

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 [Dumas and Fox, 2012, p. 1222]. Kate Moran of the Nielsen Norman Group summarily describes this type of testing as follows:

'In a usability testing session, a researcher (called a 'facilitator' or a 'moderator') asks a participant to perform tasks, usually using one or more specific user interfaces. While the participant completes each task, the researcher observes the participant's behavior and listens for feedback' [Moran, 2019].

It should be noted that there are few pure usability testing courses in undergraduate university curricula, although elements of user testing are quite frequently components of courses on human-computer interaction, interaction design, and user experience (including our *Interaction Design* course, which we run as both an on-campus and distance-learning course). However, a third-year course, *User Tests, Prototyping and Evaluation*, is entirely focused on different sorts of tests, conducted with potential users, and how tests should be reported. The course is an on-

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 Petterson [2021], who described internships in remote mode. While there are plenty of literature on on-line, on-campus and hybrid teaching modes, we contribute with some data on students' outward-facing activities when these partly went on-line.

The first instance during the pandemic of the user tests course ran in autumn 2020 and a second in autumn 2021, from the last week of August to the first week of November, ten weeks at half pace (i.e., students have one other course running concurrently). The study addresses the following three questions:

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

The scope is broadly defined to include student outcome, quality of the tests themselves, instructors' ability to supervise and follow-up, special physical arrangements, and workload. We begin by presenting our view of what should be included in a course on usability testing at the undergraduate level before sketching out our course, describing what we base our impact analysis on, and then presenting observations and reflections.

II. CONTENT FOR A COURSE IN USABILITY TESTING

One gets a rather good picture of what usability professionals regard as essential ingredients in user tests by studying some of the most referenced books on the subject internationally: *Handbook of Usability Testing* by Jeffrey Rubin and Dana Chisnell [2008], and *A Practical Guide to Usability Testing* by Joseph S. Dumas and Janice Redish [1999]. Both books provide a thorough description of usability tests, introduce and discuss the concept of *usability* and relate it to tests in which test users participate. The books describe the entire evaluation process from research questions on a product or service and its target audience(s), planning and conducting a usability test, and documentation and analysis of the results from the test. The content of this process is geared towards addressing specific open issues or providing progressive recommendations for further design development and user testing. Statistical methods are not included. In iterative design, points for evaluation should be conceived frequently, and the tests are typically used to discover flaws and generate ideas for improving the interaction design. The books also discuss and describe different usability testing techniques, when usability testing is preferred and when it is not, what a test environment should look like, and the critical characteristics of the different roles within an evaluation team.

A crucial purpose of a usability testing course should be to demonstrate the benefits of observing real people using systems (or prototypes), including gaining insight into specific human–computer interaction (HCI) problems of a particular interactive system. It is not enough that students learn the principles of good design and try to follow them. Design principles often have conflicting implications depending on how they are interpreted and implemented. In another course,

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, 2012; Reeves, 2019]. The course, which is described later, tries to capture as much as possible of the entire usability testing process, from planning a test and recruiting test participants to documenting, analysing, and drawing conclusions from the data generated. Admittedly, critiquing students' analyses is facilitated not only by screen recordings of the test sessions but also by one-on-one discussions of pilot studies between instructor and each student team. This became difficult when physical distancing protocols were in place, with some student teams preferring to conduct testing via Zoom from their homes rather than from campus and some students attending supervision meetings remotely. Before discussing the instance of the course during the pandemic, we present some notes on remote usability testing.

