

Conducting User Experiments in Recommender Systems

Bart P. Knijnenburg

bartk@clemson.edu

School of Computing, Clemson University

Clemson, South Carolina, USA

Edward C. Malthouse

ecm@northwestern.edu

Department of Integrated Marketing Communication and

Department of Industrial Engineering and Management

Science, Northwestern University

Evanston, Illinois, USA

ABSTRACT

This tutorial provides practical training in designing and conducting online user experiments with recommender systems, and in statistically analyzing the results of such experiments. It covers the development of a research question and hypotheses, the selection of study participants, the manipulation of system aspects and measurement of behaviors, perceptions and user experiences, and the evaluation of subjective measurement scales and study hypotheses. Interested parties can find the slides, example dataset, and other resources at <https://www.usabart.nl/QRMS/>.

CCS CONCEPTS

• **Human-centered computing** → *User studies*; • **Information systems** → **Recommender systems**.

KEYWORDS

recommender systems, user studies

ACM Reference Format:

Bart P. Knijnenburg and Edward C. Malthouse. 2024. Conducting User Experiments in Recommender Systems. In *18th ACM Conference on Recommender Systems (RecSys '24)*, October 14–18, 2024, Bari, Italy. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3640457.3687090>

1 AUDIENCE

The tutorial is targeted at researchers in the field of Recommender Systems (and related areas). The relative lack of knowledge of user-centric evaluation in this field makes the course suitable for students and early-career researchers as well as more experienced researchers. As we plan to present state-of-the-art statistical evaluation methods (i.e. CFA and SEM), this course may even be of interest to researchers with an established track record in user-centric recommender systems research.

No prior knowledge of user experience research is required. General knowledge of fundamental concepts of recommender systems is desirable, as is an introduction-level background in statistical evaluation.

2 MOTIVATION

Traditionally, the field of recommender systems has evaluated the fruits of its labor using metrics of algorithmic accuracy and precision. In recent years, however, researchers have come to realize that the goal of a recommender system extends well beyond accurate predictions; its primary real-world purpose is to provide personalized help in discovering relevant content or items. This realization has caused prominent recommender systems researchers to call for a broadening of the scope of research beyond algorithms and beyond accuracy- or precision-based evaluation [1, 3, 5, 6].

Despite these calls, surprisingly little recommender systems research focuses on preference elicitation, the presentation of recommendations, and/or other aspects of the “Human-Recommender Interaction”. Similarly, very few researchers evaluate their recommenders in online user experiments with subjective and experience-based metrics.

While several papers [3, 5], book chapters [4], and past tutorials [2] on the user experience of recommender systems have been instrumental in raising awareness regarding these topics, we believe that a lack of more in-depth training in user-centric design and evaluation methods remains an important reason for the relative lack of user-centric recommender systems research, making a tutorial on user experiments in recommender systems both timely and important. This tutorial provides practical training in conducting user experiments and statistical analysis of the results of such experiments, thereby helping researchers and practitioners improve the user experience of the recommender systems they develop. In the long term, this will trigger more scientific user-centric work that can grow our knowledge on how certain recommender system aspects influence the user experience.

3 OUTLINE

The tutorial consists of the following sections/activities:

Theoretical foundations: The tutorial will walk through the typical stages of designing, conducting and evaluating a user experiment in the context of recommender systems. The tutorial will begin with a discussion of existing theories of technology adoption and user experience and discusses how these theories can be used as an experimental lens for user-centric recommender systems research. It teaches how to pose theoretically motivated research questions. As an exercise, participants formulate their own research question.

Developing a research model: We will then explain how to construct a research model, i.e., a body of interrelated hypotheses, based on the posed research questions. This part of the tutorial teaches how to select system aspects to manipulate and subjective

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

RecSys '24, October 14–18, 2024, Bari, Italy

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0505-2/24/10

<https://doi.org/10.1145/3640457.3687090>

and behavioral variables to measure in order to test these hypotheses. As an exercise, participants develop a research model based on their research question

Setting up an experiment: This part of the tutorial will discuss how to select an unbiased sample of participants from a target population, how to select experimental conditions to test, and how to assign participants to conditions. As an exercise, participants develop their own experimental setup.

Determining the sample size: We will teach how to determine the optimal sample size for your study using *power analysis*, and cover sequential analyses as an economical way to run large studies. As an exercise, participants will conduct a power analysis to determine the optimal sample size for their own experiment.

Measuring demographics, perceptions and experiences: Next, we will discuss how to measure users' demographics in a way that balances data utility and privacy considerations. We will also teach how to measure perceptions and experiences using multi-item questionnaires. As an exercise, participants develop their own questionnaire(s).

Evaluating measurement scales: This part of the tutorial teaches how to evaluate the quality of the administered questionnaires using *confirmatory factor analysis* (CFA) using R package *lavaan*¹. Participants who have R installed on their laptops can follow along with a dataset that we will make available at <https://www.usabart.nl/QRMS/>.

Evaluating entire research models: The tutorial concludes by covering *structural equation modeling* (SEM), a state-of-the-art method for evaluating the results of an experiment by testing entire research models (rather than testing hypotheses one by one). We will demonstrate how to run SEMs with R package *lavaan*. Participants who have R installed on their laptops can follow along with our provided dataset.

4 MATERIALS

The tutorial slides, example dataset, and more in-depth resources are available at <https://www.usabart.nl/QRMS/>.

ACKNOWLEDGMENTS

This tutorial is supported by the U.S. National Science Foundation under Grant No. 22-32551.

REFERENCES

- [1] Michael D. Ekstrand and Martijn C. Willemsen. 2016. Behaviorism is Not Enough: Better Recommendations Through Listening to Users. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, Boston, MA, USA, 221–224. <https://doi.org/10.1145/2959100.2959179> event-place: Boston, Massachusetts, USA.
- [2] Bart P. Knijnenburg. 2012. Conducting User Experiments in Recommender Systems. In *Proceedings of the Sixth ACM Conference on Recommender Systems (RecSys '12)*. ACM, Dublin, Ireland, 3–4. <https://doi.org/10.1145/2365952.2365956>
- [3] Bart P. Knijnenburg, Saadhika Sivakumar, and Darcia Wilkinson. 2016. Recommender Systems for Self-Actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, Boston, MA, USA, 11–14. <https://doi.org/10.1145/2959100.2959189>
- [4] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. Evaluating Recommender Systems with User Experiments. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, USA, 309–352. https://doi.org/10.1007/978-1-4899-7637-6_9
- [5] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- [6] Joseph A. Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction* 22, 1-2 (April 2012), 101–123. <https://doi.org/10.1007/s11257-011-9112-x>

¹<https://cran.r-project.org/web/packages/lavaan/index.html>