

# Testing the usability of a personalized system: comparing the use of interviews, questionnaires and thinking-aloud

Lex van Velsen  
University of Twente  
l.s.vanvelsen@utwente.nl

Thea van der Geest  
University of Twente  
t.m.vandergeest@utwente.nl

Rob Klaassen  
University of Twente  
r.f.klaassen@utwente.nl

## Abstract

*Personalized systems present each user with tailored content or output. Testing the usability of such a system must take some specific usability issues and the suitability of the personalized output into account. In this study, we evaluated a personalized search engine to compare the use of interviews, questionnaires and concurrent thinking-aloud to this avail. The interview and the questionnaire are the best methods to elicit comments on usability issues. Concurrent thinking-aloud turned out to be the best method to elicit comments on the perceived relevance of search results. When testing the usability of a personalized system it is wise to use a combination of concurrent thinking-aloud and the interview or the questionnaire.*

*Keywords: personalization, usability testing, internet searching.*

## Introduction

Personalization is a new phenomenon which can be found increasingly on the internet. A personalized website is a site that can provide a tailored appearance or content to visitors. This tailored output can be based upon explicit user input (e.g., book recommendations based upon a user's explicit statement that (s)he is interested in books about Italy), in which case the website is 'adaptable'. Output can also be based upon implicit input (e.g., book recommendations based upon the assumption that a user is interested in Italy because (s)he looked at books about Rome, Venice and Leonardo da Vinci), in which case the website is 'adaptive'. These assumptions are saved in a user model. This is a personal file in which user characteristics are stored. Personalization uses adaptability and / or adaptivity.

The evaluation of personalized systems is new to many usability testers. Traditional evaluation methods rely on the assumption that presentation and content are the same for every user in every context. In the case of personalization, this assumption no longer holds. Furthermore, personalization changes the user experience

and the role of traditional usability issues must be reconsidered.

In this paper we compare interviews, questionnaires and concurrent thinking-aloud (further referred to as thinking-aloud) with regard to their suitability for evaluating personalization. Our goal is to determine which of the issues that come with personalization can be assessed best with the various methods.

## Theoretical background

A literature review of evaluations of personalized systems since 1990 has shown that questionnaires and interviews are exceedingly popular instruments [1]. This may have its consequences for the way we perceive personalization. According to Tsandilas and Schraefel, a lack of knowledge concerning the perceived usefulness of personalized systems is due to a lack of thorough evaluation studies [2]. In order to understand the way users really interact with, and appreciate personalized systems, qualitative evaluation approaches may be more useful than quantitative evaluation approaches, like questionnaires [3].

## Usability testing of traditional systems

Past research on usability evaluation of traditional systems has provided some insights on the suitability of popular evaluation methods to generate comments on usability issues. A study by Henderson et al. [4] showed that thinking-aloud identifies more usability issues than interviews, questionnaires and data log analysis. They also found the added value of using other methods in combination with thinking-aloud to be limited. Thinking-aloud may be considered a useful instrument to identify dissatisfying features of a website [5]. Interviews, and semi-structured interviews in particular, can elicit user feedback on a given, more general topic [6]. Questionnaires with Likert scales can pinpoint problem areas, but are unable to give information on the nature of the identified problem. The use of open-ended questions can solve this problem [4]. According to Kushniruk and Patel [7], questionnaires are not useful when one wants to

receive input for the redesign of a system. A similar result was found by Ebling and John [8]. Their study showed that questionnaires reveal only a small amount of critical issues compared to thinking-aloud.

Based on the studies described above, we conclude that questionnaires and interviews are suited best to generate feedback on general topics concerning a system or website. Thinking-aloud may be more suitable for detecting problems.

### Usability testing of personalized systems

Apart from traditional usability issues, personalized websites have to deal with some other issues as well. These issues are not new, but have different implications for personalization. Jameson [9] has listed the following:

1. Predictability. Users must be able to predict the consequences of their own actions.
2. Comprehensibility. Users must be able to understand how the system works.
3. Controllability. Users must be able to control the system.
4. Unobtrusiveness. Users must be able to complete their tasks without being distracted by the system.
5. Privacy. Users may not have the feeling that their privacy is being infringed upon.
6. Breadth of experience. Users may not lose the possibility of discovering something new because output only complies with their interests.
7. System competence. Users may not have the feeling that the system does not fulfil tasks successfully.

