ASSESSING THE USABILITY OF A MOBILE APP IN A REMOTE ENVIRONMENT: COMBINING THINK ALOUD WITH A USER EXPERIENCE SCALE QUESTIONNAIRE

Virginia Tiradentes Souto and Luiza Reolon Cabral
Design Department, University of Brasilia, Brazil

ABSTRACT

The aim of this study is to investigate an approach for evaluating the user experience of mobile applications in remote environments. The proposed approach includes both thinking aloud and a user experience questionnaire with three types of questions: personalized, scaled, and open-ended. The novelty of this study lies in the proposal of a new form with questions that not only focus on usability but also include other aspects of user experience, such as user characteristics, product features, and previous user experiences. The proposed App User eXperience Scale (AUXS) is based on SUS, but has many differences, including questions about content, visual appeal, content, user expectations, and an NPS question. The proposed approach is tested using a mobile application to promote artists and artisans. The results show that this approach can be useful not only for researchers, but also for designers to validate their creations and for companies offering mobile applications to test the usability of their applications.

KEYWORDS

User Experience, Mobile Application, Remote Environment, Think Aloud, Questionnaires

1. INTRODUCTION

Testing the user experience of mobile applications is important at various stages of product development, such as the interface design phase, the prototype phase and also after implementation and before launch. It is also important to consider how the researcher/designer can reach the target audience for participation in the study. Participants with certain characteristics or distances may be difficult to reach, and the use of remote tools may be a good option for researchers/designers, as well as in times of pandemic when laboratory testing may pose some risk.

There are several methods to test the user experience of a mobile application, such as: task completion, interviews, questionnaires, thinking aloud, heuristic tests, focus groups and more. Each of these tests is used for a specific purpose and has some advantages and disadvantages, so it must be selected considering the context and needs of the project. Among these methods, the think-aloud protocol stands out as one of the most popular methods in usability research (Nielsen, 2012). The think-aloud protocol is considered a highly valued (Fan et al. 2019) and effective method for usability testing to learn people's thoughts about mediating the completion of a task (Souto, 2013). It has been used for many years in various research areas such as cognitive psychology (Davison et al., 1997) and human-computer interaction (Ramey et al., 2006). Thinking aloud has also been used to test mobile apps (Nasruddin et al., 2018) and is considered a useful method for obtaining immediate feedback from end-users about their experience interacting with the app (Oliveira et al. 2022).

Besides the think-aloud protocol, another very common method for usability testing is the questionnaire. A specific type of questionnaire that uses a scale to measure responses is the System Usability Scale (SUS), which can be applied in a short period of time and measures user perceptions (Brooke, 1996). SUS is a widely used tool to determine usability aspects of various systems (Oliveira et al., 2022) and has also been used to evaluate the user experience of mobile apps. Each user experience testing method has advantages and disadvantages, and therefore different evaluation methods are more appropriate depending on the study. Qualitative methods such as the think-aloud protocol can be more useful for understanding user thoughts, gathering user feedback,

and identifying issues that need to be addressed. On the other hand, quantitative methods such as usability scaling can provide a general measure of usability (Maramba et al., 2019).

In this study, we explored a user experience evaluation approach for mobile apps using the Think Aloud protocol and a user experience scale inspired by the SUS tool. As a case study, we use an application to promote artists and artisans that is in a testing phase and has not yet been published.

2. RELATED RESEARCH

The proposed approach for evaluating mobile applications in remote environments was based on the Concurrent Think Aloud protocol, followed by a questionnaire based on the System Usability Scale tool. Both methods are described below, as well as mixed approaches to mobile application usability assessment.

2.1 Concurrent Think-Aloud Protocol

The concurrent thinking aloud method (also referred to as classic, McDonald et al. 2016) was first introduced to the literature by Ericsson and Simon (1980) and refers to a test in which participants are asked to verbalise their thoughts while performing tasks (Souto, 2013). Another type of thinking aloud, also introduced by Ericsson and Simon (1980), is retrospective thinking aloud, in which participants verbalise their thoughts after performing tasks.

