Usability and User Experience, Vol. 194, 2025, 101–111
https://doi.org/10.54941/ahfe1006672

An Investigation of Influencing Factors in Co-Participation Usability Test

Ahmed Alghamdi[1,2], Fang Chen[1], and Kun Yu[1]

[1]School of Computer Science, Faculty of Engineering and IT, UTS, Sydney, NSW, Australia
[2]Systems and Networks Department, Faculty of Computing and Information, Al-Baha University, Al-Baha, 65779, Saudi Arabia

ABSTRACT

The rise to pre-eminence of digital interfaces in everyday life means that it is crucial that those interfaces are user-friendly. In turn, the development of user-friendly systems demands effective methods for testing their usability. Methodologies for usability testing have evolved from traditional approaches, focused on single users, to co-participation techniques. Whilst co-participation testing demands more resources, the approach fosters a testing environment that is experienced as more natural, especially by children and elderly participants, and tends to result in richer qualitative data. In comparison with single-participant testing, co-participation can result in more honest feedback and identify a greater number of minor usability issues. This literature review therefore considers co-participation usability testing in particular, exploring the factors influencing testing outcomes, specifically gender, age, participant team composition and whether or not an evaluator is present. It pulls together the findings of a large number of comparative studies, which employ various methodologies to understand the effect of these factors on the results of usability testing. From the analysis, it can be shown that gender differences lead to distinct preferences, styles of interaction and priorities for navigation efficiency and how information is set out. Age differences are critical, as older adults frequently need longer to carry out tasks and tend to face different challenges to those encountered by younger users, because of differences in motor skills and cognition. Participant team composition is a significant factor for testing outcomes, as groups of people with diverse levels of expertise are better at spotting usability issues. Lastly, where an evaluator is present in the testing session, participant behaviour may be influenced by virtue of the observer effect, which can reduce result validity. This study is, then, a comprehensive review, which contributes to knowledge by demonstrating the combined influence of the factors reviewed on testing outcomes and by revealing how they interact with one another. Successful usability testing therefore depends on care in addressing these factors when designing and running usability tests. Furthermore, future research could usefully seek to develop more inclusive approaches to testing, which take these factors into account in effective and efficient testing procedures. These findings will be of especial interest to web designers, software developers and digital marketing specialists, fields where user experience has a direct effect on the success of a product.

Keywords: Usability evaluation, Co-participation testing (CP), Human-computer interaction (HCI), User experience (UX)

INTRODUCTION

Digital interactions have become an omnipresent part of everyday life – everything from using laptops for work and teaching to advancing healthcare and having smart homes. With digital interfaces now ever present, providing functional, coherent, and intuitive experiences has become essential to successful technology design. Many obstacles posed by good design force stakeholders to work toward a common goal. Using usability testing to evaluate designs ensures the seamless implementation of new technologies and prospective advantages for future generations. By investigating these factors, in this study, we investigate collaborative usability testing, where multiple users participate in a system simultaneously to provide pertinent insights, without unwarranted intervention from the test administrator. This approach cultivates feedback that is more informative and in accordance with real-world scenarios than typical single-user testing methods. Various variables that influence collaborative usability testing, including team composition, the presence of an evaluator, and demographic characteristics (age and gender), influence the results of usability tests, which in turn affects the subsequent design of the artefact. This study aims to improve usability testing and expand the breadth of methodologies by illuminating the factors that affect interview situations and, subsequently, the outcome of usability tests. Emerging insights are expected to augment the design of digital artefacts in terms of access and user contentment. This is especially important for industries such as software development, web design, digital marketing, and the like that base their success on interactions with users. Enhancing usability testing methods can enable organisations to gain better insight on how to build products that are aligned with the expectations of users. Hence, by developing findings that are meaningful for practitioners, not only provides insights to contribute to academic debate, but can truly have an impact on shaping pivotal decision-making processes around the experiences of millions of interacting users.

This review begins by exploring the lifecycle of usability engineering. Next, the review considers the various approaches to evaluation, looking especially at co-participation techniques. After that, it evaluates the most important factors that influence the outcomes of usability testing. Finally, it presents comparative studies in order to illustrate the techniques' practical applications.