Supplementary to specific usability issues, the evaluation of personalization has some other implications too. We will address two. First, adaptivity grows during interaction with a system, since more information about the user becomes available. In order for a user to notice the effect of personalization, a certain amount of interaction is needed. This asks for an investment by the user [10]. Second, evaluation of personalized systems should be done with a heterogeneous group of participants. This way, personalization is tested with different users in their own context and more (different) issues can be identified [11].

As stated before, the use of different evaluation methods has been studied on the assumption that system output is stable, which is not the case for personalized systems. As a result, the findings of previous studies can not simply be generalized to the situation of personalized systems. This study seeks to identify what kinds of results concerning personalization are generated when a personalized system is evaluated by means of interviews, questionnaires or thinking-aloud.

### Research question and hypotheses

We address the following research question in this study:

*What kind of (unique) issues do interviews, questionnaires or thinking-aloud elicit when evaluating a personalized system?*

Since no studies on the contributions of different methods for the evaluation of personalized systems have been conducted yet, our hypotheses will be based on the findings of studies done with non-personalized systems. Our first hypothesis addresses the quality of thinking-aloud to elicit dissatisfying features [5]. In order to do this, the method will have to elicit the most negative comments.

**H1** Thinking-aloud elicits more negative comments from participants than interviews and questionnaires.

Interviews and questionnaires should collect more comments on general issues. In this study, general issues comprise the specific set of usability issues for personalized systems [9] and the appreciation of personalization.

**H2** Interviews and questionnaires elicit more comments from participants on usability issues for personalized systems and appreciation of personalization than thinking-aloud.

In order to address the usefulness of a personalized system, tailored output should be useful for every user in every context of use. Since we will compare the methods by evaluating a personalized search engine, we must define usefulness in terms of perceived relevance of search results [12]. Thinking-aloud should be suited best to assess (dis)satisfaction with search results.

**H3** Thinking-aloud elicits more comments from participants on perceived relevance of search results than interviews and questionnaires.

Finally, interviews and questionnaires collect general comments and thinking-aloud focuses on more detailed issues. As a result, the collected set of issues should differ per method.

**H4** Thinking-aloud elicits a different set of issues than interviews and questionnaires.

## Method

In this section we will describe how we tested our hypotheses, by using a personalized search engine. But first we will discuss how we accounted for the implications of evaluating personalized systems.

### Accounting for implications of evaluating personalization

In our evaluation, we had to make the system ‘learn’ enough about our participants to make them notice the effect of personalization within two hours. We did this by creating a set of tasks on one topic that is well documented on the internet: city trips. Participants were asked to search for a youth hostel to their liking and for the museum of modern art in four large European cities. Cities were randomized since for some cities it is easier to find the right information than for others, and search efforts can be kept to a minimum. When this happens, the system learns less about the user, because there is less user-system interaction to base personalization on. Some participants ‘benefited’ from starting the evaluation with a badly documented tourist destination, others did not.

We were unable to gather a heterogeneous group of participants for practical reasons. This may limit the identified issues of our evaluation to a specific group of users, in our case students. However, in this study our goal is to compare evaluation methods and in this light, a group of homogeneous participants may be beneficial. Differences in issues can not be attributed to different user groups, but to the evaluation method only.

### Evaluation procedure

We tested the hypotheses by evaluating a personalized search engine, called Prospector [13]. Prospector re-ranks search results according to user interests and user ratings of earlier visited search results. A user with a high interest in sports but a low interest in arts will, when searching on ‘the Netherlands’, receive a set of search results in which results about the national soccer team will be ranked higher than results concerning the Van Gogh museum. Thus, the most relevant results are to be placed on top. Combined, indications of user interests and user ratings of pages create the user model. Prospector users can view and alter the user model.