Remote Usability Testing

Usability tests are traditionally conducted with the moderator and the test user physically present together in the same geographical location. With developing technological possibilities, especially the enhanced last-mile capacity of the Internet, it has become meaningful to consider conducting moderated user tests remotely with the aid of screen sharing and other facilities [Thompson et al., 2004; Dumas and Fox, 2012]. Remote usability testing has some advantages: test participants can perform the test tasks in their natural environment; the need for travelling is reduced, which in turn simplifies test scheduling, lessens costs significantly, and may even lower the test's impact on the global climate (although usability testing is a much less common practice than it should be, and as such, this effect is probably not significant); and above all, one can recruit individuals who would otherwise not be invited to a test as pointed out by several authors [Thompson et al., 2004; Gardner, 2007; Schade, 2013; Smith 2017]. Such authors also acknowledge a commonly cited criticism of remote usability testing, which is primarily that it is difficult for the moderator to appraise the behaviour and body language of the test participants. This suggests that it is hard for the moderator to know when it is time for a probing or follow-up question. Nevertheless, comparative studies on effectiveness in identifying usability issues between moderated in-person and moderated remote usability testing have found no differences (e.g. [Thompson et al., 2004; Andreassen et al., 2007] and references in these).

It was not the pandemic that impelled serious discussions on remote usability testing. For instance, a survey conducted in 2018 found that 72% of 169 practitioners employing think-aloud protocols also used it in remote usability testing [Fan et al., 2020]. The development of web-based videoconferencing tools with screen-sharing capabilities turned remote usability testing into a focused topic in professional circles. *Remote Usability Testing* by Inge de Bleecker and Rebecca Okoroji, published in 2018, is the first book-length manual on the subject.

An even earlier development was the practice of making interactive prototypes accessible via web interfaces, which spared participants from downloading and installing complex software. This made user feedback on a grand scale possible as a moderator is unnecessary [Albert et al., 2010, p. 4]. Naturally, automatic gathering of click streams, questionnaires, and even screen recordings may not always compensate for the absence of an observing moderator who could ask participants to explain their behaviour when the rationale behind it is unclear to the researcher [Andreassen et al., 2007; Barnum, 2011, p. 45]. Consequently, continuous infrastructure development has made remote *moderated* usability testing more common [Moran and Pernice, 2020]. Typically, unmoderated testing aims for a high number of participants, whereas moderated testing aims to garner deep insight into what participants would want from a prototyped digital service and why participants misunderstand certain features of a prototype or an existing product,

as pointed out in earlier handbooks on remote testing [Albert et al., 2010; Bolt and Tulathimute, 2010; Barnum, 2011].

Margit Neisig [2014] discusses the challenges encountered when design methods are transferred to new contexts. Although she discusses design methods for sustainability, her ‘Ten prescriptive sentences on how to transfer design methods in a situated way’ [p. 372] can serve as guidelines for most processes or methods. For our purposes, we can use her normative sentences to reflect on outcomes when the students were largely left on their own to reinterpret recommendations for traditional usability testing into guidelines for remote testing. One drawback would be that they did not know the traditional approaches well enough to be proficient at this exercise. On the other hand, an advantage would be precisely that they were not moulded by traditional approaches. Such blindness, which Neisig warns against, most likely did not exist. When something works well, the inherent need to analyse the underlying motivation behind the antecedent actions disappears. The actions themselves replace the conceptual network that once motivated them—in this regard, Neisig references the black boxing analogy used by actor-network theorists [Neisig, 2014, p. 359]. Consequently, it becomes difficult to realise that all actions should be vetted when a design method, such as formative usability testing, is transferred to a new context.

III. COURSE ARRANGEMENTS

Because our student teams are expected to recruit test participants, this course has not been included in our distance education curriculum. Campus offers a broad and varied selection of potential end users, depending on which system designs we assign the students to analyse and develop further. The campus format also allows teachers to physically attend pilot testing. We also get the students to observe other students conducting pilot studies to make them aware of the significance of the role of the test moderator.

