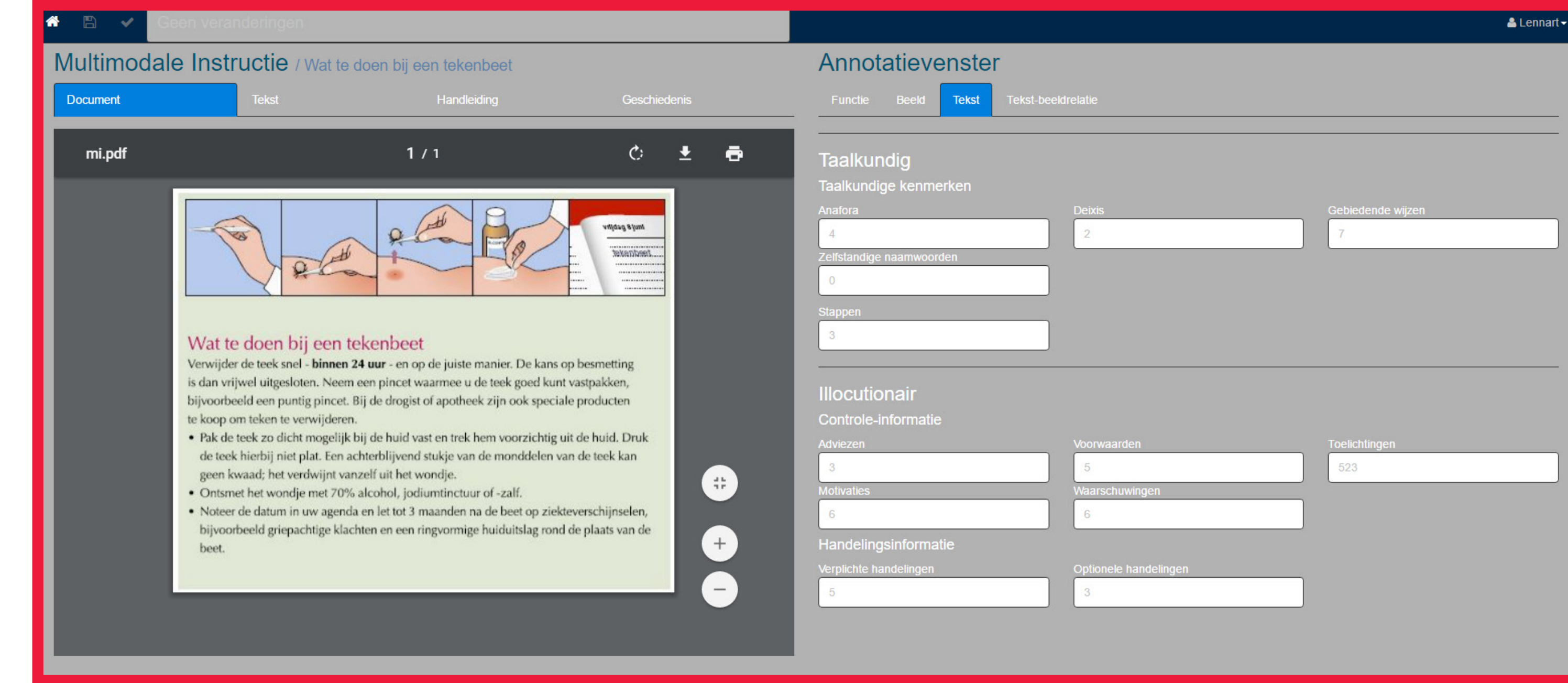


Does it work?

Applying formative evaluation methods to test the usability of the PAT Workbench



Introduction

Research question
 How do different usability evaluation methods, focused on experts and users, contribute to the evaluation of a system during an iterative design process?