The think-aloud protocol is considered a widely researched topic (Reeves, 2019). Many researchers have studied the implementation of think-aloud protocols (Fan et al. 2021), the effect of the rater on the think-aloud test (Jacobsen et al. 1998), and also the variability with which instructions are given. For example, Hertzum, Hansen, and Andersen (2009) have shown that changing the instructions given to participants on how to think aloud affects both the assessment process and the outcome.

Many studies that focus on the usability of mobile applications have been conducted using the think-aloud method. Maramba, Chatterjee, and Newman (2019), in a review of methods used exclusively in usability testing of eHealth applications (from 2014 to 2017), found that ⅓ of the studies used the think-aloud method (45 of 133 studies), which was the most commonly used qualitative method. In addition, the think-aloud method was associated with at least one additional iteration of the developed application.

2.2 Usability Scale Questionnaires

Questionnaires are considered one of the most commonly used usability methods. Among their advantages is the fact that they are easy to conduct and analyze [25]. Quizzes are also the most common usability tests found in various literature reviews on mobile apps [11; 26]). According to the literature review by Maramba, Chatterjee, and Newman (2009), the SUS was the most commonly used questionnaire (44 out of 133 studies).

The System Usability Scale (SUS) is a commonly used questionnaire to measure the usability of human-computer interaction systems and mobile applications in particular. According to Brooke (2013), the main goals of SUS are: to provide a measure of subjective perceptions of the usability of a system, and to perform this assessment in the short time available in a session.

The SUS consists of 10 questions, each with five possible answers on a 5-point Likert scale (from "strongly disagree" to "strongly agree"). The questions relate to the ease of use and simplicity of the system being learned and concern: frequency of use, system complexity, technical support, functional integration, system consistency, system reliability, learning to systemize. The ten questions of SUS are (Brooke, 1996): (1) I think that I would like to use this system frequently, (2) I found the system unnecessarily complex, (3) I thought the system was easy to use, (4) I think that I would need the support of a technical person to be able to use this system, (5) I found the various functions in this system were well integrated, (6) I thought there was too much inconsistency in this system, (7) I would imagine that most people would learn to use this system very quickly, (8) I found the system very cumbersome to use, (9) I felt very confident using the system, and (10) I needed to learn a lot of things before I could get going with this system.

More recently, some researchers have proposed adaptations of the SUS tool to test the usability of mobile apps (Kaya et al., 2019), as well as other tools developed for specific topics, such as the specific questionnaire for health-related apps: MARS - mHealth app quality rating tool (Stoyanov et al. 2015), and MAUQ - The

mHealth app usability questionnaire (Zhou et al., 2019). Some of the reasons for adapting the tool SUS are that SUS is not a specific scale for mobile apps, and while it is useful for identifying general usability issues and providing an overview, it is not intended to identify usability issues specific to mobile devices (Kaya et al., 2019). There are also some versions for specific languages, such as Spanish, to ensure conceptual, semantic, and contextual equivalence between SUS and the translated version (Sevilla-Gonzalez et al., 2020) or Indonesian adaptation through cross-cultural adaptation and reliability testing (Sharfina and Santoso, 2016), or for specific user groups, such as the SUS, adapted for testing with children (Putnam et al., 2020).

2.3 Mixing Approaches to Measuring Usability of Mobile Applications

As mentioned by Drew, Falcone, and Baccus (2018), SUS can be more useful when used in conjunction with other usability methods. There are several studies that use a mixed methods approach to mobile application evaluation. For example, O'Grady et al. (2019) used a combination of 4 methods: think-aloud tasks, task success ratings, semistructured interviews, and usability questionnaires (e.g. SUS) to assess the usability of a mobile application for guiding health professionals. To measure usability ratings of a mobile application for neuroanatomy, Ponte et al. (2019) used two types of questionnaires: SUS to determine usability and learnability; and another five questions to determine participants' perceptions of the usefulness of the application, perceived usefulness. On the other hand, Weichbroth (2019) proposes a mixed methodology for measuring and evaluating the usability of mobile applications. This methodology consists of three integrated methods: research, participant observation, and verbal protocol analysis, and aims to provide quantitative and qualitative data and a complete overview of the quality of application use and users' attitudes and perceptions.