In a usability laboratory, 32 students of communication sciences and psychology worked with Prospector. All participants started the evaluation by completing a questionnaire on demographics and computer usage. Next, they were presented with the same set of tasks to be performed with Prospector. A 4 x 8 design was applied. Participants thought aloud and completed a questionnaire (n = 8), or were interviewed after speaking their thoughts

out loud (n = 8). The remaining participants did not think out loud, but completed a questionnaire after interacting with Prospector (n = 8) or were interviewed (n = 8).

In the case of thinking-aloud, an experimenter sat next to the participant and participants were explained what they were expected to do. Thinking-aloud was practiced first by looking up a train schedule on the internet. The experimenter explained he would not answer any questions, but would only remind the participant to think-aloud if necessary.

The interviews and questionnaire both addressed usability issues for personalized systems, appreciation of personalization, intention to use and the strong and weak points of Prospector. Interviews were semi-structured: the interviewer could ask questions that came up during the interview. The questionnaire consisted of two parts. A part with closed questions, in which participants had to answer items on the afore mentioned topics by scoring 7-point Likert scales. In the second part, these topics were addressed again by means of open questions. In the analyses addressed in this paper, we only focus on the open questions.

### Data analysis

All think-aloud sessions and interviews were audio recorded. Next, the comments about interacting with Prospector were identified. For each comment the following points were determined:

1. Is the comment related to personalization and if so, to which usability issue or topic directly related to personalization (appreciation of personalization or perceived relevance of search results);
2. Is the comment positive, negative, or neutral or ambiguous;

Next, all comments that concerned the same topic (e.g., “receiving search results concerning news” or “comprehensibility concerning the compilation of the user profile”) were grouped.

## Results

Combined, the three methods helped us to collect 555 comments on personalization. The interviews elicited 166 issues, the questionnaires 227, and the thinking-aloud sessions 162.

### Positive, neutral and negative issues

Table 1 shows the average amount of positive, neutral or negative issues each session elicited on the topic of personalization. In this table, one number needs some clarification.

**Table 1. Amount of differently valued comments, elicited by each method.**

| Method         |          | Positive   | Neutral    | Negative   |
|----------------|----------|------------|------------|------------|
| Questionnaire  | Mean     | 3.94 (T)   | 1.56       | 4.88       |
|                | St. dev. | 2.54       | 1.32       | 3.14       |
| Interview      | Mean     | 6.44 (Q/T) | 3.13 (Q/T) | 4.63       |
|                | St. dev. | 2.45       | 1.82       | 2.99       |
| Thinking-aloud | Mean     | .81        | 1.00       | 8.31 (Q/I) |
|                | St. dev. | 1.05       | 1.21       | 2.75       |

Note: A letter behind a mean value means that the mean value is significantly higher than the mean value of the method the letter corresponds with. Q stands for questionnaire, I for interview, and T for thinking-aloud.

Thinking-aloud elicits on average a large number of negative comments. These comments were mostly concerned with negatively perceived relevance of search results and participants' argumentation for this statement (on average 5.38 comments per thinking-aloud participant).

Next, we performed ANOVA analyses to determine whether the amount of positive, neutral and negative

issues each method elicited differed. The results of these ANOVA analyses can be found in table 2. They show significant differences between groups in the case of positive, neutral and negative comments. In order to determine *which* groups differed, we conducted Post Hoc analyses with Bonferroni tests where we applied a 5% significance level. Results of the Post Hoc analyses can be found in table 1. The questionnaire yields more positive results than thinking-aloud. The interview collects more positive and neutral comments than the questionnaire, as well as thinking-aloud. Thinking-aloud finally, gathers more negative comments than questionnaires and interviews.

### Usability issues caused by personalization

During data analysis we determined to which usability issue concerning personalization a comment was related, or whether a comment concerned appreciation of personalization or perceived relevance of search results. Table 3 shows how many comments on each usability issue every method elicited.