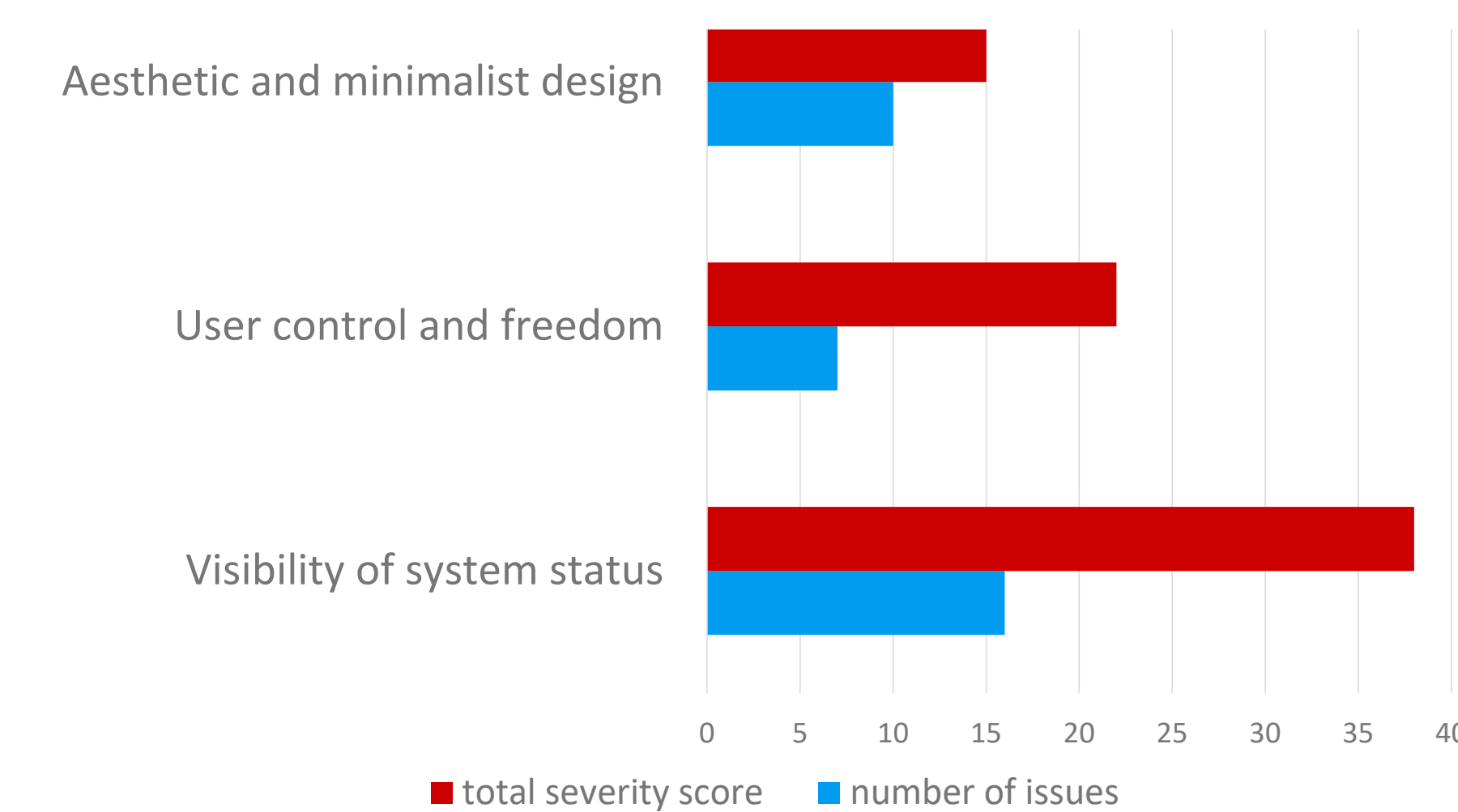
The PAT Workbench was used as a case to test usability methods. This is a tool that was developed at the department of Communication and Information Sciences of the University of Groningen, to store, annotate, retrieve and view multimodal instructions (MIs). MIs are instructions that consists of text and pictures. The goal of the PAT Workbench is to create a corpus of annotated MIs for further research. The framework used for this process was the design research cycle (Hevner, 2007).

Results

Expert review: Heuristic inspection

This inspection was based on the 10 heuristic principles by Nielsen (1995). Results showed that, given the three top tasks, the two most severe usability problems were 'visibility of system status' (e.g. lack of feedback to the user, available information and functions were not clear) and 'consistency and web standards' (e.g. clickable links styled as text, site behaved not as expected).

