

## Scientific abstract dissertation Sanne Elling

The benefits of evaluating websites among potential users are widely acknowledged. There are several methods that can be used to evaluate the websites' quality from a users' perspective. In current practice, many evaluations are executed with inadequate methods that lack research-based validation. This thesis aims to gain more insight into evaluation methodology and to contribute to a higher standard of website evaluation in practice.

A first way to evaluate website quality is measuring the users' opinions. This is often done with questionnaires, which gather opinions in a cheap, fast, and easy way. However, many questionnaires seem to miss a solid statistical basis and a justification of the choice of quality dimensions and questions. We therefore developed the 'Website Evaluation Questionnaire' (WEQ), which was specifically designed for the evaluation of governmental websites. In a study in online and laboratory settings the WEQ has proved to be a valid and reliable instrument.

A way to gather more specific user opinions, is inviting participants to review website pages. Participants provide their comments by clicking on a feedback button, marking a problematic segment, choosing a category, and formulating their feedback. There has been debate about the extent to which users are able to provide relevant feedback. The results of our study on users' abilities to review web pages showed that participants were able to provide useful feedback. The reviewing participants signalled many relevant problems that indeed were experienced by users who needed to find information on the website.

Website quality can also be measured during participants' task performance. A frequently used method is the concurrent think-aloud method (CTA), which involves participants who verbalize their thoughts while performing tasks. There have been doubts on the usefulness and exhaustiveness of participants' verbalizations. Therefore, we have combined CTA and eye tracking in order to examine the cognitive processes that participants do and do not verbalize. The results showed that the participants' verbalizations provided substantial information in addition to the directly observable user problems. There was also a rather high percentage of silences (27%) during which interesting observations could be made about the users' processes and obstacles. A thorough evaluation should therefore combine verbalizations and (eye tracking) observations.

In a retrospective think-aloud (RTA) evaluation participants verbalize their thoughts afterwards while watching a recording of their performance. A problem with RTA is that participants not always remember the thoughts they had during their task performance. We therefore complemented the dynamic screen replay of their actions (pages visited and mouse movements) with a dynamic gaze replay of the participants' eye movements. Contrary to our expectations, no differences were found between the two conditions.

It is not possible to draw conclusions on the single best method. The value of a specific method is strongly influenced by the goals and context of an evaluation. Also, the outcomes of the evaluation not only depend on the method, but also on other choices during the evaluation, such as participant selection, tasks, and the subsequent analysis.

The benefits of evaluating websites among potential users are widely acknowledged. There are several methods that can be used to evaluate the websites' quality from a users' perspective. In current practice, many evaluations are executed with inadequate methods that lack research-based validation. This thesis aims to gain more insight into evaluation methodology and to contribute to a higher standard of website evaluation in practice.

A first way to evaluate website quality is measuring the users' opinions. This is often done with questionnaires, which gather opinions in a cheap, fast, and easy way. However,

many questionnaires seem to miss a solid statistical basis and a justification of the choice of quality dimensions and questions. We therefore developed the 'Website Evaluation Questionnaire' (WEQ), which was specifically designed for the evaluation of governmental websites. In a study in online and laboratory settings the WEQ has proved to be a valid and reliable instrument.

A way to gather more specific user opinions, is inviting participants to review website pages. Participants provide their comments by clicking on a feedback button, marking a problematic segment, and formulating their feedback. There has been debate about the extent to which users are able to provide relevant feedback. The results of our studies showed that participants were able to provide useful feedback. They signalled many relevant problems that indeed were experienced by users who needed to find information on the website.

Website quality can also be measured during participants' task performance. A frequently used method is the concurrent think-aloud method (CTA), which involves participants who verbalize their thoughts while performing tasks. There have been doubts on the usefulness and exhaustiveness of participants' verbalizations. Therefore, we have combined CTA and eye tracking in order to examine the cognitive processes that participants do and do not verbalize. The results showed that the participants' verbalizations provided substantial information in addition to the directly observable user problems. There was also a rather high percentage of silences (27%) during which interesting observations could be made about the users' processes and obstacles. A thorough evaluation should therefore combine verbalizations and (eye tracking) observations.

In a retrospective think-aloud (RTA) evaluation participants verbalize their thoughts afterwards while watching a recording of their performance. A problem with RTA is that participants not always remember the thoughts they had during their task performance. We therefore complemented the dynamic screen replay of their actions (pages visited and mouse movements) with a dynamic gaze replay of the participants' eye movements. Contrary to our expectations, no differences were found between the two conditions.

It is not possible to draw conclusions on the single best method. The value of a specific method is strongly influenced by the goals and context of an evaluation. Also, the outcomes of the evaluation not only depend on the method, but also on other choices during the evaluation, such as participant selection, tasks, and the subsequent analysis.

### **De Engelstalige samenvatting op de achterzijde van het proefschrift**

The benefits of evaluating websites among potential users are widely acknowledged. There are several methods that can be used to evaluate the websites' quality from a users' perspective. In current practice, many evaluations are executed with inadequate methods that lack research-based validation. This thesis aims to gain more insight into evaluation methodology and to contribute to a higher standard of website evaluation in practice.

The studies in this thesis consist of research on three user-focused evaluation methods, which were studied within the context of municipal websites. Two methods evaluate the websites' quality by measuring the users' opinions: the 'Website Evaluation Questionnaire' (WEQ), and the 'user page review' method which invites participants to review website pages. The third method, the think-aloud method, measures website quality during participants' task performance. We studied two variants of this method combined with eye tracking: the concurrent think-aloud method (CTA), which involves participants who verbalize their thoughts while performing tasks, and the retrospective think-aloud method (RTA) with participants who verbalize their thoughts afterwards while watching a recording of their performance.