Evaluation in Human-Computer Interaction – Beyond Lab Studies

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Florian Alt
Bundeswehr Universität München
Munich, Germany
florian.alt@unibw.de

Ville Mäkelä
LMU Munich
Munich, Germany
ville.maekelae@ifi.lmu.de

ABSTRACT

Many research contributions in human-computer interaction are based on user studies in the lab. However, lab studies are not always possible, and they may come with significant challenges and limitations. In this course, we take a broader look at different approaches to doing research. We present a set of evaluation methods and research contributions that do not rely on user studies in labs. The discussion focuses on research approaches, data collection methods, and tools that can be conducted without direct interaction between the researchers and the participants.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**; **Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

Evaluation Method, Remote Studies, Datasets

ACM Reference Format:
Albrecht Schmidt, Florian Alt, and Ville Mäkelä. 2021. Evaluation in Human-Computer Interaction – Beyond Lab Studies. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3411763.3445022

1 INTRODUCTION

How do you know that your ideas for new forms of interaction, the novel interaction concepts you thought of, the alternative user interfaces your implemented, or the new interaction paradigms developed are useful and beneficial for the user? Easy – evaluate them! Looking at publications in ACM SIGCHI, most evaluations are based on studies with users, typically in a lab context. In this course we look at alternative approaches to evaluation in human-computer interaction, especially taking into account that you may not have the ability to directly interact with users.

While this course was promoted by the shut-downs that resulted from the COVID-19 pandemic, we think compiling these methods has value beyond the current situation, for example, if researchers can not interact with users for other reasons, such as health, travel or financial constraints. We discuss a number of alternatives to lab-based user studies we experienced. This collection is a work in progress, and there are other options we do not cover here.

CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05.
https://doi.org/10.1145/3411763.3445022

2 SCOPE: TOPICS AND APPROACHES

The following research approaches and data collections methods will be covered in this course.

2.1 Use Existing Data Sets

Over the recent years, many researchers have collected datasets and published them. The motivation for publishing data sets is to allow replication of research results as well as to enable secondary use of collected data to investigate further research questions. Here, it is important to look at the original data and ensure that the published data allow this type of use. Another point is that research should avoid fishing for results in data sets. When you have a specific research question or objective to investigate, it is useful to look at the bigger picture. Exactly the data you would collect yourself is probably not available, but it is quite possible that there are data sets that are similar to what you planned or where a subset of the published data fits your research question. The following resources are good starting points:

- Google data set search: https://datasetsearch.research.google.com/
- Awesome public data sets: https://github.com/awesomedata/awesome-public-datasets
- Government data sets, e.g. for Germany: https://www.govdata.de/

2.2 Create Prototypes That Can Be Experienced Remotely

Often, research prototypes we build can easily be appropriated for access and use from a remote location, for example, a web-based prototype that participants can access via the browser on their desktop computer or from their smartphone. Enabling participants to remotely experience these prototypes is as easy as providing them an URL.

For future projects, researchers might want to discuss in an early stage whether building a prototype by means of a web technology is feasible. This is also useful for mobile applications. Tools, such as PHP Mobile Detect, can be used to ensure that participants are experiencing the prototype on a mobile device. An example of a web-based prototype is the Android lock pattern screen [10] created to collect user-defined credentials.

Whereas such prototypes can be stand-alone applications, another approach could be to integrate them with survey platforms.

Many platforms provide sophisticated means to include interactive functionality, including LimeSurvey[1] and SoSci[2], both of which enable researchers to write PHP and JavaScript code. This also has the advantage that experiencing a prototype and collecting data can be tightly integrated (e.g., a questionnaire on usability or UX).

2.3 Piggyback Your Experiment Into Apps Or Web Pages

Interaction on widget scale may be independent of the specific application the widget is used in. Text input, gestures in space or on the screen, button presses, or mouse movements are elementary to many interactions. Instead of creating a test apparatus that is specifically targeted to your next text input method or your next gestural interaction, you can create a mobile phone app or web page that this interaction is an integral part of. Users that use this application or web page, provide inevitably input to the system and hence create data for collection.

One important aspect is to inform users about which data is collected and shared, and obtain their consent. Distribution of the apps can be done through app stores or your institution's website.