While some authors mention the effectiveness of their mixed-methods approaches, others only address the effectiveness of the product tested. The aim of this study is to test a methodological approach to evaluate the user experience of mobile applications. The study is based on a case study of an application for promoting artists and artisans, which is described below.

3. METHOD

To test the effectiveness of combining Thinking Aloud methods with usability questionnaires in a remote environment, an empirical experiment was conducted. The mobile application used in this study, Nicho, an application to promote artists and artisans, was developed by some of the researchers who participated in this study and has not yet been launched. The methodology and results are described below.

3.1 Case study

Testing was conducted using a concurrent think-aloud protocol followed by a usability questionnaire scale with a subjective question. The study used as research material a new application that is in its final testing phase. The application, called Nicho, is intended to be a virtual space where artisans and artists can promote their work and connect with customers. It also provides information about physical fairs taking place in the city, as well as information for artists and artisans. The app has two main user groups: Artists and artisans and the customers. The main areas of the app are: product, places, saved, follow e chat. Figure 1 show examples of Nicho application screens.

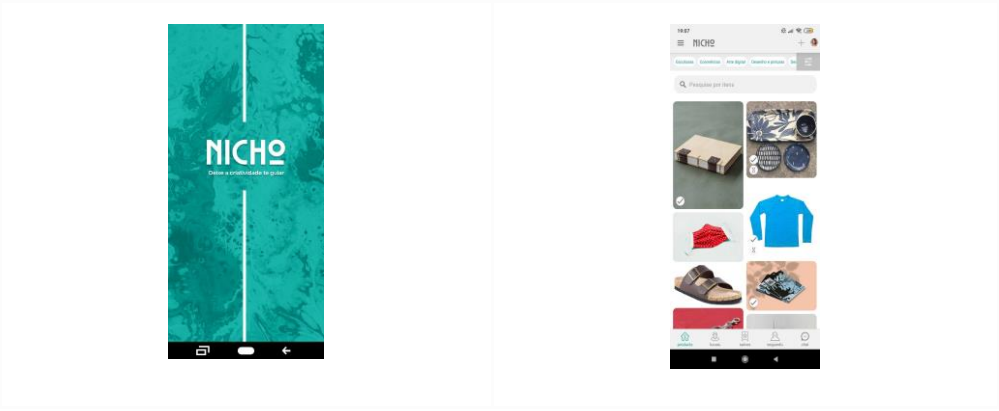

Figure 1. Screen imageshot of the app presentation (left) and the Nicho app home page

3.2 Data Collection and Participants

The purpose of the test was to verify the user experience with the application in the two established user groups. The tests were conducted in remote mode. First, the researchers visited physical fairs held in the city to recruit representative participants for the study. The study targeted different groups of participants: men and women, artists/craftspeople, and customers.

First, participants received an email or Whatsapp message (depending on the participant's choice) to schedule an appointment for the test. Then, participants were briefly informed about the experiment and asked to complete the informed consent form. To participate in this study, participants were required to use an Android phone, as the application only works on an Android platform for now. Participants were required to download the app to their cell phones. Testing was conducted in the Google Meet system, so they had access to the video call through their browser and did not need to download a communication app to connect to the researcher specifically for the test. Participants' activities were recorded during the tests, and they were asked to answer the questions and speak aloud while interacting with the application. There were two ways to access the application: one for clients and one for artists. Customers answered 9 questions and artists answered 10 questions (there was an additional task for artists: they were asked to publish a product). Examples of questions are: (1) Find ceramic plates and see what their price is; (2) Make the purchase of these dishes; and (3) Choose a product you like and save it. After participants answered the questions, the recording was stopped, and participants were asked to answer a Google form containing 4 pieces of personal information (sex, age, occupation, education), 11 scaled questions (1 to 5, strongly disagree to strongly agree), and 2 open-ended questions (one about the application and one about the test).