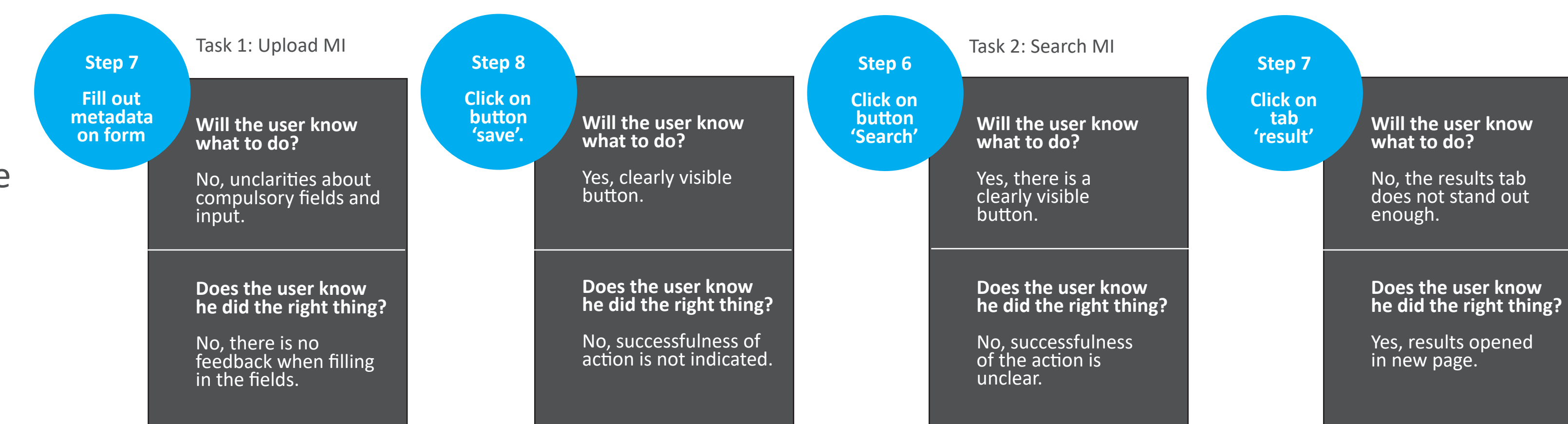
Figure 1: Key usability problems from heuristic inspection.



Expert review: Cognitive walkthrough

The cognitive walkthrough did not pose as many problems as the heuristic inspection. The top tasks were easy to accomplish by the user. The main problem was a lack of feedback or visibility of feedback and guidance for the user.

Figure 2: Problems during tasks from cognitive walkthrough.



User evaluation

Results of the first test show that participants were able to upload and search MIs. However, participants experienced problems during Task 3 - annotate an MI - with finding the annotation page (3a), as well as with saving and viewing their annotations (3b).

Figure 3: Amount of errors Test 1.