The major concern with this approach is the very limited control the researcher has over the experiment. There is little knowledge about the participant and how they conduct the study, e.g., do they use the phone on a bus or in bed? Examples show that this can be compensated by a comparably large number of participants, e.g. running a Fitts' law experiment with 10.000 users instead of 20.

Many examples exist where this method has been used, most importantly typing behavior [3], notifications [7], and authentication mechanisms [8]. Experiences from these kinds of studies were reported by [1, 4].

2.4 Run Studies in Virtual Reality

Virtual reality provides an opportunity to recreate your research environment virtually and let your participant access this from home. The researcher thereby is not limited to a lab environment but can rebuild arbitrary settings, including but not limited to public spaces, cars, homes or work environments. The environment can be made accessible through a number of different ways, for example, as a standalone application that is available to participants as download, or through online platforms such as Mozilla Hubs.

This approach supports a wide variety of data collection methods, such as logging interaction in the VR environment, letting participants do a talk-aloud walk through, interviewing them afterwards, or filling in a questionnaire in VR [9]. An investigation of this approach alongside a discussion on challenges and pitfalls is provided by Mäkelä et al. [5].

2.5 Engage With Users Through Remote Communication

Many interview-based studies that we commonly conduct in the lab can easily be conducted using remote communication technologies. This includes doing individual interview studies via Skype or running focus groups in Zoom. Also more sophisticated forms are possible that allow studying how people behave in a certain environment while concurrently observing and questioning them.

To observe users during remote tasks, the screen share features of Skype or Zoom can be used. In addition, there are several platforms available that support specific tasks, e.g., Discord for gaming or Repl.it for programming tasks.

2.6 Appropriate Your Research Question and Method to Users Who Are at Home and Where You Do Not Have Direct Access

Whereas lab environments provide the opportunity for investigations under controlled conditions (e.g., typing speed, number of errors), studying users in their natural environment creates many opportunities and challenges. On one hand, researchers may have the unique opportunity of understanding how their concepts and prototypes will be used in real life and obtain rich insights into the users' experience, their behavior, acceptance, and concerns (e.g., with regard to privacy). On the other hand, challenges include that researchers might have little control over when participants work on assigned tasks, and they might get distracted by other inhabitants, TV, or social media. Hence, researchers might want to assess the context in which participants work on their tasks. Another challenge might be compensation. In particular, if participants are residing in other countries, incentives such as online vouchers might not work. However, platforms such as PayPal work in many countries but researchers should ensure that means for remuneration comply with their institutions' regulations.

2.7 Supply Study Equipment To Users At Home

Much of our research entails letting users try out prototypes. Rather than doing this in the lab, participants could be sent the prototypes. This creates a number of challenges. Prototypes need to be easy to set up and operate. Therefore, researchers could create web-based tutorials and instructions or provide other channels through which participants can receive support, for example, Slack, WhatsApp, email or any video conferencing tool. In cases where prototypes cannot simply remain with / be disposed of by participants, another challenge arises from making the return of prototypes as easy as possible. This should create as little effort as possible, e.g., prearranging the pick-up at the participant's home.

An example where smartwatches in combination with NFC tags have been used to observe people's behavior with their smart home devices is provided by [6].

2.8 Use Analytic and Computational Evaluation

For specific questions it may be possible to evaluate a system by careful analysis or computational evaluation. Approaches like GOMS/KLM (and its many variants) allow systems to be compared and can be used to show that one approach is more efficient than another – in particular regarding the number of basic operations to be performed. Using computational approaches it is possible to prove that certain designs are more efficient than others, e.g. for text input [2].

[1] Lime Survey: https://www.limesurvey.org/
[2] SoSci: https://www.soscisurvey.de/

2.9 Study Phenomena that Happen Online

If you are open to what you want to study, you can also move the focus of your research to things that can be observed online and only online. Examples include research on crowdsourcing, social media use, or collaboration in online games.

2.10 Summary

The list shows many opportunities that are real alternatives to classical lab studies. As a field, it may be beneficial to consider the question of how we can conduct evaluations with a much broader view. This may make research more inclusive and it is likely to provide a more holistic view of the challenges and solutions.

3 INTENDED AUDIENCE

This course addresses anyone interested in conducting evaluations beyond lab studies. The course assumes basic knowledge in conducting user studies. Participants can expect to concretely benefit from the course by gaining an overview and basic understanding of out-of-the-lab research approaches, their opportunities and pitfalls, so as to facilitate the approaches' application in their own research and practical projects. To facilitate this, the course provides concrete takeaways and material, including practical examples and pointers to further relevant work.