USABILITY TESTING

Given this context of digital interfaces being important in all aspects of daily life, it is important to be able to evaluate the quality of the user experience. This is done with a range of techniques known as usability testing, the basic ideas and methods of which are explored in this section. Usability testing is one of the most effective approaches to researching usability and user experience (UX), Researchers can obtain an understanding of how real humans identify and solve problems using a product or system. It can be used to discover usability issues and any deficiencies in terms of how people solve tasks. Additionally, it can be used to gain insights into what users of the

system want and need (Nielsen, 1994a; Rubin and Chisnell, 2008a; Dumas and Redish, 1999). The goal of usability testing is to gather both qualitative and quantitative data on the extent that users can successfully complete tasks with the product. There are various forms of usability testing including moderated, unmoderated and remote testing suitable to different design stages and project goals Dumas and Redish (1999). Because researchers interact with real users, this gives designers a chance to uncover insights that might otherwise have been overlooked in a heuristic evaluation (Nielsen Molich, 1990). For these reasons, usability testing should be a necessary part of ensuring that products are useful and usable but also fun and a pleasure to use (Norman, 2013). This in turn leads to increased end-user satisfaction and a high uptake of products, thereby leading to better market performance (Sauro and Lewis, 2016).

DESIGN AND USABILITY

The extent to which usability is considered in the design of products, services and systems is central to their user-friendliness and, therefore, their effectiveness. Norman (2013) seminal work on user-centred design which still informs the work of designers today – stresses the importance of avoiding over-complication in product design. According to Norman states, flexible interfaces that respond to individual users' needs and do not force users to adapt to a rigid design should be incorporated to foster intuitive user interaction. Likewise, Krug (2006) notes that digital platforms and websites should be intrinsically easy to use, thus fostering intuitive, seamless user experiences and increasing user satisfaction. International Organization for Standardization (1999) sets out a framework for the design of user-friendly interfaces which is referred to as the usability engineering lifecycle. This lifecycle consists of four fundamental activities:

1. To understand and identify the specific situation for which the product is needed;
2. To identify the needs of both the user and the organisation/institution;
3. To produce design solutions; and
4. To evaluate the design to ensure that it fulfils user/organisational needs.

It must be noted that this an iterative design process wherein the analysis, design, implementation and assessment phases should be repeated until all usability aims are met (see Figure 1)

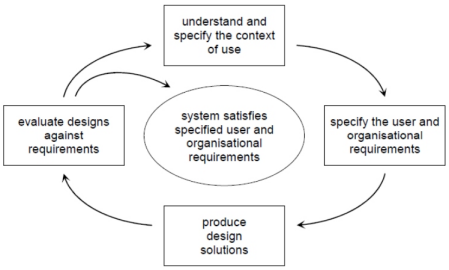

Figure 1: The usability engineering lifecycle (adapted from ISO 13407, 1999).

The current study concentrates solely on the evaluation stage of the usability engineering lifecycle. To that end, the following section examines various methods of evaluating usability.

EVALUATING USABILITY

Evaluating usability involves considering a combination of factors: namely, user interface, styles of user interaction, and the type of device being used (e.g. desktop computer, tablet, smartphone) being used Koutsabasis et al. (2007). As Sharp et al. (2019) point out, such evaluation should be carried out as early as possible because it is increasingly difficult and expensive to make changes at later stages, particularly once design and functionality are finalised. There are three basic categories of usability evaluation methods: expert-based methods, model-based methods and user-based methods Dillon (2001).

Expert-Based Methods

These methods make use of heuristic evaluation protocols, which means that experts in the field undertake thorough assessments, evaluated by means of established principles of usability (Molich & Nielsen, 1990). These approaches allow any inadequacies in the interface to be identified before ordinary users are involved, which in turn reduces risks associated with implementation (Nielsen, 1994b). The value of such pre-emptive approaches, in terms of reduced development costs, has been supported by research (Lewis, 1993).

