EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2015)
W. Aigner, P. Rosenthal, and C. Scheidegger (Editors)

Choosing the Right Sample? Experiences of Selecting Participants for Visualization Evaluation

Simone Kriglstein[1] and Margit Pohl[1]

[1]Institute for Design & Assessment of Technology (IGW), Vienna University of Technology, Austria

Abstract

Conducting and reporting evaluation studies has become more and more popular over the last few years in the information visualization community. A big challenge is to describe such studies in a way such that the investigations are repeatable and comparable with other studies. This not only includes the description of methodology, tasks, and procedure of the study but also information about the participants – including the reasons for their selection – to make the work reproducible and to assess its validity. In this paper we give a short overview about our research that we conducted in the past to show in which context and situations which types of test persons (e.g., students or experts) were considered.

Categories and Subject Descriptors (according to ACM CCS): H 5.2. [Information Interfaces and Presentation(e.g., HCI)]: User Interfaces—Evaluation/Methodology

1. Introduction

Reproducibility of research results is an important aspect of scientific research in many different areas [IIC*13]. To attain reproducibility, different procedures have to be followed in different disciplines. In social sciences and the experimental areas of Human-Computer Interaction (HCI), reproducibility implies that various aspects of the experimentation process are reported in great detail so that investigations can be repeated easily and readers are able to assess their validity without any problems. Such aspects might be, for example: hypotheses investigated, methods of investigation, description of the sample, details of the procedure of the study (e.g. where was an investigation conducted, how much time did it take, etc.), tasks used in the investigation, methods of analysis, and results. It is well known that such details are often lacking in evaluation studies in information visualization [IIC*13].

One important issue in such a reporting procedure is the sample used for the investigation. Investigations in HCI and information visualization often use students as participants. There is an ongoing controversial discussion about this issue (cf. [BD06]). In general, it is seen as problematic to use students as extensively as is currently the case. On the other hand, it is well accepted that it is difficult to convince experts to take part in comprehensive evaluation studies. In addition, it is not advisable to use some kinds of study methods with experts. Experts might, for example, be offended if they have to solve long series of challenging cognitive tasks. On the other hand, such tasks might provide researchers with important insights into cognitive processes. Therefore, it is an open question whether to use experts or students for evaluations in information visualization. This question is also related to other important aspects of experimental processes. As mentioned above, some kinds of tasks and methods are not appropriate for testing experts.

In this paper, we discuss research we conducted in the past to clarify the conditions where it is more appropriate to use experts or students for the evaluation of information visualizations.

2. Analysis of Previous Research Work

We conducted several studies concerning the application of information visualizations in the medical domain, partly with physicians as participants and partly with students. We also conducted a study to compare the results of physicians with those of students [RAM*11b]. This study indicated that the differences between the results of experts and of students were not highly significant. The most obvious difference was that students took more time to interact with the system and therefore identified more usability problems than the physicians. Students were slightly more favorable con-

DOI: 10.2312/eurorv3.20151146

cerning novel visualizations and interaction techniques, but this result was not very pronounced. All other results indicated that there were no differences between students and experts. We did not analyze the insights gained with the tools. We expect that in this area the differences would have been really significant, but as far as usability problems were concerned the differences were negligible.

We also conducted a larger study of another system in the medical domain [RPW*07] in which the sample only consisted of students. Therefore, we were able to investigate a larger number of participants (32 students). The students got an introduction of about one hour into the system and the goal of the investigation and then, a few days later, worked for approximately 3 hours to solve the given tasks. The students had to solve a considerable number of tasks and then had to provide meta-information about the insights they got (screenshot, confidence rating, etc.). Such a complex study design would not be possible with physicians. In another study with 10 physicians we also used predefined tasks, but the main part of the investigation was the observation of the interaction of the physicians with the tool to detect usability errors and an interview conducted after the test. Such a study design is accepted by physicians [RAM*11a].

For the development of an ontology visualization tool [KW13] we conducted evaluation studies with domain and ontology experts. In total, 44 domain and 3 ontology experts participated. The evaluation studies with domain experts showed that they critically analyzed if the visualization was helpful for their purpose. In addition, they allowed us to verify if the ontology was correctly visualized. Furthermore, they carefully investigated the tool in regard to usability problems and design. They were very motivated to improve the tool since the visualization will support them in their tasks in the future. For the ontology experts information and features were of interest which are relevant to support them in the development tasks of ontologies (e.g., they missed the internal types of the datatype properties). Although the number of asked ontology experts who participated in the study was rather small, the evaluation sessions were very valuable since in-depth questions could be answered that had also a positive impact on the design of the ontology which was also of interest for the domain experts.