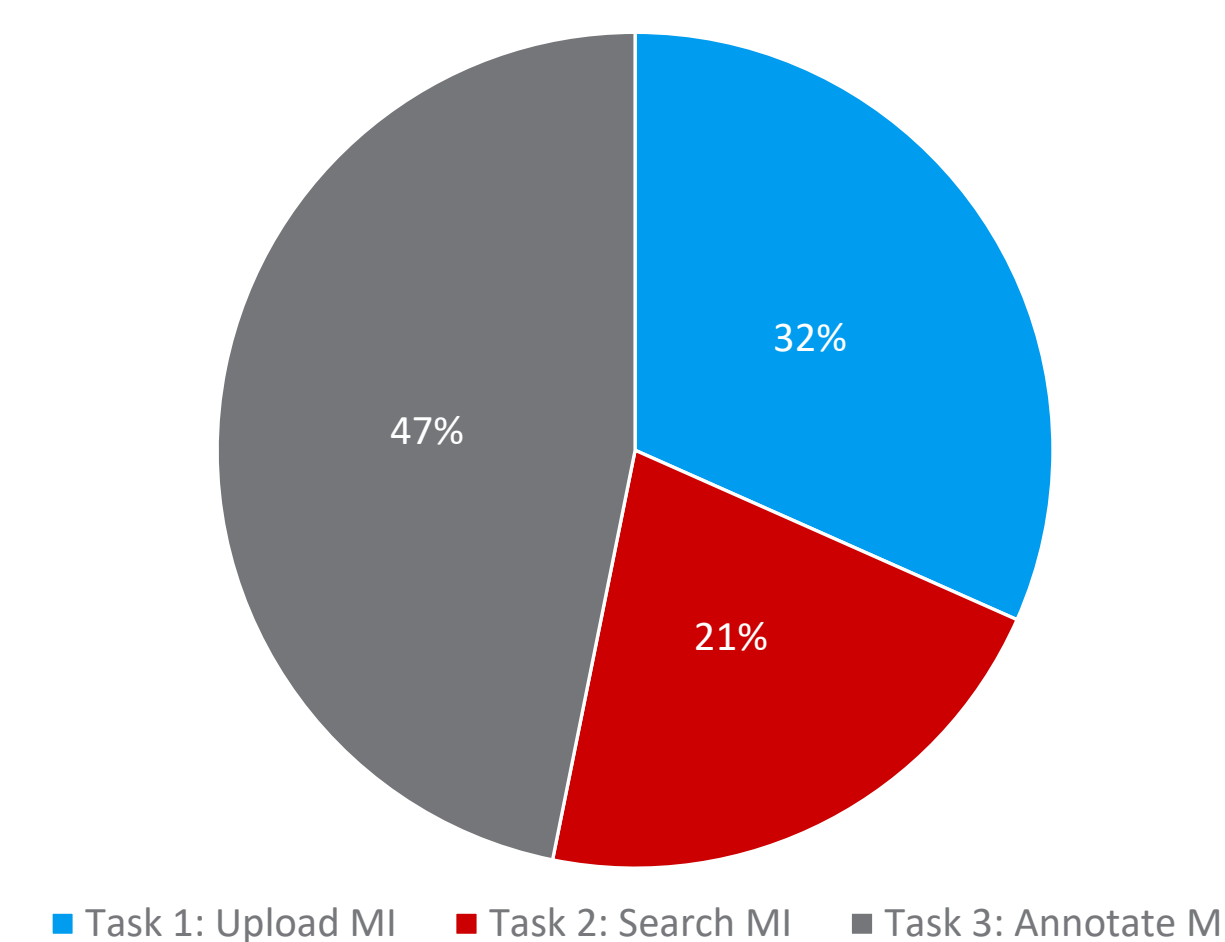