In agreement with Kaya et al. (2019) that SUS is not a specific scale for mobile applications and therefore may not be as helpful in identifying specific user experience issues for mobile devices, a user experience scale form is proposed that includes questions more related to mobile applications. The user experience tool used in this test included 4 questions adapted from the SUS tool (questions 1, 2, 3, and 5), with the same meaning and almost the same words as the original (only adapted to be more consistent with the other questions, including the term app instead of system), others revised to be more direct and adapted to the context of use, one based on an NPS question, but instead of a scale from 0 to 10, the scale 1 to 5 was kept. The proposed scaled questions of the form (translated from Portuguese) and the open-ended questions were:

App User Experience Scale (AUES):

1. I would use this app often.
2. I found the app easy to use.
3. I think I would need the help of a person with more experience to use it.
4. I think the features of the app would be useful to me.
5. I can imagine learning how to use this app quickly.
6. How likely are you to recommend this app to others?
7. I found the app to be visually appealing.
8. I think this app stands out from other apps like Etsy, Instagram, and Pinterest.

9. I like the content of the app.
10. It took me a long time to find anything.
11. Overall, my expectations for the app were met.

Open questions:

1. Is there anything else you would like to say about the app?
2. What do you think of our survey?

To ensure that the researchers proceeded in the same way for each of the participants, a research route was planned. As Hertzum, Hansen, and Andersen (2009) noted, the variability used to explain the instructions in the think-aloud test may affect the results of the study. Also, considering that the test is conducted in a remote environment, many steps are required, as well as technical equipment (cell phones, operating systems, recording software, contact software, etc.), and prior contact, downloading the application, and more. Considering all these aspects, the research journey for this study was divided into two main phases: the selection of users and the day of the interview.

The test was conducted with 10 participants, but two results were discarded due to problems in recording the test. For this, the data of 8 participants were analyzed: half were artists/craftsmen and the other half were customers. Regarding sex, four were female, 3 were male, and one preferred not to provide any information. The age group of the participants was very diverse: half were between 20 and 20 years old and the others were between 30 and 60 years old or older. The participants had different professions: the clients were civil servants, students, college professors and entrepreneurs, while the artists described themselves as artists, designers and copywriters, plastic artists and illustrators. All of them had at least a college degree.

3.3 Think-Aloud Results and Discussion

Participants took about 10 minutes to answer the questions (minimum 7, maximum 13 minutes). Of all the answers, only one participant could not solve one task (Find the product you previously saved). This indicates that participants had no problems with accuracy when using the app. In general, the time to answer the questions decreased as the test progressed, indicating that familiarity with the app is helpful to the user experience.

It was observed that participants expressed their feelings, preferences, and navigation behaviors when using the app by speaking aloud. For example, one participant found it strange to create a storage folder and suggested the Instagram app as an example of use. In terms of navigation, the results show that thinking aloud helps designers understand where users look first, what grabs their attention, and which path users prefer. This is because users talked out loud about where they look, how they navigate, and what they pay attention to.

Participants' insights or suggestions for improving the Niche app include: increasing the size of the profile photo, linking artists, decreasing the number of levels to reach the purchase screen, and increasing the size of the font in some areas. In addition, it was recognized that the method of thinking aloud captured the emotions of the participants. The joy of finding something they like, or the disappointment of not finding it, or still having difficulty finding an answer to a question, or an estrangement with a function different from what they are used to is captured by this method.

3.4 App User Experience Scale (AUES): Results and Discussion

As explained earlier, the App user experience scale (AUES) proposed in this study includes 11 questions and uses a scale from 1 to 5 (strongly disagree to strongly agree). The proposed questionnaire is based on SUS, but has many differences, including questions about content, visual appeal, content, user expectations, and an NPS question. The AUXS calculation method is the same as for SUS: (1) add the score contributions of each item (from 1 to 5), (2) for positive items related to the application, the score is the position of the scale minus 1, for negative items (items 3 and 10), the score is 5 minus the position of the scale, (3) multiply the sum of the scores by 2.5 to get the total score. The score ranges from 0 to 100.

In terms of frequency, the majority of participants found the app easy to use and that they would use the app frequently. The third question was about the need for help in using the app. But unlike the original question in SUS tools, which asks if the participant thinks he/she "would need the support of a technical person to be able to use this system," it was adapted in this form to ask for help from "a person with more experience." This is meant to sound more common, as users have quicker access to people around them with more experience

who can help them with the app than assistance from a "technical person". In this test, most participants felt they did not need help from a person with more experience, confirming the ease of use of the app.