This course includes four usability (a.k.a. *user experience*) evaluations assigned to groups of 3–4 students: a plain test (which may involve ‘thinking aloud’ [Hertzum, 2016]), a video-prototyping exercise with a focus group, a Wizard-of-Oz, and an eye-tracking test. Each test is preceded by a pilot test in which the students trial their test design and social skills. The pilot studies also provide proof-of-concept (or evidence to the contrary) that their test design generates the sort of data they need to answer their research questions. The students have precious little time for each test, approximately two half-time weeks, to simulate real-world conditions. Usability testing is often a marginal phenomenon that is easily abandoned if designers or developers think it is complicated. Consequently, the requirement concerning pilot testing is limited to just one individual (i.e. one session). Real tests and repeated testing will make students increasingly aware of the utility of their test designs. For each test, we require five participants. As is well known, in an iterative design cycle, five participants are sufficient [Nielsen, 2000].

For each evaluation, each student team writes its own report, which is presented to the entire class or half the class for a general sharing and critiquing of all aspects of the conducted study. Furthermore, the course instructors review and comment on the test reports within a week as a formative assessment. Therefore, teams that were not overly sloppy would not need to rewrite the report (cf. the intertwined nature of ‘feed up’, ‘feed back,’ and ‘feed forward’ discussed by Hattie and Timperley [2007, pp. 86–90]). However, we follow up on the next reports to ensure that the same weaknesses are not present. All students can peruse all assessed reports. The test reports and the intermediate tutoring sessions make it possible for us to track the level of engagement in these assignments, even at the level of the individual students.

At IT consultancy companies, individuals are often expected to check the reports of colleagues before dispatching them to the client. Therefore, this course includes an assignment, between the first two evaluations and the last two evaluations, that requires students to review and comment on the test report of another evaluation team and present their observations on the reviewed report before the entire class. They must also present lessons learned based on their own report writing. Over the years, this has proved valuable to students. By critically scrutinising a report written by someone else, students become aware of issues, such as insufficient clarity and non-sequiturs, and they learn how to make improvements.

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.

During the course, several supervision meetings were held in which the evaluation teams could vet questions on test plans, planned or conducted pilot studies, and report writing. These meetings provided insight into how work proceeded in each team. Supervision meetings were held in hybrid mode: each student was allowed to participate in the format they desired. This approach was also used occasionally before the pandemic due to illness or students residing outside Karlstad. We observed certain events during the course in person, especially a test based on eye tracking, because the equipment for this was at the usability lab on campus.

V. OBSERVATIONS FROM TWO COURSE INSTANCES DURING THE PANDEMIC

Before explaining how the course was conducted during the COVID-19 pandemic, it should be noted that for 2019, for example, among the reports we examined for this case study, none of the teams explicitly stated that a test or a specific session was conducted remotely. However, in some instances, there may be deficiencies in the reporting on the type of data collection approach (online vs. offline), which is the type of issue we typically comment on in our written comments. Furthermore, the students probably accounted for the location of their data collection approach orally during their report presentations.

Table 1 presents our post hoc analysis of 2019, which is based only on instructor-reviewed reports. We excluded Assignment 3 from this analysis because it is the assignment in which the students assess the test report of another team. Assignment 3 does not include interactions with test participants, unlike the other assignments. Assignment 4 is the eye-tracking assignment conducted at the usability laboratory on the university campus.

Table 1: 2019 Test Location *In person* or *Remote* before the pandemic

2019	Assignment 1	Assignment 2	Assignment 4	Assignment 5
	<i>Conducted in lab</i>			
Group 1	In person	In person	In person	In person
Group 2	In person	In person	In person	In person
Group 3	In person	In person	In person	In person
Group 4	In person	Not apparent from the report	In person	In person
Group 5	In person	In person	In person	In person
Group 6	In person	Not apparent from the report	In person	Prob. In person
Group 7	In person	In person	In person	In person
Group 8	In person	In person	In person	In person

Because much of what we considered to be critical—that is, changes (deterioration) in tutoring during the pandemic due to less direct contact with many students—we begin our account with changes to the course design that affected the use of the usability laboratory. Normally, students taking this course conduct only a few test sessions in the lab. Instead, they use seminar rooms around the campus, including glass-walled seminar rooms to keep observers out of earshot of study participants. However, the execution of pilot studies often takes place in the lab. The same applies to the discussion of test data with the course instructors. In addition, our eye-tracking equipment is at the lab, which is why we conduct Assignment 4 at the lab, as indicated in Table 1.