Model-Based Methods

These methods employ predictive frameworks to assess interactions between users and the system under test. In particular, the GOMS framework (John & Kieras, 1996) and the Keystroke Level Model (Card et al., 1983) allow the quantitative evaluation of the efficiency of interaction before implementation.

User-Based Methods

Centring on user experience, user-based evaluation methods concentrate on obtaining user feedback on the functionality and ease of use of interfaces. Unlike model-based methods, which depend on preset frameworks and anticipating user behaviour, user-based methods produce observations of actual user behaviour to indicate user preferences and difficulties experienced when interacting with an interface. A key advantage of user-based methods is their ability to yield unexpected results by revealing how actual users interact with a system; as Nielsen (1993) stresses, even a limited number of test subjects may unanticipated usability issues. This category of usability evaluation methods includes interviews and surveys Soares et al. (2022), as well as several behaviour observation techniques which are detailed below.

Co-Participation Method

The co-participation method (CP), alternatively referred to as constructive interaction or the collaborative method Dumas and Redish (1999), is gaining popularity in the field of usability testing. This method uses two participants who co-operate in exploring the test object and carrying out tasks, while verbalising their thoughts aloud as they explore the system and interact with their partner. Introduced by O'Malley et al. (1985) in the mid-1980s to investigate how humans interacted with computers, the CP method is considered a valuable tool for helping testing to feel more natural for participants (Nielsen 1993; Van Den Haak et al., 2004). According to Nielsen (1993), the CP method works particularly well for usability testing involving children because it can more effectively encourage them to verbalise their thoughts than does the classic TA technique. While individual participants involved in usability testing are often hesitant to express criticism of an interface Hourcade (2007), CP creates an environment where participants are more comfortable expressing negative reactions Als et al. (2005). The disadvantage of the CP method, however, is that it requires two people per test, which makes it more expensive to implement and raises potential problems with recruiting adequate numbers of participants Als et al. (2005).

ORGANIZING USABILITY TESTS

Usability Testing Stages

Usability testing necessarily involves various stages. There are variations in this structure (e.g. Lazar, 2006; Rubin & Chisnell, 2008) (see Table 1), but the basics involve consistent procedural aspects, from participant recruitment to the administration and analysis of the test.

Table 1: Usability testing stages from two different sources.

Lazar (2006)	Rubin and Chisnell (2008)
Recruit participants representative of users.	Formulate a plan for the test.
Choose the setting for the test.	Prepare the test environment.
Choose the tasks for participants to perform.	Recruit suitable participants.
Choose the kind of data to be collected.	Get the test materials ready.
Preparation, e.g. obtaining informed consent.	Run test sessions.
Run the test.	Debrief participants afterwards.
Debrief participants afterwards.	Analyse the data observed.
Compile the findings and propose improvements.	Report the results and make recommendations.

The participant recruitment process is usually the responsibility of a moderator or usability expert (Dumas & Loring, 2008) and depends on stringent adherence to selection criteria to provide a sample that conforms to the required demographic profile, expertise in the specific

field and technical ability (Tullis & Albert, 2013). This is paramount, so that, although convenience sampling is generally considered acceptable for usability research, it is methodologically unsound to recruit such participants as students, where they do not match the target profile, unless they have suitable knowledge of the field and practical experience. There is only one exception to this and that is in the case of comparisons between cultures or languages, where discrete testing cohorts are required.

FORMATIVE AND SUMMATIVE TESTING

There are two main approaches to usability testing: formative and summative (Lazar et al., 2017c). The first of these, formative testing, takes place at an early stage of system or interface development and uses exploratory methods and simple, often paper-based prototypes, in order to generate explicit feedback from potential users (Snyder, 2003). Formative testing usually entails an informal style of interaction with participants and prioritises the collection of qualitative data, often using think-aloud protocols, on participants' perceptions of the design (Rubin & Chisnell, 2008). The second approach, summative testing assesses highly developed prototypes and evaluates more strategic design issues, frequently using task performance measurement and quantitative validation against benchmarks (Dumas & Fox, 2007). However, these are not absolute distinctions and there is debate about the boundaries, but there appears to be a consensus that formative testing is about identifying usability issues from observation and summative testing is about quantitative assessment of performance (Lewis, 2006). However, the evident flexibility in testing methods is appropriate to the need for usability testing to be adaptable to particular contexts and design objectives.