4 PREREQUISITES AND PREPARATION

The course is self-contained and assumes a general knowledge and interest in methods in human computer interaction. The information presented is on an overview level and aims at helping to identify appropriate approaches to research. Learning about the details of the specific methods is beyond the scope of this course and requires further reading.

5 COURSE OVERVIEW

The course is presented in one module of 75 minutes length.

(1) Motivation: Why look beyond lab studies?
(2) Overview of different research approaches
(3) Discussion in breakout groups to identify positive and negative aspects
(4) Deep dive into several specific methods
(5) Concluding discussion

6 RESOURCES

All of the presentation material for this course will be made available at:https://hci-lecture.org/methods/.

7 INSTRUCTORS

Albrecht Schmidt is a full professor of Human Centered Ubiquitous Media in the Computer Science Department at LMU Munich. The focus of his work is on novel user interfaces to enhance and amplify human cognition. He is working on interaction techniques and intelligent interactive systems in the context of ubiquitous computing. In 2018 he was elected to the ACM CHI Academy.

Florian Alt is an full professor of Usable Security and Privacy at the Bundeswehr University Munich. His research focuses on how security can be made an integral part of user-centered design and how ubiquitous computing technologies can help building systems that are both usable and secure. Specific research topics include but are not limited to behavioral biometrics, usable security in smart homes, security concepts based on physiological data, and secure AR and VR interfaces. He chaired the CHI subcommittee on *Usability and User Experience* in 2020 and 2021.

Ville Mäkelä is a postdoctoral researcher in the Human Centered Ubiquitous Media group at LMU Munich. His research focuses on virtual reality, ubiquitous computing, and games. His primary research topic is about how virtual reality can be utilized for research purposes, including how studies can be conducted remotely during a pandemic. He is an associate chair (AC) in the *Usability and User Experience* subcommittee at CHI '21.

ACKNOWLEDGMENTS

This course is partly funded the EU (ERC Project AMPLIFY, no.683008) and by BMBF (Hive-lab, 16SV8183).

REFERENCES

[1] Susanne Boll, Niels Henze, Martin Pielot, Benjamin Poppinga, and Torben Schinke. 2011. My App is an Experiment: Experience from User Studies in Mobile App Stores. *Int. J. Mob. Hum. Comput. Interact.* 3, 4 (Oct. 2011), 71–91. https://doi.org/10.4018/jmhci.2011100105

[2] Anna Maria Feit. 2019. SIGCHI Outstanding Dissertation Award: Assignment Problems for Optimizing Text Input. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3290607.3313773

[3] Niels Henze, Enrico Rukzio, and Susanne Boll. 2011. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 133–142. https://doi.org/10.1145/2037373.2037395

[4] Niels Henze, Alireza Sahami Shrazi, Albrecht Schmidt, Martin Pielot, and Florian Michahelles. 2013. Empirical Research through Ubiquitous Data Collection. *Computer* 46, 6 (June 2013), 74–76. https://doi.org/10.1109/MC.2013.202

[5] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376796

[6] Sarah Prange, Christian Tiefenau, Emanuel von Zezschwitz, and Florian Alt. 2019. Towards Understanding User Interaction in Future Smart Homes. In *Proceedings of CHI '19 Workshop on New Directions for the IoT: Automate, Share, Build, and Care* (Glasgow, UK) *(CHI '19 Workshop)*. ACM, New York, NY, USA, 5 pages. https://www.unibw.de/usable-security-and-privacy/publikationen/pdf/prange2019iot.pdf

[7] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-Scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[8] Stefan Schneegass, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2014. SmudgeSafe: Geometric Image Transformations for Smudge-Resistant User Authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 775–786. https://doi.org/10.1145/2632048.2636090

[9] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300590

[10] Emanuel von Zezschwitz, Malin Eiband, Daniel Buschek, Sascha Oberhuber, Alexander De Luca, Florian Alt, and Heinrich Hussmann. 2016. On Quantifying the Effective Password Space of Grid-Based Unlock Gestures. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (Rovaniemi, Finland) *(MUM '16)*. Association for Computing Machinery, New York, NY, USA, 201–212. https://doi.org/10.1145/3012709.3012729