Impact on the Test Implementation Environment and Supervision Rooms

The usability lab consists of three rooms connected in a sequence: *control room* → *test room* → *reception room*, with internal doors between each room and doors leading from each room out into the corridor. With these many doors, participants do not have to exit the same way they entered the test room. This means that test participants waiting in the reception room do not see previous participants leaving the test room. This setup was, of course, particularly useful during the pandemic, as the students could easily ensure that participants did not crowd around the doorways. Therefore, in the autumn of 2020 and 2021, testing with participants who were physically present took place primarily at the lab instead of in locations scattered across the university. Because it is usually university employees and students who are recruited as test subjects, it was a challenge that the university was very much deserted during the pandemic. The students had to conduct participant recruitment with a little more foresight, and the teams assigned members to accompany participants to the lab premises.

Figure 1 shows the test room in which the test participant sits during the actual test session. The test moderator is either in the test room or the control room and, together with observers, exercises oversight over what goes on in the test room via a *one-way window* (the side facing the test room behaves like a mirror). Communication equipment allows the participant and the test team to communicate. A mandatory part of the course requires physical presence on campus, during which course participants observe first-year students from our *Web Developer* programme conducting pilot studies in the usability lab. The control room is large enough to accommodate four people at the back of the room without any crowding.



Figure 1: Two test stations in the test room

Test subjects usually gather for a briefing in the reception room before a test. During the pandemic, however, the student teams did not assemble their test participants in this room, but had them sit outside the lab instead, where there is a larger space with study seats. In this way, they could maintain physical distancing in keeping with prevailing restrictions. However, students conducted post-test interviews in the reception room, as these are done individually (sometimes using a questionnaire). The reception room and the test room were also used for supervision discussions. The reception room has two TV screens on which students can present screen recordings and draft test reports during supervision.

To further minimise the risk of infection, hand sanitisers were placed in all three rooms. As shown in Figure 1, there were also disinfectants and tissue paper in the test room for wiping the keyboard, mouse, table surfaces, chair frames, and door handles. Many hand sanitisers and disinfectants were used. In the test room, face masks were also available for interested students and test subjects. However, they hardly used masks. This may be because the evaluation teams and the course instructors left the windows and the doors facing the corridors open during test and tutoring sessions, respectively. The circulation was apparently sufficient that people did not feel the needed to wear masks. Both course rounds were conducted during the first ten weeks of the autumn semester and the weather allowed intensive airing, which is not typical for our country.

Tests Conducted in the First Course Instance of the Pandemic: Autumn 2020

The first corona instance of the course was held early in autumn 2020. Table 2 presents data on how the students chose to conduct their tests: in-person or remote. As mentioned previously, one assignment compulsorily required on-campus activity, as the eye tracker was in the usability lab. The eye tracker is required not only for data collection but also for subsequent data processing. To use the lab, each team had to schedule a lab session. For Assignment 4, a specific schedule calendar was published on the course platform Canvas.

Table 2: 2020 Test Location *In person* or *Remote*

2020	Assignment 1	Assignment 2	Assignment 4	Assignment 5
Group 1	Remote	In person	In person	Remote
Group 2	In person	Remote	In person	Remote
Group 3	In person	In person	In person	In person
Group 4	In person	In person	In person	In person
Group 5	Remote	In person	In person	Remote
Group 6	Not apparent from the report	In person	In person	Remote
Group 7	In person	In person	In person	Partially Remote (2 observers + wizard remote)
Group 8	In person	In person	In person	Remote

As is apparent from Table 2, two teams performed Assignment 1 fully remotely, one team conducted Assignment 2 fully remotely, five teams conducted Assignment 5 fully remotely while a sixth team conducted Assignment 5 partially remotely. This may be attributable to students initially feeling a little uncertain about test arrangements and later choosing to perform their tasks remotely when they felt more secure. Furthermore, based on the university's recommendations, first-year students should be given priority for on-campus teaching, which may have influenced whether our third-year students worked remotely.

