The HCI User Studies Toolkit: Supporting Study Designing and Planning for Undergraduates and Novice Researchers in Human-Computer Interaction

Valentin Schwind
Frankfurt University of Applied Sciences
Frankfurt am Main, Germany
valentin.schwind@fb2.fra-uas.de

Stefan Resch
Frankfurt University of Applied Sciences
Frankfurt am Main, Germany
stefan.resch@fb2.fra-uas.de

Jessica Sehrt
Frankfurt University of Applied Sciences
Frankfurt am Main, Germany
jessica.sehrt@fb2.fra-uas.de

ABSTRACT

Planning and organizing user studies can be a challenge for novices and student researchers in human-computer interaction (HCI). We present a web-based study planning toolkit for researchers to support their planning and organization with human subjects. The free and open toolkit contains multiple interactive research applications for (1) the selection of quantitative and qualitative research methods in HCI, (2) automated study design planning for experimental user studies, (3) an automated informed consent generation for studies with human subjects, (4) calendar booking for experimental user studies in shared laboratories, and (5) automated participation confirmation certificates. The toolkit can be helpful for Ph.D. students in supervising novices and undergraduates who need support with their planning and study design choices. The toolkit was actively used over two semesters in multiple HCI lectures and formatively evaluated. We contribute with the whole toolkit and encourage other researchers to contribute and further improve the content.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in HCI**; **HCI design and evaluation methods**.

KEYWORDS

Study Design Planning, HCI Toolkit, User Studies, Organization

ACM Reference Format:
Valentin Schwind, Stefan Resch, and Jessica Sehrt. 2023. The HCI User Studies Toolkit: Supporting Study Designing and Planning for Undergraduates and Novice Researchers in Human-Computer Interaction. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585890

1 INTRODUCTION AND BACKGROUND

Conducting user studies is a crucial part of research in human-computer interaction (HCI) and of the education of undergraduate students in the field. This allows young researchers to learn how to gather data and how people interact with novel technologies and their prototypes. However, despite the supervision of the student research projects, the planning, and organization of scientific experiments for the first time can be extremely challenging due to a lack of experience or time for screening literature. Student researchers and novices supposed to conduct a user study in HCI then often struggle with selecting appropriate research methods, designing user studies, generating informed consent documents, and scheduling experimental sessions [6, 14, 20]. The overwhelming amount of research methodologies in HCI alone causes difficulties for inexperienced researchers to narrow these options down. Typically, research methods are the subject of every HCI lecture, but undergraduates or novice students conducting an HCI user study rather desire to know which is the best method or study design to answer *their* research question.

Conducting a deep literature review likely answers all the questions but can take up a great deal of time in turn for student researchers. However, if students do not have the time to familiarize themselves with all HCI methods due to their study or semester plan, the supervisors such as doctoral students and professors have to make decisions instead about the design of the studies. This then affects the amount of time student projects must be intensively supervised and the students may not consider and learn alternative study designs when others do make their decisions. More importantly, too tight supervision situation calls into question the independence of the student's research work and the pedagogical value of conducting the students' research. Additionally, conducting user studies involving human subjects requires compliance with ethical and regulatory guidelines, which can be challenging for those who are not familiar with and new to these requirements (e.g., exchange students). All these difficulties can lead to undesired study results, negatively affect the time required for the research, and even impede the work in a research institute.

During our work with students at our university, we found that most questions were less about content and more about the organization of their studies, particularly options for their study designs, and selecting the correct statistical method for analyzing the results. For example, students who wanted to test their prototypes did not want to deal with the statistical finesse of parametric and non-parametric tests – they desired to know the statistical test for *their* research. The repetitive inquiries initially gave us the idea of making the most frequently asked questions (FAQ) available in a dedicated section on our online learning platform, but we found out that most of the questions depended on the research question

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585890

of the students. We integrated discussions into the exercises of our lectures, but here too the issue came up that the requests were repetitive, and took a lot of time to explain or justify the details (e.g., of a Fitts' law task [9, 15, 16]). We desired not to limit the creativity (and time) of our students while building new devices, user interfaces, or prototypes through additional discussions and bundled a set of tools to help them for answering their research questions.