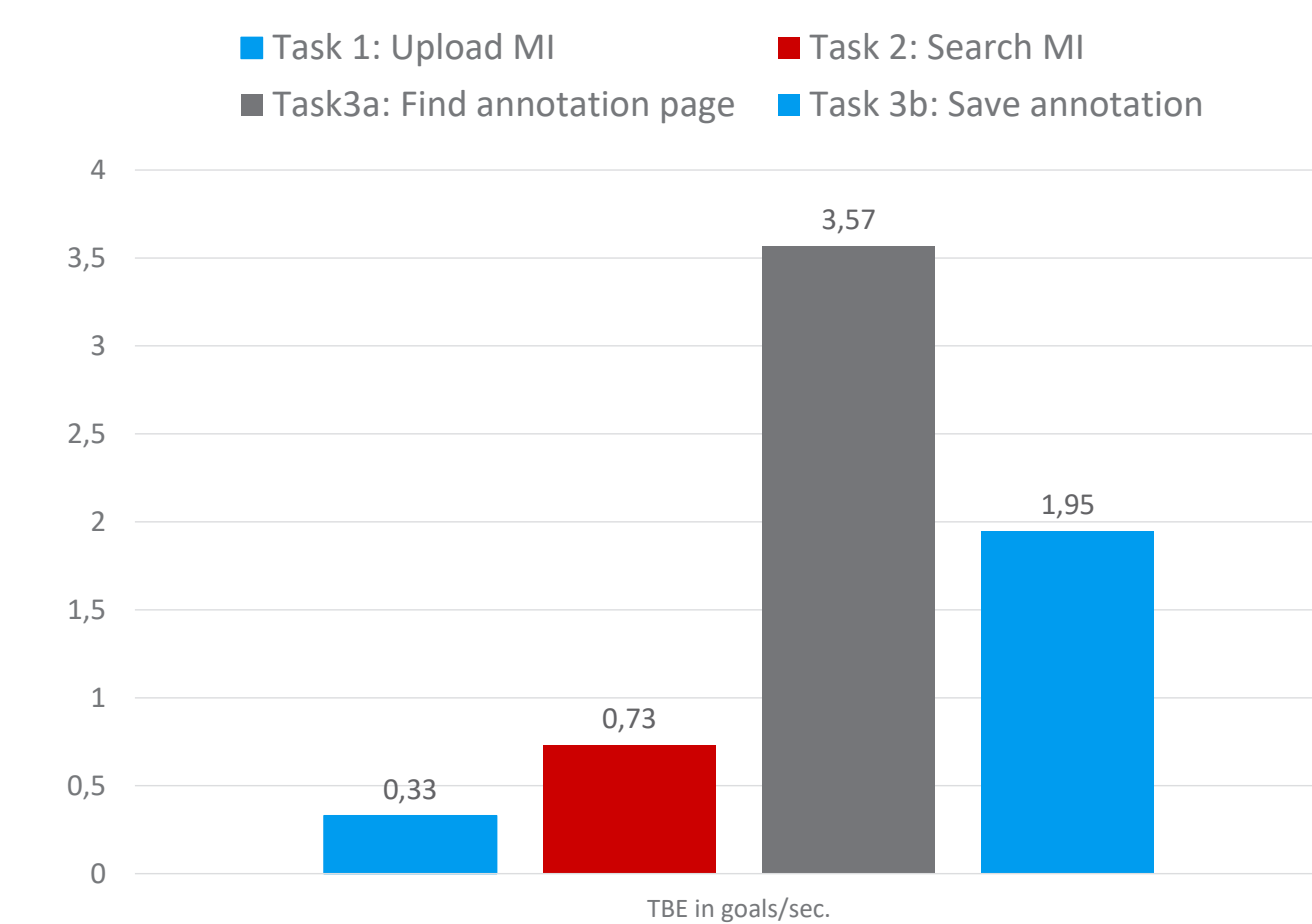