One team, Group 3, expressed concerns about the pandemic situation and tried to work remotely as much as possible. Group 3 participated in tutorials and presentations almost exclusively remotely. Paradoxically, they performed the test tasks with the test participants in person. For the other test teams, there were, of course, elements of remote work, even if the teams performed the tests in person. It is noteworthy that *in person* does not necessarily mean *on campus*. Moreover, some course participants may have supervised the in-person tests remotely via Zoom, as Group 7 states in their Assignment 5 report, although judging by the other reports, this is unlikely. Analysis of the collected data was most likely performed at home. (Table 2 is only about data collection during the different assignments and does not cover the entire work for each assignment.)

Presentations to the entire class (or half the class) by the student teams are crucial for the teams to learn from each other's mistakes and exemplary practices. Remote participation was always offered, but absenteeism was disallowed, except for illness. Supervision was offered (with some exceptions) with options for remote participation by individuals. Thus, some tutorial sessions were in a mixed format.

Tests Conducted in the Second Course Instance of the Pandemic: Autumn 2021

The second instance of the course during the pandemic was held in early autumn 2021. The course design was the same as that of the first pandemic instance of the course. Table 3 presents data on how the students conducted tests or test sessions, as both in-person and remote approaches could be adopted for different sessions of the same test.

Table 3: 2021 Test Location *In person* or *Remote*. Proportions refer to test participants and not the activity of team members

2021	Assignment 1	Assignment 2	Assignment 4	Assignment 5
Group 1	3/5 Remote	In person	In person	Partially Remote (2 observers remote)
Group 2	Remote	In person	In person	In person
Group 3	Remote	In person	In person	In person
Group 4	1/6 Remote	Remote	In person	Remote
Group 5	Remote	In person	In person	Remote
Group 6	In person	In person	In person	In person
Group 7	In person	In person	In person	In person
Group 8	3/7 Remote	In person	In person	In person

Six out of eight teams conducted the test in Assignment 1 fully or partially remotely. In contrast, interactions with test participants for Assignment 2 were conducted in-person by all teams except one. Finally, after the mandatory in-person test conducted on campus for Assignment 4, two teams conducted tests remotely for Assignment 5, and a third team had a couple of team members who observed the test remotely.

We observed earlier that for 2020, remote work became more common towards the end of the course. Arguably, we can say the opposite for 2021. The Omicron variant of COVID-19 had not yet been discovered, and the infection situation looked bright, which may have contributed to the reduction of physical distancing.

We should also note that before the commencement of the course in 2021, some students had applied for the course with the belief that this course and all other bachelor's courses would be taught entirely online due to the pandemic. Our department has an entire bachelor's programme offered in a

distance learning format. However, some of the local distance learning students choose to take part in the on-campus offering during their final year. In 2021, however, it was our distance learning students from remote parts of Sweden who applied for the user test course. The eye-tracking aspect of the course places certain constraints that demand physical presence, which discouraged a handful of applicants from actually enrolling in the course. Furthermore, the on-campus requirement also made one student decide to drop out partway through the course.

Other Observations

We were able to make several observations during the course rounds. The detailed test reports written after each evaluation during the course enable us to follow the commitment of individual students. The oral presentations of the reports after each completed assignment make it possible for us to direct questions to each team about the distribution of work, arrangements of data collection and analysis, lessons learned and so on. In addition, as the students often worked remotely, it became natural to ask what each one contributed during the supervision sessions; it was also natural for the students to add such comments to their reports and into initial test plans, which they uploaded to Canvas and often discussed with us.

The impression we instructors have is that the presentations of reports went well during the pandemic and so did the tutoring occasions, both in terms of the depth of the discussions around how the students chose to set up their data collection (in person or remotely) and the format of the presentations (hybrid mode). Overall, most groups conducted one to three tests remotely out of the four assignments that included data collection. In both years, two out of eight teams did not employ any remote interaction with test participants during data collection. Because of the full-class or half-class presentations, we are sure that all students got a sense of remote testing.