Researchers have already provided a number of research tools, particularly for designing studies [22]. For example, Masson [17] provided a balanced Latin square generator based on the algorithm by James Bradley [5] on his website[1] to counter-balance order effects in within-subject designs. More complex designs can be developed using the ExpDes package for R [7], where researchers can identify statistical tests according to their research design. However, multiple problems with such packages have been addressed [1, 2], particularly since the researcher must have a lot of prior knowledge of the study and the statistical software (R). To overcome these issues, graphical assessments of experimental designs were developed [19]. Another popular example is the research randomizer[2] by Urbaniak and Plous [21] to randomize conditions. There are also useful help sites for planning studies such as at scribbr.com [4]. Nevertheless, students often lack the connection between theoretical content and the practical implications for their studies.

In our work with 117 students during the summer and winter semesters in our HCI lecture from 2022 to 2023, we developed a set of open-source tools to make the organization and study design planning of them easier and faster. In this paper, we present and release that toolkit to the public[3] and invite other researchers to contribute. All tools are web-based ensuring cross-platform availability and mobile usage and were developed using HTML/JavaScript/R. We explicitly emphasize that our tools do not replace the didactic measures of good support and advice of high-class researchers in HCI. Nevertheless, the tools can be an exploratory mean for the design of users studies and a basis for discussions with students, so that they learn to weigh up other alternatives. We discuss our evaluation with the students and the possibilities and limitations of our toolkit.

2 THE HCI USER STUDIES TOOLKIT

2.1 Methods and Measures

Before starting their user studies, new researchers are first supposed to explore the wide range of methods in HCI available to answer their research question(s). The review of all methods can take time despite the widely-available literature and research sources, which is why a tool can help to ask for a series of questions to quickly narrow down that search and the selection of methods. The goal of the web-tool was to limit the workload of conducting the process of operationalization from concept to measure and to find a path for necessary methods to answer individual research questions. Using the tool, the researchers are answering questions about their goals and are then able to explore and search for more detailed explanations of fitting analytical and empirical research methods. After answering the questions, the researchers get a list of all methods available (in the system) able to answer those research questions. Clicking on one of the methods or measures provides further details on how to apply and obtain data. Each accordion-based entry contains a definition and short description, the source paper (when available), variations of the method, and links to follow more detailed guidelines on the web. The tool (Figure 1) is in progress insofar as direct downloads (e.g. the corresponding questionnaire as PDF) are not yet included and use web-based sources or refer to the original articles. The students are first asked for the purpose of their study: (1) Designing or improving a system, interaction, or method, (2) Testing and understanding devices or people, or (3) Reviewing and summarizing original literature.

Based on the answers to the questions, the selection narrows down the options without displaying the answers but further questions. In the case, the students want to start a design or explore a design process (1), they are able to choose if they want to start the design from scratch (solution-oriented) or if they like to identify problems with a current design or design process (problem-oriented). When the researchers are interested in starting a new design or exploring the design space from scratch, solution-oriented options of research methods related to designing and exploring design issues or user requirements with invited experts or users are being displayed (e.g., contextual inquiry, affinity diagramming, card sorting, storyboarding, etc.). If students are interested in the final outcome of a solution-oriented method, they receive a summary of common prototyping options available in HCI (paper prototyping, wireframes, rapid prototyping, video, and image mockups, etc.). To identify problems with systems that already exist, students are able to select two approaches such as under option (2): analytical and empirical methods.

Selecting testing and understanding devices or people (2) in the tool opens the question if the research is based on logic and analytics (analytical methods) or observation (empirical methods). In the case that students want to analyze their system analytically they can choose between data-driven (task analysis, personas, public statistics, etc.) and model-driven (GOMS, KLM, etc.) analytical research methods. Selecting empirical methods continues with the sample size and if the sample is a single incident (case studies), a special subset (expert-based studies), or a random/representative population (user studies). While case studies refer to scenarios of use and other case studies, selecting a special subset of expert-based or special users inspection methods (cognitive walkthrough, heuristic evaluation, consistency inspection, etc.) and qualitative methods related to receiving feedback from participants with special knowledge (semi-structured interviews, think-aloud protocols, etc.). The respective selection in the tool textually indicates whether the method selection is suitable for users and/or experts. If the empirical method is supposed to draw data from a representative sample, the tool displays three further options: quantitative and qualitative methods, objective and subjective measures, as well as a list of concepts related to HCI.