Figure 4: Time based efficiency Test 1.



During Test 2 participants had encountered only a few problems while executing the tasks. However, overall satisfaction had decreased. From the interviews it became apparent that working with a system while it was being developed influenced the participants workload and quality of work.

Figure 5: Satisfaction from Test 1 to Test 2.

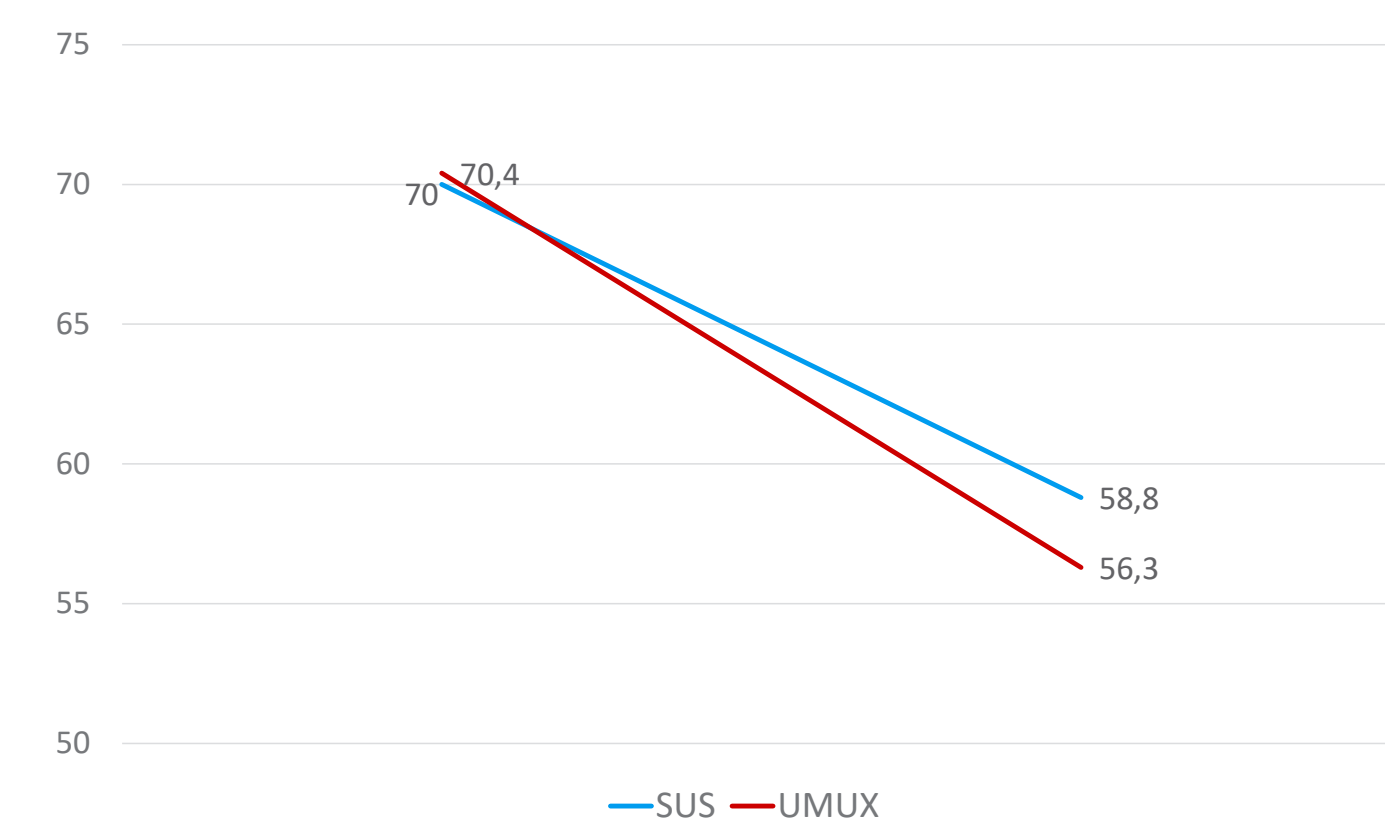
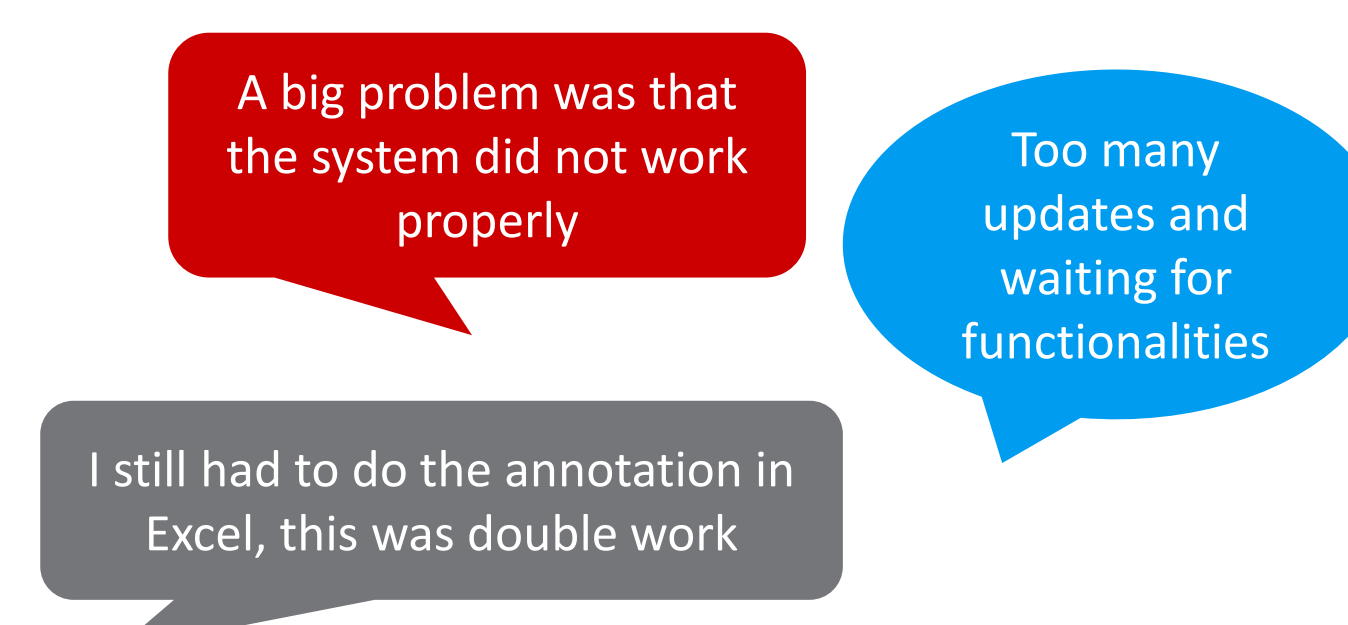


Figure 6: Reactions from participants during interview Test 2.



Conclusion

An iterative design process allows for the use of various evaluation methods, which contribute to the evaluation of a system in different ways:

Expert evaluation

- is cheap,
- offers a detailed system description,
- helps to overcome obvious issues in a more expensive user evaluation.