Again, we performed ANOVA analyses to detect differences between the numbers of issues each method

**Table 2. Results of the ANOVA analyses for differences among differently valued comments**

|                 |                | Sum of squares | DF | Mean square | F-value | Sig. |
|-----------------|----------------|----------------|----|-------------|---------|------|
| Positive issues | Between groups | 254.17         | 2  | 127.08      | 28.13   | <.01 |
|                 | Within groups  | 203.31         | 45 | 4.52        |         |      |
|                 | Total          | 457.48         | 47 |             |         |      |
| Neutral issues  | Between groups | 38.79          | 2  | 19.40       | 8.94    | <.01 |
|                 | Within groups  | 97.69          | 45 | 2.17        |         |      |
|                 | Total          | 136.48         | 47 |             |         |      |
| Negative issues | Between groups | 135.88         | 2  | 67.94       | 7.74    | <.01 |
|                 | Within groups  | 394.94         | 45 | 8.78        |         |      |
|                 | Total          | 530.81         | 47 |             |         |      |

**Table 3. Number of comments on usability, elicited by each method**

| Method         |          | Predictability | Comprehensibility | Controllability | Unobtrusiveness | Privacy  | Breadth of experience | System competence |
|----------------|----------|----------------|-------------------|-----------------|-----------------|----------|-----------------------|-------------------|
| Questionnaire  | Mean     | 1.00           | 2.19              | .75 (T)         | 1.06 (T)        | 1.50 (T) | 1.44                  | 1.00              |
|                | St. dev. | .37            | .75               | .45             | .68             | .63      | .89                   | .63               |
| Interview      | Mean     | 1.19           | 2.19              | 1.13 (Q/T)      | 1.56 (T)        | 1.63 (T) | 2.19 (T)              | 1.81 (T)          |
|                | St. dev. | .54            | 1.05              | .50             | 1.03            | 1.03     | 1.17                  | 1.38              |
| Thinking-aloud | Mean     | .69            | 1.50              | .06             | .06             | .06      | 1.00                  | .89               |
|                | St. dev. | .87            | 1.32              | .25             | .25             | .25      | .97                   | .50               |

Note: A letter behind a mean value means that the mean value is significantly higher than the mean value of the method the letter corresponds with. Q stands for questionnaire, I for interview, and T for thinking-aloud.

**Table 4. Results of the ANOVA analyses for differences among the amounts of usability issues and topics directly related to personalization**

| Usability issue / topic               |                | Sum of squares | DF | Mean square | F-value | Sig. |
|---------------------------------------|----------------|----------------|----|-------------|---------|------|
| Predictability                        | Between groups | 2.04           | 2  | 1.02        | 2.57    | n.s. |
|                                       | Within groups  | 17.88          | 45 | .40         |         |      |
|                                       | Total          | 19.92          | 47 |             |         |      |
| Comprehensibility                     | Between groups | 5.04           | 2  | 2.52        | 2.23    | n.s. |
|                                       | Within groups  | 50.88          | 45 | 1.13        |         |      |
|                                       | Total          | 55.92          | 47 |             |         |      |
| Controllability                       | Between groups | 9.29           | 2  | 4.65        | 27.20   | <.01 |
|                                       | Within groups  | 7.69           | 45 | .17         |         |      |
|                                       | Total          | 16.98          | 47 |             |         |      |
| Unobtrusiveness                       | Between groups | 18.67          | 2  | 9.33        | 17.64   | <.01 |
|                                       | Within groups  | 23.81          | 45 | .53         |         |      |
|                                       | Total          | 42.48          | 47 |             |         |      |
| Privacy                               | Between groups | 24.13          | 2  | 12.06       | 23.93   | <.01 |
|                                       | Within groups  | 22.69          | 45 | .50         |         |      |
|                                       | Total          | 46.81          | 47 |             |         |      |
| Breadth of experience                 | Between groups | 11.54          | 2  | 5.77        | 5.60    | <.01 |
|                                       | Within groups  | 46.38          | 45 | 1.03        |         |      |
|                                       | Total          | 57.92          | 47 |             |         |      |
| System competence                     | Between groups | 14.04          | 2  | 7.02        | 6.80    | <.01 |
|                                       | Within groups  | 46.44          | 45 | 1.03        |         |      |
|                                       | Total          | 60.48          | 47 |             |         |      |
| Appreciation of personalization       | Between groups | 15.50          | 2  | 7.75        | 17.66   | <.01 |
|                                       | Within groups  | 19.75          | 45 | .44         |         |      |
|                                       | Total          | 32.25          | 47 |             |         |      |
| Perceived relevance of search results | Between groups | 314.29         | 2  | 157.15      | 61.13   | <.01 |
|                                       | Within groups  | 115.69         | 45 | 2.57        |         |      |
|                                       | Total          | 429.98         | 47 |             |         |      |