The students can easily operationalize their concepts by selecting the interface options. Showing the list of concepts is one of the core features of the tool: Only the methods relevant and available to the research question are displayed. Besides the demographics options and guidelines (required in all studies), the tool currently

[1]https://cs.uwaterloo.ca/~dmasson/tools/latin_square/
[2]https://www.randomizer.org/
[3]https://hci-studies.org

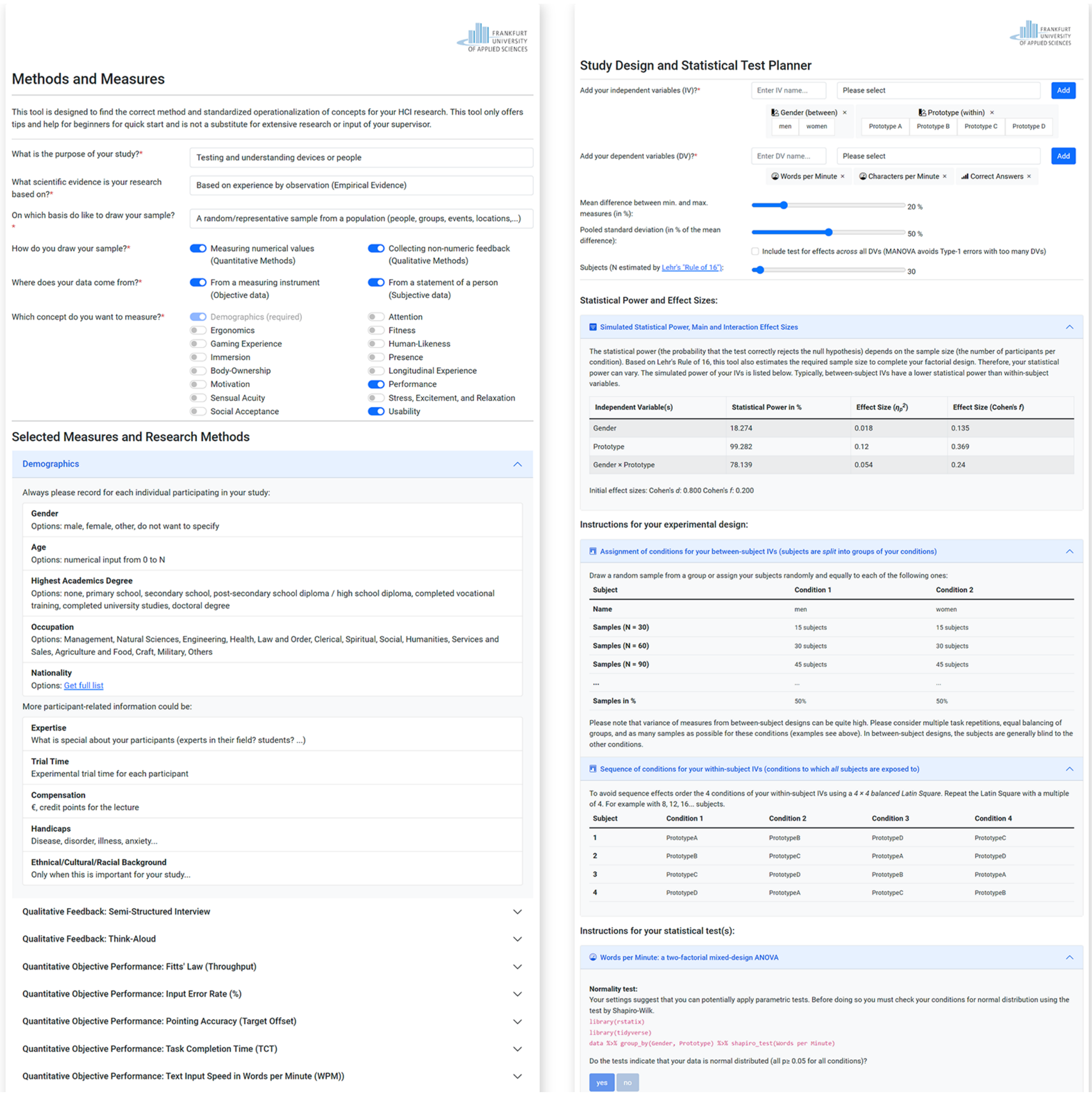

Figure 1: Screenshot of the Methods and Measures tool (left) and the Study Design and Statistical Test Planner (right). The tools are core parts of the HCI user studies toolkit and provide options and recommendations to set up a study and get a list of potential options to conduct their studies. In the Study Design and Statistical Test Planner, the researchers provide some details on the desired dependent and (between/within) independent variables to estimate statistical power, generate the conditions assignment and experimental sequence, as well as code for the statistical evaluation in R.

