mobile applications available on the market, intended for young people, for managing home finances were used.

In order to prove and investigate the negative impact of the cognitive load resulting from environmental factors and the graphical interface of the applications themselves, user testing was carried out. A total of 12 participants took part in the study. They were divided into groups of 6 people. Each of these groups took part in the study of both applications in laboratory and field conditions.

The collected data allowed for a thorough analysis of user environments and behavior, including the use of t-Student statistical analysis for independent groups.

The first obtained and easily measurable result was proving the problem of cognitive load associated with the user interface. By demonstrating imperfections in the graphical interface, proving differences in task execution time and user frustration depending on the applications used, it is easy to draw conclusions about the need for a thorough analysis of the end-users of the potential application.

Another aspect examined in this work was the cognitive load and behavior of users in various applications. These results were not predictable. Just as an easy assumption and simple to prove is the recognition that cognitive load will always have a negative impact on perception and action across the system, a few exceptions emerged in the course of research.

The analysis showed that the average number of actions in the groups of applications are statistically insignificant. This makes it possible to conclude that in the case of a well-designed interface, where messages and labels are clear and transparent, the user under the influence of unfavorable weather conditions, disturbances from real life, still is able to complete his/her task.

An interesting discovery was the independence of the *NPS* and *SUS* survey results from environmental conditions. Of course, the differences between *App1* and *App2* ratings remain unchanged. In addition, it can be stated that if the application is unsatisfactory under favorable laboratory conditions, it will be just as unattractive under worse field conditions.

As part of the analysis of the results, the statistical difference between the *App1* and *App2* applications was demonstrated. *App1* showed an advantage in each of the collected measures over the other tested software. In the process of creating this application, the user interface was carefully developed, standard mobile application design patterns were used and, most of all, target audience was examined. *App2*, despite providing similar functionalities to *Add1*, had errors in their bad nomenclature. Additionally, the designed user path was not intuitive and did not have any useful help.

References

1. Leppink, J.: Cognitive load theory: practical implications and an important. J. Taibah Univ. Med. Sci. **12**(5), 385–391 (2017). https://doi.org/10.1016/j.jtumed.2017.05.003
2. de Jong, T.: Cognitive load theory, educational research, and instructional design: some food for thought. Instr. Sci. **38**, 105–134 (2010). https://doi.org/10.1007/s11251-009-9110-0
3. Bernacki, J., Błażejczyk, I., Indyka-Piasecka, A., Kopel, M., Kukla, E., Trawiński, B.: Responsive web design: testing usability of mobile web applications. In: Nguyen, N.T., Trawiński, B., Fujita, H., Hong, T.-P. (eds.) ACIIDS 2016. LNCS (LNAI), vol. 9621, pp. 257–269. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49381-6_25
4. Błażejczyk, I., Trawiński, B., Indyka-Piasecka, A., Kopel, M., Kukla, E., Bernacki, J.: Usability testing of a mobile friendly web conference service. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 565–579. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_52
5. Krzewińska, J., Indyka-Piasecka, A., Kopel, M., Kukla, E., Telec, Z., Trawiński, B.: Usability testing of a responsive web system for a school for disabled children. In: Nguyen, N.T., Hoang, D.H., Hong, T.-P., Pham, H., Trawiński, B. (eds.) ACIIDS 2018. LNCS (LNAI), vol. 10751, pp. 705–716. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75417-8_66
6. Myka, J., Indyka-Piasecka, A., Telec, Z., Trawiński, B., Dac, H.C.: Comparative analysis of usability of data entry design patterns for mobile applications. In: Nguyen, N.T., Gaol, F.L., Hong, T.-P., Trawiński, B. (eds.) ACIIDS 2019. LNCS (LNAI), vol. 11431, pp. 737–750. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14799-0_63
7. Waloszek, S., Zihisire Muke, P., Piwowarczyk, M., Telec, Z., Trawiński, B., Nguyen, L.T.T.: Usability study of data entry design patterns for mobile applications. In: Nguyen, N.T., Hoang, B.H., Huynh, C.P., Hwang, D., Trawiński, B., Vossen, G. (eds.) ICCCI 2020. LNCS (LNAI), vol. 12496, pp. 888–901. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63007-2_70
8. Weichbroth, P.: Usability of mobile applications: a systematic literature study. IEEE Access **8**, 55563–55577 (2020). https://doi.org/10.1109/ACCESS.2020.2981892
9. Bashirl, M.S., Farooq, A.: EUHSA: extending usability heuristics for smartphone application. IEEE Access **7**, 100838–100859 (2019). https://doi.org/10.1109/ACCESS.2019.2923720
10. Da Costa, R.P., Canedo, E.D., De Sousa, R.T., De Oliveira Albuquerque, R., Garcia Villalba, L.J.: Set of usability heuristics for quality assessment of mobile applications on smartphones. IEEE Access **7**, 116145–116161 (2019). https://doi.org/10.1109/ACCESS.2019.2910778
11. Harrison, R., Flood, D., Duce, D.: Usability of mobile applications: literature review and rationale for a new usability model. J. Interact. Sci. **1**, 1 (2013). https://doi.org/10.1186/2194-0827-1-1
12. Ejaz, A., Rahim, M., Khoja, S.A.: The effect of cognitive load on gesture acceptability of older adults in mobile application. In: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019, pp. 0979–0986 (2019). https://doi.org/10.1109/UEMCON47517.2019.8992970
13. Sunardi, Desak, G.F.P., Gintoro: List of most usability evaluation in mobile application: a systematic literature review. In: Proceedings of 2020 International Conference on Information Management and Technology, ICIMTech 2020, pp. 283–287 (2020). https://doi.org/10.1109/ICIMTech50083.2020.9211160
14. Paas, F., van Merriënboer, J.J.G.: Instructional control of cognitive load in the training of complex cognitive tasks. Educ. Psychol. Rev. **6**(4), 351–371 (1994). https://doi.org/10.1007/BF02213420
15. Sun, G., Yao, S., Carretero, J.A.: A pilot study for investigating factors that affect cognitive load in the conceptual design process. In: Proceedings of the Human Factors and Ergonomics Society, pp. 180–184 (2015). https://doi.org/10.1177/1541931215591037