Regarding the NPS (Net Promoter Score) question: How likely are you to recommend this app to others? most participants (87.5%) answered 5, and one participant (12.5%) answered 4. This means that it is valid to spread the application to others. The NPS is considered a simple and effective method for measuring and monitoring customer satisfaction (Sasmito et al., 2019). Moreover, positive NPS (likelihood to recommend) scores indicate that users' willingness to install an app also increases (Wohllebe et al., 2020).

In this study, all participants found the app visually appealing. The beauty of the app was also mentioned by some participants during the Think Aloud test and in the open-ended questions.

This form also differs from that of the SUS tool in that it includes a question about the content of the app. How close or interested users are in the content of the app seems to have an impact on their experience with the app. According to Stoyanov et al. (2015), one of the 5 app quality criteria is the information quality of the app (the other four are: engagement, functionality, aesthetics, and subjective quality categories). The authors divide information quality into seven sub-criteria: accuracy of app description, goals, quality of information, quantity of information, visual information, credibility, evidence base. The Nicho app not only establishes a contact between artists, artisans, and consumers, but also provides information about the artists so that they can be informed about certain laws, rights, deities and opportunities. It also provides a space to inform consumers and interested parties about events such as fairs in a particular city.

The last question was related to the users' general expectation of the app. This question is also important to understand the overall picture of the app for users. Even if users find some points of improvement or difficulties in the user experience, the overall impression may be positive or negative, and the problems found may or may not affect the overall impression and desire to continue using the app. In the case study, participants felt that their overall expectations of the app were I (91.37 out of 100).

3.5 Insights from the Opening Questions

The test contained two opening questions: one related to the app and the other to the survey. These are not required answers to complete the form. Regarding the app, the form included a question asking if the participant had anything else to say about the app. Six of the eight participants answered this question. Some of the participants mentioned good things about the app, such as: easy to learn, very intuitive user interface, great logo. However, others took the opportunity to make other suggestions for the app, such as: Repeat the idea of creating "sub-filters" and gave an example: "Accessories - > silver accessories | gold accessories | with stone...; Ceramics - > painted | nature | vases | dishes | mugs (in this case you could select more than one, such as painted ceramics and mugs" (Participant's text translated from Portuguese)".

The last question of the form was designed to clarify participants' perceptions of the questionnaire, which contained three types of questions: Multiple choice, scaled, and open-ended questions. Five of the eight participants answered this question. All provided positive feedback about the survey, such as: easy to understand and participate in, agile, well-structured and straightforward, calm, and very direct.

4. CONCLUSIONS

The proposed approach to evaluating the user experience of mobile applications in remote environments includes both thinking aloud and a user experience questionnaire with personalized, scaled, and open-ended questions. While it can be considered common to mix thinking aloud and questionnaires in this type of test, the novelty of this study is the description and explanation of the whole approach (including the research journey), as well as the proposal of a new form with questions that not only focus on usability but also include other aspects of user experience, such as user characteristics, product features, and previous user experiences. In addition, the proposed approach focused on the evaluation of mobile applications in remote environments.

This study confirms other studies that claim that the Thinking Aloud method can help understand users' thoughts and identify usability problems when using the app (Nasuddin et al. 2018; Cho et al., 2019). In addition, the results show that the Thinking Aloud method promotes participants' spontaneity and sincerity. The participants' emotions could be captured during the test, with their engagement, difficulties, and surprises.

Finally, the thinking aloud method can also help designers compare their app with the competition, as participants mention features they know from other apps or would like to find in the app they are using.

