

USABILITY TESTING OF INTERACTIVE SYSTEMS CONDUCTED BY UNIVERSITY STUDENTS DURING A PANDEMIC

John Sören Pettersson
Karlstad Business School
Karlstad University
john_soren.pettersson@kau.se

Gunnar Olsson
Karlstad Business School
Karlstad University
gunnar.olsson@kau.se

Henrik Andersson
Karlstad Business School
Karlstad University
henrik.andersson@kau.se

Abstract

This case study describes a usability testing course in which students learn by practicing several evaluation methods. The on-campus format makes it possible for teachers and students to meet to discuss recorded test sessions and students can observe other students' execution of pilot studies conducted on campus. The COVID-19 pandemic placed new demands on this course. In-person activities were avoided by some students and many test participants. Some student teams tried remote usability testing. Interestingly, screen recordings (with sound) of the test sessions show that remote testing sometimes helped the students focus more on observation and less on (inappropriately) guiding the test subjects. Another effect was that the students found it easier to recruit participants than during the previous years when the university was teeming with students, lecturers, and non-academic staff. However, the recruited participants were often notably limited to the students' circles of friends.

Keywords: usability testing, prototyping, focus groups, remote testing, user participation

I. INTRODUCTION AND BACKGROUND

Usability tests, also called user tests, constitute a common type of testing in which the purpose is to evaluate the ease of use of various interactive systems with the help of people, who should preferably be representative of the target end users of the system, to obtain test results with high validity. Usability testing has no fixed format but can be described as a set of techniques or methods, with the common denominator being that potential users trial a system or prototype

campus course because there is certain equipment, especially an eye tracker, that students should use during the course and also because students need to interact with people outside of the course in the different usability tests. Thus, we have an opportunity to observe and guide students as course instructors, especially during the initial phases of each test. The COVID-19 pandemic necessitated remodelling the course to allow some students to conduct most assignments remotely. Recruiting participants at the deserted university would be a rather pointless chore. There was authorisation from the university management to conduct courses on campus, but not all the students or their test participants could be expected to want to come to the university. However, neither could the students expect participants to come to their homes.

This study describes our observations regarding the impact of partial physical distancing protocols on students' modus operandi. In contrast to the many papers in Van Slyke et al. [2021], this study brings up students' approaches to people *outside* a course. Few studies seem to have this focus. When external relations are brought up, the teacher is the actor. For instance, Harindranath and Panteli [2021] included industry speakers by video recorded interviews conducted by teachers, not by students. An interesting exception is Petersson [2021], who described internships in remote mode. While there are plenty of papers on remote work, there are few on remote learning.

Interaction Design, based on Benyon [2019], we include a couple of assignments in which a test is not constructed. Instead, real use is observed, and interviews are conducted with potential users to generate data on which to base a needs analysis.

Similar to a general HCI course, a usability testing course should discuss biases in evaluation within user tests, which is done in the two books mentioned earlier. It is essential that the practical aspect of the evaluation work is learned by doing practical coursework, which includes results analysis (i.e., data analysis). In our experience, students cannot understand qualitative data analysis merely by reading about warnings of biases in test design and the downsides of cherry-picking data [Dumas and Fox, 8 (dat)-1.1y(ox)Uat)-12.6 (c)-6s(,)-1.1 (8 (da(c)-8.1()-12.3 (and F)-5.4 (ox)-8 b/s2 T
bda127.916i2 b (des)-8-8.17 onse T02 Tw 19.494 0 T916 (s)17 T337i doi ysp(

The final examination for this course is conducted in regular examination halls at the university. However, the examination does not involve writing an essay on how to conduct the perfect usability study given a scenario XYZ. Instead, we provide a real test report from an actual research project or a commissioned work that is anonymised and salted and peppered with flaws. The students then play the role of the colleague who proofreads a report written by a teammate before it goes to the client. They check claims made in the report against data tables. There is also a requirement that they make at least eight references to the course book, the *Handbook* by Rubin and Chisnell [2008], which they are allowed to bring to the exam, with explanations of the applicability of the referenced statements to the exam report. The students get points for each identified issue in the report against which they provide a suitable argument. The issues identified may concern imprecisely described test objects or the test itself, or improper use of test data in the conclusions and recommendations made in the report.

IV. METHODS FOR EVALUATING THE EFFECT OF DISTANCE WORK

Several methods and sources have enabled the evaluation of the effect of a transition to a predominantly distance education format for this course.

The most important instrument with which to evaluate the COVID-19-invoked arrangements is the detailed and sometimes extensive test reports that each evaluation team authored. Assessing these reports gave us good insight into how the students structured and planned their work. However, when starting to write on this article in spring 2022, we only had access to the students of autumn 2021 to clarify when reports were not fully explicit on how testing was conducted.

We had sound and screen recordings of many of the evaluation sessions conducted by the students. When originally assessing and commenting the reports, these recordings made it possible to observe how participative the test moderator was during a recorded session, that is, if the moderator influenced the test participant in some way. The recording also revealed how well the data account in the report covers the sequence of events during the test session.

Course evaluations collected by an automated questionnaire system set up by the university several years ago ought to poll the opinions of the students on the course as a whole. However, the low response rates make us only briefly discuss response contents later: in 2021, 8 of 33 registered students responded to the automated survey, in 2020, 8 of 30, and in the year before the pandemic, 13 of 29.