elicited. The results of these analyses can be found in table 4. They show that there are significant differences among the methods for all usability issues, except predictability and comprehensibility. The different methods all elicit the same numbers of comments on these two topics. For the remaining five usability issues, we conducted Post-Hoc analyses by means of Bonferroni tests to ascertain which groups differed. As before, we applied a 5% significance level. The results of these tests can be found in table 3. We found that for the remaining five usability issues the interview always elicited more comments than thinking-aloud. In the case of controllability, the method also elicited more comments than the questionnaire. The questionnaire itself gathered more comments than thinking-aloud on the topics of controllability, unobtrusiveness and privacy.

### Appreciation and perceived relevance

Besides a usability issue, a comment could also be classified as a statement on appreciation of personalization or the perceived relevance of search results. Table 5 shows the numbers of comments questionnaires, interviews and thinking-aloud elicited on these topics.

In table 4, the results of ANOVA analyses that were conducted in order to identify differences among the amount of comments each method elicited on these topics are included. They show that these amounts differ for both topics. Post-Hoc analyses using Bonferroni tests and a 5% significance level show that for the topic of appreciation of personalization, the questionnaire and the interview elicit more comments than thinking-aloud.

**Table 5. Numbers of comments on appreciation of personalization (Appr.) and perceived relevance of search results (relevance), elicited by each method.**

| Method         |          | Appr.    | Relevance  |
|----------------|----------|----------|------------|
| Questionnaire  | Mean     | 1.00 (T) | .44        |
|                | St. dev. | .82      | 1.03       |
| Interview      | Mean     | 1.50 (T) | 1.00       |
|                | St. dev. | .73      | 1.03       |
| Thinking-aloud | Mean     | .13      | 6.13 (Q/I) |
|                | St. dev. | .34      | 2.36       |

Note: A letter behind a mean value means that the mean value is significantly higher than the mean value of the method the letter corresponds with. Q stands for questionnaire, I for interview, and T for thinking-aloud.

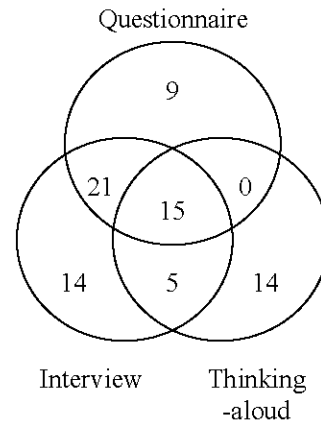
However, in the case of perceived relevance of search results a different result was found. Here, thinking-aloud gathered more comments than the questionnaire and the interview.

### Identified issues

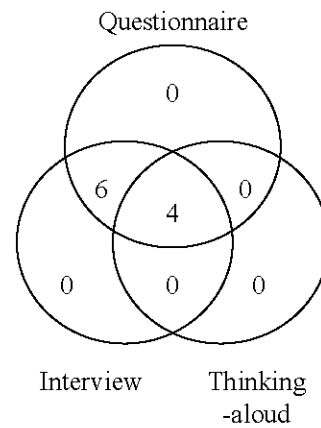
After looking at the amounts of comments each method gathered, it is interesting to look at the content of these comments. We have categorized the comments into topics. Examples of these topics are “receiving the same search result more than once in one set of search results” or “comparisons with Google”. In figure 1 we have displayed the contribution of each method to the total collection of topics that were raised during the evaluation. It shows that each method has contributed a unique set of topics. A considerable number of topics was covered by two or three methods.

A closer look at the contributions of each method to the collection of identified topics that are differently valued shows the same picture. Each method helps to discover a unique set of topics and there is again a large set of topics that is covered by multiple methods, whether we focus on positive, neutral or negative comments.