supports concepts such as attention, ergonomics, fitness, gaming experience, human-likeness, immersion, presence, body-ownership, longitudinal experience, motivation, performance, sensual acuity, stress/excitement/relaxation, social acceptance, and usability. If a research group e.g., only wants to measure objective performance and subjective usability, the typical metrics (Fitts' law, typing speed, system usability scale, etc.) are shown with the corresponding sources and a link to the guidelines on how to obtain the measure. New options can easily be added and edited using an XML file within the tool folder. If a method or measure is connected with a certain task, the procedure of the task is given in the description of the entry. The currently available list was added at the beginning of the summer semester '22 and based on the proposed 32 research projects and requests from doctoral students at our institute. The tool was then iteratively improved during the winter semester 22/23. To include more conceptual measures we plan to conduct a large field survey among HCI researchers and an extensive literature review to provide a holistic approach.

2.2 Study Design and Statistical Test Planner

Another core feature of the HCI Studies Toolkit is the Study Design Planner for planning empirical studies and controlled experiments without extensive knowledge about experimental designs and hypothesis testing. The tool begins with the following procedure: First, it asks the researchers to enter the name and the kind of factor/independent variables (IV) the study has (e.g., the prototype or gender). If it is a nominal variable (based on categories such as prototypes) the researchers must then enter the comma-separated names of levels and whether the IV is either within or between-subject (divided into groups or done by all participants). The tool allows mixed designs and up to three IVs. After entering at least one IV, the researchers can then enter the name of a measure/dependent variable (DV) and the kind of data (ordinal, interval, ratio) they want to obtain in their study (e.g., the items found in the measures and methods tool). Nominal and qualitative data is currently not supported. To compute effect sizes, the users can then enter the expected mean difference and pooled standard deviation (SD).

The Study Design Planner automatically sends an AJAX request to the OpenCPU platform[4] via API call to a running R instance on our server to return the corresponding study design and R code to evaluate any data based on those study designs. The tool not only suggests the recommended study design (e.g. permutations, balanced Latin square or pseudo-randomization) to avoid sequence effects but also returns the suggested number of samples (participants) to obtain at least 80% statistical power. In accordion-based entries, the framework returns statistical information (effect sizes in different units, statistical power, and other sample size estimators based on Lehr's Rule of 16 [13] or more precisely power analysis of the Superpower package [11]) and also offers a number of hints when carrying out the experiment (e.g., group sizes for between-subjects designs) and the proposed sequence of the conditions (e.g., in a balanced Latin square).

Moreover, the tool also returns the working R code to evaluate the data when a data frame with the same names of IVs and DVs is available from an experiment (e.g., as CSV) or through manual data. In the exercise, the students learned to create data tables manually so that they were able to anticipate how the data in their research project should be formatted and stored. Depending on their study design and if the data is normally distributed and parametrical or not (yes/no question in the user interface), the researchers can then see the corresponding statistical tests, post hoc analysis, and functions to compute the final effect sizes in their study. The tool currently supports parametric and nonparametric paired tests, one-way and multi-variate ANOVAs, MANOVAs, regressions, mixed designs, and implications for null-hypothesis significance tests (NHST). Bayes statistics and the support of automatic reporting descriptive and inferential statistics are planned. The current version provides three examples for student researchers to see what potential results and answers on their research questions can look like.

As students desired to extract the feature of the tool to generate balanced Latin square designs to quickly assess the length of a study, we developed a dedicated tool. We used the JavaScript implementation by Masson [17] based on the algorithm by James Bradley [5] to allow students to easily name, generate, and download their study designs as CSV file. In addition to the original tool by Masson, we implemented a copy-and-paste function to enter and save the experimental conditions more easily. Interestingly, students also desired the statistical decision tree for our interactive Study Design Planner to comprehend the choices made by the tool. The tree is based on consensus among books on statistical tests in inferential statistics and typical approaches in HCI [3, 8, 10, 12, 18]. Thus, we also provided a detailed and interactive infographic visualizing the options and relations of statistical test decisions included in the system.

2.3 Informed Consent Generator

Ethical considerations are part of every research with human subjects. One of the most important steps in taking ethics into account is to provide informed consent. The paper or digital form requires researchers to clearly and transparently inform the participants about the purpose and goal of the study, identifies the researchers, and potential risks and benefits of participation. However, the scope and workload to write an informed consent are highly underestimated. In many cases, the students can hardly overview the legal situation or principles of ethics in working with human subjects and rely on the feedback and a template of their supervisors.