USABILITY TESTING SETTINGS

Various settings are suitable as usability testing environments. These range from dedicated laboratory facilities to remote communication. The laboratory option conventionally involves two rooms. Participants sit in one and the moderator and any observers in the other. Th rooms are provided with microphones, cameras, screen-recording facilities and, frequently, a two-way mirror. Figure 2 shows such an arrangement from Sauro (2018).

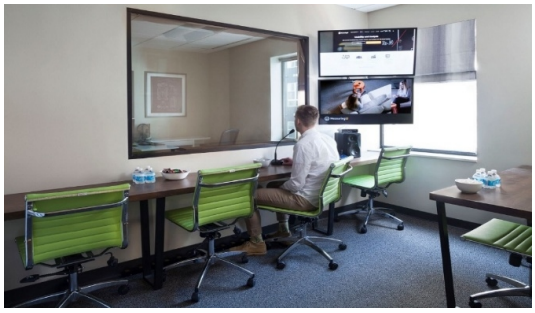

Figure 2: Usability testing laboratory (adapted from Sauro, 2018).

The laboratory arrangement works well testing on desktops and with mobile devices, but there can be practical issue which demand another solution. Testing at participants' workplaces or homes gives greater context veracity and allows for better participant access, especially for those with mobility issues (Lazar et al., 2017). The down side is that there are technical challenges and potential access issues for the moderator. Similarly, remote testing has become more important due to the availability of modern internet-based systems. Here, the pool of potential participants is expanded significantly and there is great potential for simultaneous testing. On the other hand, remote testing cannot capture contextual interactions and non-verbal cues as well, so that it is more suited to the quantitative assessment of finished products than early-stage exploratory testing of designs (Dray & Siegel, 2004). At the end of the day, what kind of testing environment is selected depends on such factors as data collection needs, participant access and the resources available. Each kind of setting has its own set of pros and cons for usability research.

FACTORS RELATED TO USABILITY TESTING

As many researchers have found, the process and results of usability testing can be influenced by a wide range of factors, the most significant being team composition, evaluator presence, age, gender, accessibility, design aesthetics, cultural considerations, environmental factors, and cumulative testing factors. Reviewed together, these studies highlight the importance of taking a holistic approach to usability testing that addresses as many of these factors as possible for maximum effectiveness and inclusivity. The following sections spotlight on these factors in details.

Team Composition, diverse teams with varied skills and backgrounds are more likely to identify a broader range of usability issues (Kuniavsky, 2003). Teams combining technical experts and UX specialists, along with both novice and experienced evaluators, provide more comprehensive evaluations (Nielsen, 2012; Rubin & Chisnell, 2008).

Evaluator Presence, the observer effect can significantly impact participant behavior during testing (Draper, 1999). While remote testing may reduce this effect, it presents other challenges such as technical issues and limited observation of non-verbal cues (Andreasen et al., 2007).

Demographic Factors Age, older users often face different challenges due to cognitive and motor skills variations (Sonderegger et al., 2016), requiring more time for task completion and different interface preferences (Czaja & Lee, 2007). Gender: Men and women may exhibit distinct interaction styles and preferences, with women typically prioritizing detailed information organization while men focus on navigation efficiency (Moss et al., 2006; Burnett et al., 2011).

Environmental Considerations, testing environment factors such as lighting, noise levels, and equipment consistency can significantly affect results (Nielsen, 1994). A controlled, comfortable testing environment is essential for reliable outcomes. Additional Influences - Accessibility requirements for diverse user capabilities (Schmutz et al., 2017) - Cultural and linguistic considerations for global audiences (Sauer et al., 2010) - Design aesthetics impact on perceived usability (Sonderegger & Sauer, 2010).

