Evaluation in Human-Computer Interaction - Beyond Lab Studies

Albrecht Schmidt[1] (Ludwig-Maximilians-Universität München)
Florian Alt[2] (Universität der Bundeswehr, München)

Abstract. In this paper we present a set of approaches to evaluation in human computer interaction that offer an alternative to lab studies. The discussion focuses on research approaches, data collection methods, and tools that can be conducted without direct interaction between the researchers and the participants.

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 is useful and beneficial for the user? Easy - you do an evaluation! Looking at publication in ACM SIGCHI, the majority of evaluations presented is based on studies with users, typically in a lab context. In this paper 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.

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

Using Existing Data Sets

Over 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

---

[1] albrecht.schmidt@ifi.lmu.de

[2] florian.alt@unibw.de

original data and ensure that the published data allow this type of use. A further 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 for looking at existing data:

- 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/

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[3] 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 von [Zezschwitz et al., 2016] 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 quite sophisticated means to include interactive functionality, including LimeSurvey[4] and SoSci[5], 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 user experience).

Piggybacking 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

---

[3] http://mobiledetect.net/

[4] https://www.limesurvey.org/

[5] https://www.soscisurvey.de/

app or web page where 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 being collected and shared and obtain their consent. Distribution of the apps can either be done through AppStores or also 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 the use the phone while on a bus or while in bed. Examples have shown that this can be compensated by a comparably large number of participants that take part, 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 [Henze et al., 2011a], notifications [Shirazi et al., 2014], and authentication mechanisms [Schneegass et al., 2015]. Experiences from this kind of studies were reported by [Henze et al., 2011b; Henze et al., 2013].

Run Studies in Virtual Reality

Virtual reality provides an opportunity to recreate your research environment in virtual reality 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 also as Online plattform, such as Mozilla Hubs[6].

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 [Schwind et al., 2019].

An investigation of this approach alongside a discussion on challenges and pitfalls is provided by [Mäkelä et al., 2020].

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 that allow people to study how people behave in a certain environment while concurrently observing and questioning them are possible.

---

[6] https://hubs.mozilla.com/#/

To observe users during remote tasks, the screen share features of Skype or Zoom could be used. In addition, there are several platforms available that support specific tasks, e.g., Discord[7] for gaming or Replit for programming tasks[8].

Appropriate Your Research Question And Method To Users Who Are At Home And Where You Have Not 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 on 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, that 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.

Supply Study Equipment To Your Users At Home

Much of the research we are conducting in HCI 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, for example, pre-arranging the pick-up at the participant's home.

An example where smart watches in combination with NFC tags have been used to observe people's behavior with their smart home devices is provided by [Prange et al. 2019]

---

[7] https://discordapp.com/
[8] https://repl.it/

Analytic and Computational Evaluation

For specific questions it may be possible to evaluate a system by careful analysis or computational evaluation. Simple approaches like GOMS/KLM (andy its many variants) allow a comparison between systems and can be used to show that one approach is more efficient than another - in particular with regard to the number of basic operations that need to be performed.

Using computational approaches it is possible to prove that certain designs are more efficient than others, e.g. for text input [Feit, 2018].

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.

Conclusion

In this short paper we present an overview of approaches to doing research and in particular evaluation in human-computer interaction, without direct access to the user. The list shows many opportunities that are real alternatives to classical lab studies. As a field it may be beneficial to consider the question how to evaluate with a much broader view. This may make research more inclusive and it is also likely to provide a more holistic view of the challenges and the respective solutions.

References

[Feit, 2018] Feit, Anna Maria. Assignment Problems for Optimizing Text Input. Aalto University publication series DOCTORAL DISSERTATIONS, 103/2018

[Henze et al., 2011a] Henze, Niels, Enrico Rukzio, and Susanne Boll. "100,000,000 taps: analysis and improvement of touch performance in the large." Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. 2011.

[Henze et al., 2011b] Henze, N., Pielot, M., Poppinga, B., Schinke, T., & Boll, S. (2011). My app is an experiment: Experience from user studies in mobile app stores. International Journal of Mobile Human Computer Interaction (IJMHCI), 3(4), 71-91.

[Henze et al., 2013] Henze, N., Shrazi, A. S., Schmidt, A., Pielot, M., & Michahelles, F. (2013). Empirical research through ubiquitous data collection. Computer, (6), 74-76.

[Mäkelä et al., 2020] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt und Florian Alt. Virtual Field Studies: Conducting Studies on

PublicDisplays in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20. ACM, New York, NY, USA

[Prange et al., 2019] S. Prange, C. Tiefenau, E. von Zezschwitz, and F. Alt, "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, New York, NY, USA, 2019.

[Schneegass et al., 2015] 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 (UbiComp '14). Association for Computing Machinery, New York, NY, USA, 775–786. DOI:https://doi.org/10.1145/2632048.2636090

[Schwind et al., 2019] 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 (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 360, 1–12.

[Shirazi et al., 2014] Sahami Shirazi, Alireza, et al. "Large-scale assessment of mobile notifications." Proceedings of the SIGCHI conference on Human factors in computing systems. 2014.

[Zezschwitz et al., 2016] 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(MUM '16). Association for Computing Machinery, New York, NY, USA, 201–212. DOI:https://doi.org/10.1145/3012709.3012729