To provide student researchers an easy way to write their own informed consent form we developed the Informed Consent Generator, which generates documents for participants based on the Declaration of Helsinki, the European General Data Protection Regulation (GDPR), the legal regulations in our country and internal rules of our institution (house and hygiene regulations). The JavaScript-based tool (see Figure 2) is based on a set of ready-made texts from our ethical compartment and the legal requirements as provided by our institute. The tool prompts the student researchers to answer a few questions about their user study. One key differentiator of the application is whether the user study is conducted online, field, or in a lab, as well as the implications of using the data that is collected (e.g., photos, videos, audio, physiological data, motion tracking data, etc.). In addition to formal information about taking part in a study (duration, sample size, principal investigator,

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

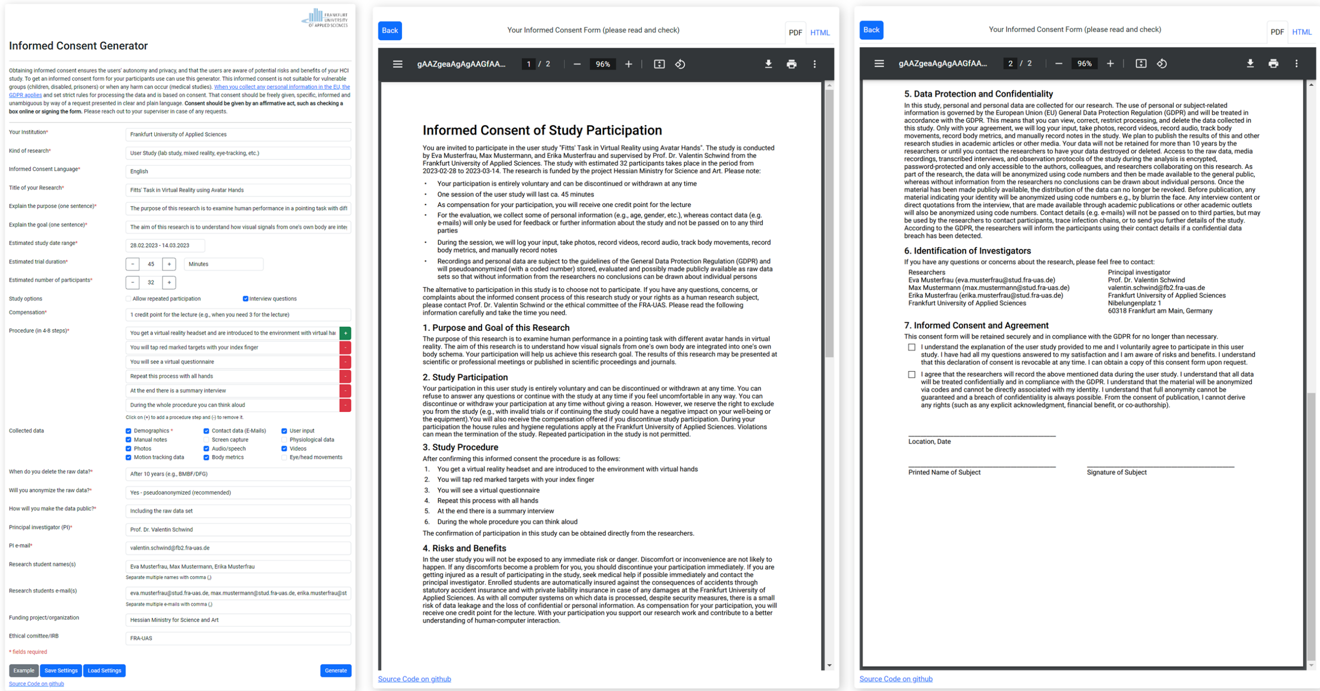

Figure 2: Informed Consent Generator (left) and screenshots from the generated PDF (right). After entering details about the study, the system offers a one-click solution to generate a full consent form for participants in online, field, or lab studies.

etc.), the researchers are also asked to explain the way of compensation (e.g., money or credit points) and details of the experimental procedure to their participants. Details on data protection and health insurance are part of legal rights and our institutions and consequently not editable. Parts of the text are modular and can be changed when aspects (e.g., the anonymization of data) are only addressed when the researchers indicate so.