User evaluation

- displays multiple aspects of the system, which an expert may overlook,
- provides useful insights in time based efficiency, error count and task completion,
- benefits from Think Aloud Protocols (although with concurrent TAP participants continuously need reminders to verbalise their thoughts).

In longitudinal studies, added value in iterative tests may be gained from fresh participants in addition to the original ones.

In between testing, it is advised to not let participants use a beta version of the product for their own work.

In both types of evaluation good communication and collaboration between developer and tester are crucial.

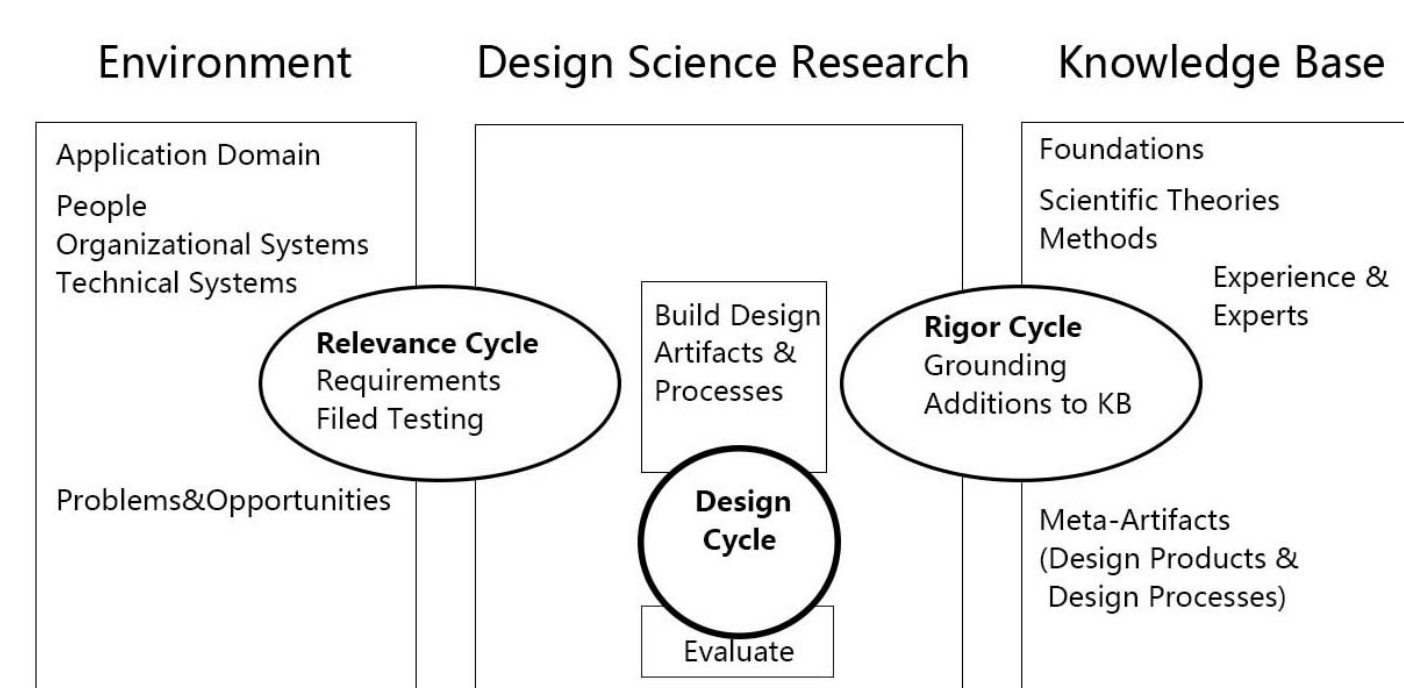
References

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, 19(2), 4.

Nielsen, J. (1995, January 1st). *10 Usability Heuristics for User Interface Design*. Retrieved from <https://www.nngroup.com/articles/ten-usability-heuristics/>

Methods

Design research cycle



Expert review

- Functional analysis
- Heuristic inspection
- Cognitive walkthrough w.r.t. PAT's top tasks (upload MI, search MI and annotate MI)

User evaluation

Two tests with real users, students enrolled in the master course on multimodal instructions in Communication and Information Sciences.

- Test 1 at beginning of course, with 9 users
- Test 2 after seven weeks, with 4 users
- Three top tasks while thinking aloud
- A questionnaire after each task
- Two general usability questionnaires
- Interview with the participants.