COMPARATIVE STUDIES

Building on the methodologies and design factors discussed above, comparative studies can reveal how theory is transformed into practice. Comparative findings show the complex interplay evident between the diverse design factors previously described. This is particularly true with team composition and the presence or absence of an evaluator.

Numerous comparative methodological studies have enriched the field of usability testing by exploring the practical implications, participant experience, and effectiveness associated with different methods. as these studies are relevant to my proposed study, a few examples will be discussed in details, Alhadreti and Mayhew's (2018) study assessing a library website demonstrates the value of CP for enhanced problem detection and user experience, although it does involve greater time and effort on the part of researchers. Of particular interest is Alhadreti's (2021) novel comparison of co-discovery (CD) and concurrent TA (CTA) methods in the context of Saudi Arabia, which comprehensively assesses problem identification, task performance, and participant experience. This study suggests that CD is overall the more effective method, as it identifies more diverse, small usability issues of layout and functioning and provides a more relaxing testing experience. However, the study finds no notable distinctions between CD and CTA concerning task performance, which implies that the advantages of CD are primarily qualitative (Alhadreti, 2021). These recent studies are consistent with prior research, including work by Van Den Haak et al. (2004), Als et al. (2005), and Adebesin et al. (2009), which emphasise the subtle trade-offs experienced when choosing between methods using single versus paired participants. Methodological considerations, not least the influence of how well participants know the system being tested and the "evaluator effect," are critical to the results of these studies. Adebesin et al.'s (2009) study of testing the usability of e-learning software stresses that it is vital for testing frameworks to be adjusted to the specific characteristics of a system's applications, and proposes a hybrid of TA and CD to be the most reliable means of providing a comprehensive understanding of usability issues.

Conversely, Sinabell and Ammenwerth's (2022) systematic review and case study of usability testing for digital health platforms aimed at the elderly concludes that a combination of remote testing and TA works best for elderly users. Based on a comparison of CD, CP, and a combination of remote user testing and TA, this study posits that CD and CP are less effective for usability testing with elderly participants. Furthermore, it offers 24 recommendations for dealing with the challenges identified in order to improve the usability, accessibility and acceptance of digital health platforms among elderly people.

The studies outlined here highlight the importance of methodological and contextual flexibility in arriving at reliable usability findings and refining usability testing methods. However, they also show the continually evolving nature of research in this area and the need for further research on the variables affecting the efficacy and efficiency of usability testing methods

There are direct, practical implications from these findings. 1) A careful consideration of team composition is needed in order to maximise issue detection. 2) There is a balance between the value of having the evaluator present and not restricting for natural user behaviour. 3) Both interface development and test design need to be consider differences in age and gender. 4) Reliable results depend on the ability to control for factors in the testing environment.

CONCLUSION

This review of the literature has considered the many theoretical and empricial factors affecting usability testing of digital interfaces. The review shows that there are many issues in addition to technical matters. The demographic profile of participants, team composition and the presence or absence of an evaluator are key determinants of success in testing. There are also still important gaps in the research, especially in respect of how multiple test variables interact and how the cultural context influences outcomes in an increasingly globalised HCI field. Key areas for future research include: 1) the development of standard testing protocols which are still adaptable to diverse participant characteristics, 2) the evaluation of testing methodologies over time and 3) the exploration of the impact on testing practices of emerging technologies. This literature review has contributed to research by pulling together the existing knowledge, as well as identifying topics needing further research. The overall conclusion is that there is a need for a holistic approach, which brings together human and technical factors in harmony to promote effective usability testing methods.

REFERENCES

Adebesin, F., De Villiers, R., Ssemugabi, S. (2009) 'Usability testing of e-learning: An approach incorporating co-discovery and think-aloud', Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists.