Similarly, it would appear that the results of the final exam give some indication of learning effects, with a caveat, as the report given to students to correct cannot cover all evaluation design methods learned during the course. Interestingly, there are no differences in throughput for those students who passed all assignments during the course. The following student data include re-exams:

- 2021: 31 students completed all assignments, 30 attended written exams, and 29 passed the exams. (August 2022: the numbers include two re-exams.)
- 2020: 28 students completed all assignments, 27 attended written exams, and 25 passed as of 2021. The first written exam was a home exam, but a handful of students were ill with Covid-19. However, they passed the first re-exam in December of the same year.
- 2019: 29 students completed all assignments, 28 attended written exams, and 28 have passed (the last one in a re-exam in August 2022).
- 2018: 25 students completed all assignments, 25 attended written exams, and 25 have passed.
- 2017: 26 students completed all assignments, 26 attended written exams, and 25 have passed.

Therefore, exam results will not be used in subsequent discussions.

44 students enrolled, forming 12 teams, which strongly indicates there are no circulation among students of bad rumours about the course.

VI. REFLECTIONS ON LESSONS LEARNED

We scrutinised our course instances in autumn 2020 and autumn 2021 to be able to answer three questions:

1. What worked well?
2. What worked less well?
3. What is worth retaining for future instances?

It is obvious that the teams worked remotely with their user tests to various degrees. Digitally supported distance work and presentations are something that students were already used to, albeit to varying degrees for the autumn 2020 class. Before the pandemic, there were elements of distance learning in certain courses. In addition, the teachers have experience in distance education because the Information Systems subject has a distance learning undergraduate programme. The teachers have also worked with remote testing and remote co-design in their research [Petterson et al., 2019; Wik and Khumalo, 2020] and with mobile co-design with walking test participants both within this course and in other situations [Wik and Bergkvist, 2022]. Naturally, all these facts may have contributed to the smooth transition to remote testing. Moreover, based on the literature references in section II, there was no fear among instructors that the student teams would get very skewed results when they let people participate via Zoom.

During these two years, we learned that it is quite feasible to follow the students' work and to comment on mistakes thanks to screen recordings with sound, the students' written reports and the interaction during supervision and presentation, regardless of the location of each student. Question 1 is thus answered with 'basically everything' and question 2 with 'nothing', except that the requirement for campus presence for Assignment 4 makes the course unattractive for distance students.

What is worth keeping in future instances of this course? Remote usability testing was a growing method even before the pandemic and we should, of course, encourage students to try it out. However, it is hard to add a remote usability test as yet another assignment because the course is already content rich and many of the students do not plan to become usability experts. Instead, they take the course to have an orientation in this field, just as they take a course in interaction design immediately after this course to obtain the basics of user interface design principles. This helps them communicate in their professional careers with various stakeholders in development projects. We intend to allow groups to choose between remote and in-person tests in future instances of the course. The joint presentations will continue to be used for comparison between these two methods of performing usability tests.

The answers to questions 2 and 3 led us to rethink the assignment we felt we could not be run on distance because the needed equipment requires the students, as well as their test participants, to be on campus. Some students left or simply did not register because of this requirement. An alternative solution that we intend to try out is to have an assistant physically located in the lab, to manage both the equipment and test participants, while a team of students design the study and follow the test sessions via Zoom. Through remote control, the team can later use the equipment to do the analysis. Naturally, this will increase the burden on the assistant, but this arrangement can be restricted to students enrolled in the distance programme if the workload gets too costly.

Finally, one can question the selected course reading. Should a newer handbook not be used with a better account of remote testing? Unfortunately, we do not feel that the book by De Bleecker and Okoroji [2018] fully constitutes an alternative. The title of their book suggests a demarcation against anything other than remote usability testing and our students would not have time to read two books. Second, there are certain weaknesses in *Remote Usability Testing* [ibid.], weaknesses that we suspect one would have to look out for in any niched literature that advocates for a new technique. The black box Neisig [2014] warns against can, unfortunately, be found here. As an

Conference on Information Systems Education and Research, Article 1.
<https://aisel.aisnet.org/siged2021/1/>

Hattie, J. and H. Timperley (2007) "The Power of Feedback", *Review of Educational Research*, Vol. 77:1, pp. 81-112.

Hertzum, M. (2016) "A Usability Test is Not an Interview". *Interactions*, Vol. 23:2, pp. 82-84.

Moran, K. (2019) "Usability Testing 101", Nielsen Norman Group,
<https://www.nngroup.com/articles/usability-testing-101/>

Moran, K. and K. Pernice (2020) "Remote Moderated Usability Tests: How to Do Them", Nielsen Norman Group, <https://www.nngroup.com/articles/moderated-remote-usability-test/>

Neisig, M. (2014) "Transferring Methods to New Contexts", in Simonsen J., C. Svabo, S.M. Strandvad, K. Samson, M. Hertzum & O.E. Hansen (eds.), *Situated Design Methods*, Cambridge, MA: MIT Press, pp. 357-376.

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

Pettersson, J. (2021) "Student Perception of Learning During On-Site and Online Internships", *Proceedings of the 20th European Conference on E-Learning ECEL 2021*, pp. 364-370.

Pettersson, J.S., M. Wik, and H. Andersson (2018) "GUI interaction interviews in the evolving map of design research", in Paspallis, N. et al. (eds.) *Advances in Information Systems Development*, Cham: Springer, pp. 149-167.

Reeves, S. (2019) "How UX Practitioners Produce Findings in Usability(n U)0.9 .4 (.),e8 Tw 0.554au.2

s, ci2021) (A)D1.1 ()0.7 (al)3.6.4 (2()TJ3.6.3 (cie 20t)ne(,)-1.1 ()2l)3.6.4 (2J0 Tc 0 Tw08Td()T-1.145 Td[MC /P k4)643 Tw 0.59 0 Td[(i)3.1

ABOUT THE AUTHORS

Henrik Andersson is a PhD student with the Swedish national research school MIT—Management and IT since the end of 2021. Previously, he was in charge of the User Laboratory at Karlstad University, and he also worked as a teacher for several undergraduate courses in