2.1. Literature Analysis

In addition to the experiences gained from our evaluations studies in our projects, we also gained interesting insights in our previous literature analysis [KPS14, PKSW11] about the selection of participants and their description. For example, we observed that in many publications relevant information about the user studies (e.g., which methodology was used, sample size, details about the participants, and duration of the study) was not sufficiently provided which makes reproducibility difficult. For example, 15% from 68 publications did not report the sample size (cf. [PKSW11]). This lack of information makes it difficult to generalize the results, to compare them with other results, and to derive tentative recommendations or guidelines for the design of visualizations. Another interesting observation was that the recruitment of experts was often a big challenge and therefore the sample size did not often exceed 10 experts.

3. Discussion and Conclusions

Our literature analysis as well as research works like [IIC*13] and [LBI*12] showed that user studies have become more and more prevalent in the information visualization and visual analytics community. However, as also observed by Isenberg et al. [IIC*13] many publications lacked in their description of their studies and hence make reproducibility difficult. Especially the reporting about the participants is still neglected in publications. However, already the guidelines developed by Forsell and Cooper [FC12] and also the recommendations by Isenberg et al. [IIC*13] pointed out that the information about the participants and their characteristics depending on the purpose of the study are essential for reporting of evaluation studies. This can, perhaps, partly be attributed to page limitations of publications which make trade-offs necessary. However, also the selection and recruitment of participants, especially of experts, is often a big challenge. Therefore, visualization systems are often evaluated with students which is not clearly reported in the publications. Our experiences with our previous evaluation studies showed us that the purpose and goal of the evaluation studies are very important for selecting participants. It makes a differences which types of test persons, for example students or experts, take part in evaluation. Although experts are important to verify if the visualization is useful for the domain, presents all relevant content, and fits their workflow, our experiences in the evaluation studies showed us that students are very good candidates to identify usability issues. Since the mechanisms of human cognition (e.g., visual search, color vision, process of seeing, perception) are similar for all humans, students are also suitable, for example, for evaluating general cognitive processes. Especially for such evaluations it is often necessary to conduct controlled experiments which require a large number of participants and students are easier to recruit than experts.

Based on our observations and experiences in evaluation studies and literature analysis, we recommend to precisely express the goal, purpose, and output of evaluation studies. This includes being clear about the selection process of the experimental group.

4. Acknowledgments

This work is conducted in the context of the CVAST (Centre of Visual Analytics Science and Technology) project. It is funded by the Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative (project number: 822746).

References

[BD06] BORTZ J., DÖRING N.: *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4th ed. Springer, 2006. 1

[FC12] FORSELL C., COOPER M.: A guide to reporting scientific evaluation in visualization. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2012), AVI '12, ACM, pp. 608–611. 2

[IIC*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMAIR M., MOLLER T.: A systematic review on the practice of evaluating visualization. *Visualization and Computer Graphics, IEEE Transactions on 19*, 12 (Dec 2013), 2818–2827. 1, 2

[KPS14] KRIGLSTEIN S., POHL M., SMUC M.: Pep up your time machine: Recommendations for the design of information visualizations of time-dependent data. In *Handbook of Human Centric Visualization*, Huang W., (Ed.). Springer, 2014, pp. 203–225. 2

[KW13] KRIGLSTEIN S., WALLNER G.: Human centered design in practice: A case study with the ontology visualization tool knoocks. In *Computer Vision, Imaging and Computer Graphics. Theory and Applications*, Csurka G., Kraus M., Mestetskiy L., Richard P., Braz J., (Eds.), vol. 274 of *Communications in Computer and Information Science*. Springer, 2013, pp. 123–141. 2

[LBI*12] LAM H., BERTINI E., ISENBERG P., PLAISANT C., CARPENDALE S.: Empirical studies in information visualization: Seven scenarios. *Visualization and Computer Graphics, IEEE Transactions on 18*, 9 (Sept 2012), 1520–1536. 2

[PKSW11] POHL M., KRIGLSTEIN S., STACHL C., WIENER S.: *Cognitive Aspects of the Design and Evaluation of the Information Discovery Process using well-known Techniques from the Literature*. Report, Vienna University of Technology, 2011. 2

[RAM*11a] RIND A., AIGNER W., MIKSCH S., WILTNER S., POHL M., DREXLER F., NEUBAUER B., SUCHY N.: Visually exploring multivariate trends in patient cohorts using animated scatter plots. In *Proceedings of the 2011th International Conference on Ergonomics and Health Aspects of Work with Computers* (2011), EHAWC'11, Springer, pp. 139–148. 2

[RAM*11b] RIND A., AIGNER W., MIKSCH S., WILTNER S., POHL M., TURIC T., DREXLER F.: Visual exploration of time-oriented patient data for chronic diseases: Design study and evaluation. In *Information Quality in e-Health*, Holzinger A., Simonic K.-M., (Eds.), vol. 7058 of *LNCS*. Springer, 2011, pp. 301–320. 1

[RPW*07] RESTER M., POHL M., WILTNER S., HINUM K., MIKSCH S., POPOW C., OHMANN S.: Evaluating an infovis technique using insight reports. In *Proceedings of 11th International Conference on Information Visualization* (2007), IV '07, IEEE Computer Society Press, pp. 693–700. 2