Alhadreti, O. (2021) 'Comparing the co-discovery and concurrent think-aloud methods in usability testing: a study of e-services in Saudi Arabia', Behaviour & Information Technology, 40(3), pp. 271–285.
Alhadreti, O., Mayhew, P. (2018) 'Rethinking thinking aloud: A comparison of three think-aloud protocols', Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.
Als, B. S., Jensen, J. J., Skov, M. B. (2005) 'Comparison of think-aloud and constructive interaction in usability testing with children', Proceedings of the 2005 Conference on Interaction Design and Children.
Andreasen, M. S., Nielsen, H. V., Schrøder, S. O., Stage, J. (2007) 'What happened to remote usability testing?: an empirical study of three methods', Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
Bevan, N. (2009) 'International standards for usability should be more widely used', Journal of Usability Studies, 4(3), pp. 106–113.
Burnett, M., Fleming, S. D., Iqbal, S., Venolia, G., Rajaram, V., Farooq, U., Grigoreanu, V., Czerwinski, M. (2011) 'Gender differences and programming environments: across programming populations', Proceedings of the 2011 ACM SIGSOFT Symposium.
Card, S. K., Moran, T. P., Newell, A. (1983) The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.
Dillon, A. (2001) 'The evaluation of software usability', in Karwowski, W. (ed.) Encyclopedia of Human Factors and Ergonomics. London: Taylor and Francis.
Dumas, J. S., Fox, J. E. (2007) 'Usability testing: Current practice and future directions', in Sears, A., Jacko, J. A. (eds.) The Human-Computer Interaction Handbook. New York: Lawrence Erlbaum Associates, pp. 1129–1149.
Dumas, J. S., Loring, B. A. (2008) Moderating usability tests: Principles and practices for interacting. Burlington, MA: Morgan Kaufmann.
Dumas, J. S., Redish, J. C. (1999) A practical guide to usability testing. Exeter: Intellect Books.
Hourcade, J. P. (2007) 'Interaction design and children', Foundations and Trends in Human-Computer Interaction, 1(4), pp. 277–392.
Hwang, W., Salvendy, G. (2010) 'Number of people required for usability evaluation: The 10±2 rule', Communications of the ACM, 53(5), pp. 130–133.
International Organization for Standardization (1999) ISO 13407:1999 Human-centred design processes for interactive systems. Geneva: ISO.
John, B. E., Kieras, D. E. (1996) 'The GOMS family of user interface analysis techniques: Comparison and contrast', ACM Transactions on Computer-Human Interaction, 3(4), pp. 320–351.
Koutsabasis, P., Spyrou, T., Darzentas, J. (2007) 'Evaluating usability evaluation methods: Criteria, method and a case study', Human-Computer Interaction. Interaction Design and Usability, pp. 569–578.
Krug, S. (2006) Don't make me think!: A common sense approach to web usability. Berkeley, CA: New Riders.
Lewis, C., Rieman, J. (1993) Task-centered user interface design: A practical introduction. Boulder, CO: University of Colorado.
Lewis, J. R. (2006) 'Usability testing', in Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics. Hoboken, NJ: John Wiley & Sons, pp. 1275–1316.
Nielsen, J. (1993) Usability engineering. San Francisco: Morgan Kaufmann.
Nielsen, J. (1994) 'Estimating the number of subjects needed for a thinking aloud test', International Journal of Human-Computer Studies, 41(3), pp. 385–397.

Nielsen, J., Landauer, T. K. (1993) 'A mathematical model of the finding of usability problems', Proceedings of ACM INTERCHI'93 Conference, pp. 206–213.

Norman, D. A. (2013) The design of everyday things: Revised and expanded edition. New York: Basic Books.

O'Malley, C. E., Draper, S. W., Riley, M. S. (1984) 'Constructive interaction: A method for studying human-computer-human interaction', Proceedings of INTERACT '84, pp. 269–274.

Rubin, J., Chisnell, D. (2008) Handbook of usability testing: How to plan, design, and conduct effective tests. 2nd edn. Indianapolis: Wiley Publishing.

Skov, M. B., Stage, J. (2012) 'Training software developers and designers to conduct usability evaluations', Behaviour & Information Technology, 31(4), pp. 425–435.

van den Haak, M. J., de Jong, M. D. T., Schellens, P. J. (2004) 'Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: A methodological comparison', Interacting with Computers, 16(6), pp. 1153–1170.