Handling the generator and creating documents for their projects is the subject of the students' exercise lessons. The Informed Consent Generator was probably the most well-received tool among students and can easily be reviewed by senior researchers certified to work with human subjects. Moreover, the tool (and form) can be easily edited by replacing texts according to the regulations of another institution and country. As desired by our exchange students, more institutions can easily be added. Our tools support instant document generation, formatting, and printing as well as PDF export for signable documents.

2.4 Shared Laboratory Booking Calendar

In laboratories and co-working spaces, there are typically multiple seats and workstations, devices, and areas in which the researchers develop and prepare their user studies. Flexible room planning and booking calendars are required for the coordination of using multiple users and groups in those laboratories and if bookings were requested for user studies blocking the whole room. In addition, student groups should be able to book and delete slots via common access, but not be able to change slots made by other students or staff members. To simplify scheduling and avoid collisions while planning and booking lab times, we have developed an online tool that uses a PIN-based drag-and-drop application to reserve workstations for development time and user studies. The application is based on the FullCalendar framework[5] for JavaScript. The system has a simple front and back end for tutors and staff to enroll students and make changes. After a student or group has been enrolled students can, behind a password-protected web page, set a slot and PIN to book time for their preparation and user study at a seat or an entire room for their user study. The user study slot automatically blocks all seats in a room. Even the work on certain devices (e.g., 3D printers) can and is also being coordinated via the booking system in our laboratory. The system was particularly well received by the tutors because it dramatically reduced their email traffic. The tutors can see at one glance when which group is coming and when they have to be on-site to supervise work outside the lecture. In the next iteration of the application, students should be able to tag if they require support from the tutors or supervisors during their booking time.

2.5 Automated Confirmation of User Study Participation

To confirm their participation, attendees typically receive a short form with a certificate of participation signed by the researchers. Depending on the length of the study they receive a number of credit points. Issuing and signing certificates for participation in user studies can be a time-consuming act. Thus, we implemented a

[5]https://fullcalendar.io/

system that automatically generated those certificates for students. The benefit of automatically and digitally creating those certificates is that they can easily be reproduced, signed, and printed. The system supports regulations of more than one institution, automated signatures, and an overview record of participation for the experimenters. This is used in our lecture as a cross-check for participation to prevent students from participating in the same studies more than once (or that experimenters can fake the participation of users). In the future, we aim to submit the confirmation of user study participation to our learning platform (Moodle) in order to fully digitize the process.

3 FORMATIVE EVALUATION

The tools were formatively evaluated by our students using anonymized feedback questionnaires. We use that feedback to improve the tools and plan to integrate them permanently into the HCI courses at our institute. Preliminary results of the first qualitative evaluation round (N=6) indicate that the particularly the Methods and Measures tool was "helpful" (P1) and provides a "very good first overview" (P6) of the student's research. The Study Design Planner was perceived as "easy to use with good results" (P2), but we also identify space for improvements, e.g., to "add more options to combine variables" (P2). The participants found the Balanced Latin Square Generator was "very helpful while working with many conditions" (P5) with a "clear presentation" (P1). While the Statistical Decision Tree was considered as "not so clear and not user-friendly but sufficient" (P2), the Informed Consent and the Participation Confirmation Generator both were perceived as "actually pretty good" (P2), but obviously, the students preferred a "nicer document design" (P1). While the Booking Calendar was perceived as "easy and uncomplicated" (P5), some participants complained about the "overbooking by some groups" (P1), and that "booked appointments have been deleted" (P3) (accidentally by the tutors), which must be prevented until the next evaluation.

4 DISCUSSION AND CONTRIBUTION

With our tools, we address the organizational issues that initially arise in shared research and lab environments in HCI research as well as with novice and student researchers. We contribute with all the tools via GitHub[6] under an open-source MIT license and warmly invite other researchers to participate and improve the tools and content. In addition to the herein presented toolkit, the tools are part of a number of other materials that are used in our lectures and exercises on our open-access HCI web page[7] (lecture and exercise slides, JupyterLab, Wiki, slides/paper/thesis template for Latex). However, while the toolkit supports novice HCI researchers and may be applicable in various educational systems, the current tool is mainly designed to address the authors' institution's needs. Thus, we will release the tool on our website for public use and observe its capabilities with additional survey mechanisms, and extend the formative evaluation using external feedback to assess how the tool affects students' learning outcomes in other institutions and their skill development over time.