16. Liang, Y., Liang, W., Qu, J., Yang, J.: Experimental study on EEG with different cognitive load. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018, pp. 4351–4356 (2019). https://doi.org/10.1109/SMC.2018.00735
17. Ferreira, E., et al.: Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults. In: IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain, Proceedings, pp. 39–48 (2014). https://doi.org/10.1109/CCMB.2014.7020692
18. Fast Budget - Expense & Money Manager: https://play.google.com/store/apps/details?id=com.blodhgard.easybudget&hl=en&gl=US. Accessed 15 Jan 2021
19. Family Budget: https://play.google.com/store/apps/details?id=pl.com.cierniak.android.familybudget&hl=en&gl=US. Accessed 15 Jan 2021
20. Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., Kankainen, A.: Usability testing of mobile applications: a comparison between laboratory and field testing. J. Usability Stud. **1**(1), 4–16 (2005)
21. Alshehri, F., Freeman, M.: Methods for usability evaluations of mobile devices. In: Lamp, J.W. (eds.) 23rd Australian Conference on Information Systems, pp. 1–10. Deakin University, Geelong (2012)
22. Nielsen, J.: Why You Only Need to Test with 5 Users: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/. Accessed 15 Jan 2021
23. Macefield, R.: How to specify the participant group size for usability studies: a practitioner's guide. J. Usability Stud. **5**(1), 34–45 (2009)
24. Faulkner, L.: Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav. Res. Methods Instrum. Comput. **35**(3), 379–383 (2003)
25. Screen Recorder: https://play.google.com/store/apps/details?id=com.fragileheart.screenrecorder&hl=en_US&gl=US. Accessed 15 Jan 2021
26. Net Promoter Score. A Comprehensive Introduction. https://www.genroe.com/net-promoter-score. Accessed 15 Jan 2021
27. Pollack, B.L., Alexandrov, A.: Nomological validity of the net promoter index question. J. Serv. Mark. **27**(1), 118–129 (2013). https://doi.org/10.1108/08876041311309243
28. Sauro, J.: Measuring Usability with the System Usability Scale (SUS): https://measuringu.com/sus/. Accessed 15 Jan 2021
29. Thomas, N.: How to Use the System Usability Scale (SUS) to Evaluate the Usability of Your Website: https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/. Accessed 15 Jan 2021
30. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Int. J. Hum.-Comput. Interact. **24**(6), 574–594 (2008)
31. Brooke, J.: SUS: a retrospective. J. Usability Stud. **8**(2), 29–40 (2013)
32. Wieckowska, B.: User Guide - PQStat. Downloaded from https://download.pqstat.pl/UserGuideGeo.pdf. Accessed 15 Jan 2021