The proposed scaled questionnaire (called AUXS - App User eXperience Scale) showed to be an effective approach to evaluate the user experience of a mobile application. Like the SUS tool (Brooke, 1995), the proposed tool is easy to answer and quick to apply. The differences in terms of questions, as explained earlier, are in the evaluation of aspects of the user experience that are not mentioned in SUS, but others, such as: User characteristics (information about the user's age, gender, education and profession), internal state (question about the user's general expectations), cultural differences (the form was adapted to the Portuguese language and the context of use by the participants), product characteristics (questions about the usefulness of the functions and the content), brand image (question about visual appearance) and previous experience (question comparing very well-known similar applications). The questionnaire contains different types of questions: multiple choice (for user characteristics), scaled questions (from 1 to 5, strongly disagree to strongly agree), and open-ended questions. It also contains an NPS question aimed at indicating the level of potential recommendation of the app and, related to this, the possibility of installing it (Wohllebe et al., 2020).

Regarding the limitations of this study, it is important to highlight that the study was conducted with few participants. Although studies suggest that as few as 5 participants are sufficient for usability testing (Nielsen, 2000), it seems that the proposed form should be tested in different contexts and with more people to further verify the order of questions and the number of positive and negative statements, among others. Another limitation was that the selected case study so far only works on the Android system (which limited the number of participants). Since it was difficult to find volunteers and organize the tests with them, it was not possible to check if there were differences between the different groups of participants, for example between men and women, young and elderly. Therefore, the tests should be conducted with more participants, different groups of participants and different case studies (with different types of applications).

ACKNOWLEDGEMENT

We would like to thank all the participants who took part in this research. This research was supported by the Department of Design - Art Institute - University of Brasília.

REFERENCES

Brooke, B. 1996. SUS: A 'Quick and Dirty' Usability Scale. In: Usability Evaluation In Industry. Edited by: dited By Patrick W. Jordan, B. Thomas, Ian Lyall McClelland, Bernard Weerdmeester. CRC Press.

Cho H, Powell D, Pichon A, Kuhns LM, Garofalo R, Schnall R. 2019. Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. Int J Med Inform. 2019 Sep;129:366-373. doi: 10.1016/j.ijmedinf.2019.07.010.

Davison, G. C., Vogel, R. S., & Coffman, S. G. (1997). Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. Journal of Consulting and Clinical Psychology, 65(6), 950–958. https://doi.org/10.1037/0022-006X.65.6.950

Drew, M.R., Falcone, B., Baccus, W.L. 2018. What Does the System Usability Scale (SUS) Measure?. In: Marcus, A., Wang, W. (eds) Design, User Experience, and Usability: Theory and Practice. DUXU 2018. Lecture Notes in Computer Science(), vol 10918. Springer, Cham. https://doi.org/10.1007/978-3-319-91797-9_25

Ericsson, K. A. and Simon, H. A. 1980. Verbal reports as data. _Psychological Review_ 87 (3):215-251. DOI 10.1037/0033-295x.87.3.215

Fan, M., Zhao, Q. and Tibdewal, V.. 2021. Older Adults' Think-Aloud Verbalizations and Speech Features for Identifying User Experience Problems. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 358, 1–13. https://doi.org/10.1145/3411764.3445680

Fan, M., Lin, J., Chung, C. and Truong, K. N. 2019. Concurrent Think-Aloud Verbalizations and Usability Problems. ACM Trans. Comput.-Hum. Interact. 26, 5, Article 28 (October 2019). https://doi.org/10.1145/3325281

Hertzum, M., Hansen, K. D., and Andersen, H. H. K. 2009. Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload?, Behaviour & Information Technology, 28:2, 165-181, DOI: 10.1080/01449290701773842

Jacobsen, Niels & Hertzum, Morten & John, Bonnie. (1998). The Evaluator Effect in Usability Studies: Problem Detection and Severity Judgments. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 42. 1336-1340. 10.1177/154193129804201902.

Kaya, A., Ozturk, R., Altin Gumussoy, C. 2019. Usability Measurement of Mobile Applications with System Usability Scale (SUS). In: Calisir, F., Cevikcan, E., Camgoz Akdag, H. (eds) Industrial Engineering in the Big Data Era. Lecture Notes in Management and Industrial Engineering. Springer, Cham. https://doi.org/10.1007/978-3-030-03317-0_32

Maramba, I.D., Chatterjee, A., & Newman, C. 2019. Methods of usability testing in the development of eHealth applications: A scoping review. International journal of medical informatics, 126, 95-104. DOI:10.1016/J.IJMEDINF.2019.03.018

McDonald, S., Zhao, T. and Edwards. H. M. 2016.Look Who's Talking: Evaluating the Utility of Interventions During an Interactive Think-Aloud, in Interacting with Computers, 28 (3), 387-403, May 2016, doi: 10.1093/iwc/iwv014.