Both years we noted that the recordings of the test sessions showed that remote testing sometimes helped the students focus on observing. Inexperienced test moderators may sometimes be too eager to answer questions from test subjects or to help them in other ways. Remote testing seems to make inexperienced test moderators less likely to lead. This can be compared with a recent finding from another study with a focus on group interaction in online learning: '1) the threshold for contacting (or interrupting) other team members was raised' [Sjølie et al., 2022, p. 9]. In a usability testing course, the raised threshold benefits students. Whether this transfers to in-person testing later in life is an open question. Hopefully, an already acquired behaviour will feel natural to employ also later. However, the promise of remote testing to add actual use environments, as mentioned in section II, was not possible to fulfil here. In both years, the assignments focused on evaluating the university library's search functions for academic literature. Such use is found both at home and on campus, whether users are students or professors.

Another interesting but less surprising 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. There was, however, a notable drawback. The recruited participants were often decidedly limited to the circle of friends of the students. The campus environment, in contrast, offers a more varied population (in non-pandemic times, that is). Here was a notable lack of possibility to develop catch-as-catch-can skills, but then again, university campuses offer untypically easy working conditions for general recruitment, so training for more foresighted recruitment, as during the pandemic, has its advantages.

In the course evaluations mentioned earlier, students in this course often voice opinions about the course workload. This also happened in autumn 2020 and autumn 2021, but no one raised any specific distance-related factors as problematic or negative. Admittedly, the response rate is low and the answers are anonymous, so one cannot draw any firm conclusion on what the remote testing students felt about their modus operandi. However, the absence of remarks on this and the very voluntary nature of the choice between in-person and remote testing formats strongly suggest that many students found this an interesting alternative rather than a necessary burden imposed upon them by the pandemic in combination with our tight deadlines. For autumn 2022,

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

example, under the heading ‘The participant is not getting it’ (p. 81 in the chapter ‘Running a Remote Moderated Study’), the text explains that it is better to abort a session than continue when ‘it becomes clear that the session is a waste of time long before the session time is up’ [ibid.]. It is then pointed out that how this should be communicated to the participant depends on how far into the session one has come, but no examples are given of how to handle such delicate situations (a hint is given for the case when the session is continued after a fail [ibid., p. 84], but it is just a hint). A text for beginner moderators would need content about what makes participants feel valued when single tasks or whole sessions are aborted.

This last little analysis shows the value of being vigilant when new literature is offered within a new field, a field that, due to its newness, apparently lacks good coverage in existing standard works. To handle the present situation, we see no problem with the pedagogical arrangement in which the student teams choose themselves in a class with several teams. Presentations discussed by the entire class, and all reports available for all students of the class, make it possible for students to see the effects of different ways of conducting studies they have conducted themselves. We could possibly extend this with short reminder lists or video demonstrations. For the eye-tracking assignment, we will have to elaborate on a method to have the students control the teachers’ handling of equipment and participants in the usability lab, while we simultaneously carefully guide the students.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the input from Malin Wik on the literature discussed. A shorter version in Swedish was published by the university’s unit for pedagogical development.