We see the potential of the toolkit to support undergraduates in HCI lectures, research projects, as well as theses. We highlight that further explorations are required to determine how the toolkit can be improved to aid students in making informed decisions and comprehending the methods they have selected. It is also important to acknowledge the potential risks of using such tools as they may also lead to overlooking innovative and investigative research designs. Moreover, the use of such tools may hinder critical reflection on the proposed methods or study designs as well as in-depth ethical reflections. In our institute, we address these issues through milestone presentations by the students and regular feedback from senior researchers. However, it is essential to thoroughly examine the long-term consequences of using such tools and continue conducting research to understand their impact on the learning situation in other research institutions with different (cultural) backgrounds. It is also important to note that these tools should not replace the role of good support and supervision from senior researchers. For future examinations, we release the toolkit as a resource to improve the learning experience and situation with novice researchers in HCI.

REFERENCES

[1] E Arnhold. 2014. Easyanova: Analysis of variance and other important complementary analyzes. *R package version* 4 (2014).

[2] Emmanuel Arnhold et al. 2013. Package in the R environment for analysis of variance and complementary analyses. *Brazilian Journal of Veterinary Research and Animal Science* 50, 6 (2013), 488–492.

[3] Arthur Aron and Elaine N Aron. 1999. *Statistics for psychology.* Prentice-Hall, Inc.

[4] Rebecca Bevans. 2022. Guide to Experimental Design | Overview, Steps, & Examples. https://www.scribbr.com/methodology/experimental-design/

[5] James V. Bradley. 1958. Complete Counterbalancing of Immediate Sequential Effects in a Latin Square Design. *J. Amer. Statist. Assoc.* 53, 282 (1958), 525–528. http://www.jstor.org/stable/2281872

[6] Jeffrey C Carver, Letizia Jaccheri, Sandro Morasca, and Forrest Shull. 2010. A checklist for integrating student empirical studies with research and teaching goals. *Empirical Software Engineering* 15, 1 (2010), 35–59.

[7] Eric B Ferreira, Pórtya P Cavalcanti, Denismar A Nogueira, et al. 2014. ExpDes: an R package for ANOVA and experimental designs. *Applied Mathematics* 5, 19 (2014), 2952.

[8] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering statistics using R (2012).* Vol. 958.

[9] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[10] Dennis Howitt and Duncan Cramer. 2007. *Introduction to statistics in psychology.* Pearson education.

[11] Daniel Lakens and Aaron Caldwell. 2021. Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science* 4, 1 (2021), 251524592095150. https://doi.org/10.1177/2515245920951503

[12] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction.* Morgan Kaufmann.

[13] Robert Lehr. 1992. Sixteen S-squared over D-squared: A relation for crude sample size estimates. *Statistics in Medicine* 11, 8 (1992), 1099–1102. https://doi.org/10.1002/sim.4780110811 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.4780110811

[14] Gondy Leroy. 2011. *Designing user studies in informatics.* Springer Science & Business Media.

[15] I. Scott MacKenzie. 2018. Fitts' law. *Handbook of human-computer interaction* 1 (2018), 349–370.

[16] I. Scott MacKenzie and William Buxton. 1992. Extending Fitts' Law to Two-Dimensional Tasks *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 219–226. https://doi.org/10.1145/142750.142794

[17] Damien Masson. 2023. Balanced Latin Square Generator. https://cs.uwaterloo.ca/~dmasson/tools/latin_square/

[18] Judy Robertson and Maurits Kaptein. 2016. *Modern statistical methods for HCI.* Springer.

[6]https://github.com/stars/valentin-schwind/lists/hci-user-studies-toolkit
[7]https://hci-studies.org/

[19] Pieter Schoonees, Niel Le Roux, and Roelof Coetzer. 2016. Flexible graphical assessment of experimental designs in R: The vdg package. (2016).

[20] Anam Sohail. 2020. *Challenges in Recruiting Participants for Studies in HCI.* Ph. D. Dissertation. RWTH Aachen University.

[21] Geoffrey C. Urbaniak and Scott Plous. 2013. Research Randomizer (Version 4.0). https://www.randomizer.org/

[22] Arnold P. O. S. Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User Experience Evaluation Methods: Current State and Development Needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (Reykjavik, Iceland) *(NordiCHI '10)*. Association for Computing Machinery, New York, NY, USA, 521–530. https://doi.org/10.1145/1868914.1868973