Nasruddin, Z.A., Markom, A., Abdul Aziz, M. (2018). Evaluating Construction Defect Mobile App Using Think Aloud. In: Abdullah, N., Wan Adnan, W., Foth, M. (eds) User Science and Engineering. i-USEr 2018. Communications in Computer and Information Science, vol 886. Springer, Singapore. https://doi.org/10.1007/978-981-13-1628-9_1

Nielsen, J. 2000. Why You Only Need to Test with 5 Users. https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Nielsen, J. 2012. Thinking Aloud: The #1 Usability Tool. https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

Oliveira, E.R., Branco, A.C., Carvalho, D. et al. An Iterative Process for the Evaluation of a Mobile Application Prototype. SN COMPUT. SCI. 3, 262 (2022). https://doi.org/10.1007/s42979-022-01153-6

Ponte, R. P., Sanders, L. L. O., Peixoto Júnior, A. A., Kubrusly, M., Marçal, E. 2019. Development and Usability Assessment of a Mobile Application for Neuroanatomy Teaching: A Case Study in Brazil. Creative Education, Vol. 10 No.3. DOI: 10.4236/ce.2019.103043

Ramey, J., Boren, T., Cuddihy, E., Dumas, J., Guan, Z., van den Haak, M. J. and De Jong, M. D. T.. 2006. Does think aloud work? how do we know? In CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06). Association for Computing Machinery, New York, NY, USA, 45–48. https://doi.org/10.1145/1125451.1125464

Reeves, S. 2019. How UX Practitioners Produce Findings in Usability Testing. ACM Trans. Comput.-Hum. Interact. 26, 1, Article 3 (February 2019), 38 pages. https://doi.org/10.1145/3299096

Sasmito, G. W., Zulfiqar, L. O. M. and Nishom, M., 2019. Usability Testing based on System Usability Scale and Net Promoter Score, International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2019, pp. 540-545, doi: 10.1109/ISRITI48646.2019.9034666.

Sevilla-Gonzalez M. D. R., Moreno Loaeza L., Lazaro-Carrera L. S., Bourguet Ramirez B., Vázquez Rodríguez A., Peralta-Pedrero M. L., Almeda-Valdes P. 2020. Spanish Version of the System Usability Scale for the Assessment of Electronic Tools: Development and Validation. JMIR Hum Factors. 2020 Dec 16;7(4):e21161. doi: 10.2196/21161.

Sharfina, Z. and Santoso, H. B. 2016. An Indonesian adaptation of the System Usability Scale (SUS), International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2016, pp. 145-148, DOI: 10.1109/ICACSIS.2016.7872776.

Souto, V. T. (2013). Users' perceptions of the arrangement of links in government websites: an investigation using think-aloud and interview methods. InfoDesign - Revista Brasileira De Design Da Informação, 8(2), 01–14. https://doi.org/10.51358/id.v8i2.122

Stoyanov S. R., Hides L., Kavanagh D. J., Zelenko O, Tjondronegoro D, Mani M. 2015. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. JMIR Mhealth Uhealth 2015;3(1):e27. DOI:10.2196/mhealth.3422

Weichbroth, P. 2019. A mixed-methods measurement and evaluation methodology for mobile application usability studies. Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 101–106. DOI: 10.15439/2019F299.

Wohllebe, A., Ross, F. & Podruzsik, S. 2020. Influence of the Net Promoter Score of Retailers on the Willingness of Consumers to Install Their Mobile App. International Association of Online Engineering. Retrieved August 23, 2022 from https://www.learntechlib.org/p/218399/.

Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. 2019. The mHealth App Usability Questionnaire (MAUQ): Development and Validation Study. JMIR Mhealth Uhealth 2019;7(4):e11500. DOI: 10.2196/11500