LIST OF REFERENCES

- Albert, W., T. Tullis, and D. Tedesco (2010) *Beyond the Usability Lab: Conducting Large-scale Online User Experience Studies*. Boston, MA: Morgan Kaufmann.
- Andreasen, M.S. et al. (2007) “What Happened to Remote Usability Testing? An Empirical Study of Three Methods”, *Proceedings of Conference on Human Factors in Computing Systems 2007 (CHI 2007)*, pp. 1405-1414.
- Barnum, C.M. (2011) *Usability Testing Essentials*. Burlington, MA: Morgan Kaufmann.
- Benyon, D. (2019) *Designing User Experience* (4th ed.; 2014 *Designing Interactive Systems* 3rd ed.). Harlow, UK: Pearson Education.
- De Bleecker, I. and R. Okoroji (2018) *Remote Usability Testing*. Birmingham, UK: Packt Publishing Ltd.
- Bolt, N. and Tulathimutte, T. (2010) *Remote Research: Real Users, Real Time, Real Research*. Brooklyn, NY: Rosenfeld Media.
- Dumas, J.S. and J. Fox (2012) “Usability Testing” In Jacko, J.A. (ed.) *The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (3rd ed). Boca Raton, FL: CRC Press, Taylor and Francis, pp. 1221-1242.
- Dumas, J.S. and J.C. Redish (1999) *A Practical Guide to Usability Testing* (Revised ed.) Exeter, UK: Intellect Books.
- Fan, Shi et Truong (2020) “Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey”, *Journal of User Experience*, Vol. 15:2, pp. 85-102.
- Gardner, J. (2007) “Remote Web Site Usability Testing - Benefits Over Traditional Methods”, *International Journal of Public Information Systems*, Vol. 3:2, pp. 63-73.
- Harinadranath, G. and N. Panteli (2021) “Engaging Industry Speakers in the Newly Transitioned Virtual Classroom During COVID-19”, *Proceedings of the 2021 AIS SIGED International*

- 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.
- Petterson, 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 Testing", *ACM Transactions on Computer-Human Interaction*, Vol. 26:1, Article 3, pp. 1-38.
- Rubin, J. and D. Chisnell (2008) *Handbook of Usability Testing, Second Edition: How to Plan, Design, and Conduct Effective Tests*, Indianapolis, Wiley Publishing Inc.
- Schade, A. (2013) "Remote Usability Tests: Moderated and Unmoderated", Nielsen Norman Group, <https://www.nngroup.com/articles/remote-usability-tests/>
- Sjølie, E., T. Ch. Espenes, and R. Buø (2022) "Social Interaction and Agency in Self-organizing Student Teams During their Transition from Face-to-face to Online Learning", *Computers & Education*, Vol. 189, Article 104580, pp. 1-12.
- Smith, C. J. (2017) "Getting the Most Out of Remote Research and Testing", *Interactions*, Vol. 24:2, pp. 82-84.
- Thompson, K.E., E.P. Rozanski, and A.R. Haake (2004) "Here, There, Anywhere: Remote Usability Testing That Works", *CITC5 '04: Proceedings of the 5th Conference on Information Technology Education*, pp. 132-137.
- Van Slyke, C., H. Topi, and M.J. Granger (2021) "Special Section: COVID-19, Learning, Pedagogy, and Educational Systems", *Communications of the Association for Information Systems*, 48, pp. 476-486. <https://doi.org/10.17705/1CAIS.04841>
- Wik, M. and L. Bergkvist (2022) "Exploring Mobile Co-design in the Context of Use: Continuous Elicitation and Evaluation of Design Suggestions", *HCI 2022. Lecture Notes in Computer Science*, LNCS 13302, Springer, pp. 324-342.
- Wik, M. and A. Khumalo (2020) "Wizardry in Distributed Participatory Design. From design to implementation", *HCI 2020, Lecture Notes in Computer Science*, LNCS 12181, Springer, pp. 172-186.

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 Information Systems before he became a research student. He is also the programme coordinator for the undergraduate distance programme in Information Systems.

Gunnar Olsson has taught undergraduate courses in Information Systems since 2017, especially process development, systems development, and usability. Together with Pettersson, he developed a course on the accessibility of digital services and documents. Before coming to Karlstad University, he had worked for twenty years in technical communication.

John Sören Pettersson is a full professor in Information Systems at the Business School at Karlstad University, specialising in user-centred systems development and usability. In particular, within the areas of ICT for Development, IT support for crisis management training, and usable privacy (i.e. usability questions around privacy-enhancing technologies and accountability for cloud computing). For more than 20 years, he has been involved in distance education in parallel with his engagement in campus education.