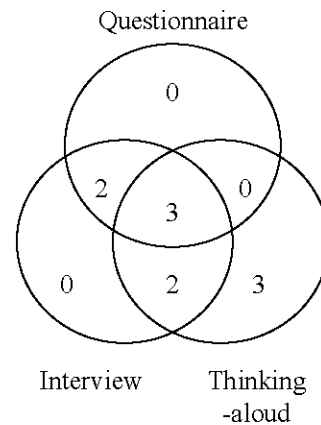
A different picture is disclosed when we concentrate on the ten topics that received the most positive comments. As figure 2 shows, thinking-aloud does not contribute any unique topics. Four of these topics are elicited by all methods, six of them by the questionnaire as well as the interview. The top ten topics that received the most negative comments, on their turn, show a different picture than their positive counterparts. Figure 3 shows a picture in which thinking-aloud is the only method contributing unique topics.



**Figure 1. Topics covered by each method**



**Figure 2. The ten topics that received the most positive comments, identified by each method**



**Figure 3. The ten topics that received the most negative comments, identified by each method**

The two topics that were not elicited by thinking-aloud were uncovered by the interview as well as the questionnaire.

## Conclusions

The results demonstrate the different weak and strong points of each method. When one chooses an evaluation method, one should take them into account. For example, when one wants to assess feelings of privacy, using thinking-aloud will not result in a lot of comments on the topic of interest.

The results support our first hypothesis. Thinking-aloud generates more comments on dissatisfying features than questionnaires and interviews.

Our second hypothesis partly holds. Questionnaires elicit more comments on some usability issues and appreciation of personalization than thinking-aloud. Interviews elicit more comments on most usability issues and appreciation of personalization than thinking-aloud.

Hypothesis 3 is accepted. This is partly the result of the fact that thinking-aloud elicited most negative comments. In our evaluation, perceived relevance was heavily influenced by negative experiences with the Prospector search engine and participants experienced the personalized output while they were thinking-aloud.

Finally, our fourth hypothesis is also partly supported. Thinking-aloud uncovers a unique set of issues with the system. However, a large part of issues demonstrated by thinking-aloud were also detected by questionnaires or interviews.

## Characteristics of the three methods

The open questions of the survey yielded more comments on three usability issues than thinking-aloud. Based on this result, one should prefer the questionnaire with open-ended questions above thinking-aloud for the assessment of usability issues of personalization.

The interview appears to be the most fruitful instrument to elicit comments concerning the usability issues, since it collects more comments than thinking-aloud. For the issue of controllability, the interview is also more fruitful than the questionnaire. When one wants to elicit many comments on the usability issues concerning personalization, the interview is the best method. Nevertheless, the interview has one downside. The comments appear to be positively biased. This may be due to the wish of interviewees to keep face by, for example, not admitting that they do not understand the system, or they may want to give socially desirable answers.

Thinking-aloud is of little value when one wants to receive comments on usability issues or appreciation of personalization. Since these topics are abstract, they do

not come to the mind of the participant while interacting with the system. There is one topic of feedback for which thinking-aloud was the best elicitation technique in this study: perceived relevance of search results. In the domain of personalized searching, this may be the most important condition for user acceptance. Therefore, we must conclude that thinking-aloud is an important part of the iterative design process, even though the surplus value may be limited to just one issue.

## Matching instrument characteristics and evaluation goals

The iterative design process of a personalized system can include a stage in which one has to detect problems in order to generate input for redesign. At this stage, thinking-aloud should be combined with either questionnaires or interviews. Thinking-aloud was found to be the sole supplier of a considerable amount of important topics that hindered the participant to work with the personalized search engine to their satisfaction. The remaining issues that thinking-aloud does not uncover can then be assessed by questionnaires or interviews.

At a later stage of the design process, when prototyping is done and a full system is developed, one may be more interested in the user perception of the usability issues of personalization. Although the interview elicited most comments on these topics, the questionnaire may be a more valid means. Interviews give a positively biased view of these psychological constructs and hence, can misinform the evaluator.

## Future work

In a next stage, we will analyze the comments on issues not related to personalization in the same manner. We will then be able to draw a comparison between the two sets of data. As a result, we can make statements on the suitability of the three methods for system evaluation in which issues concerning personalization and other issues (e.g., lay-out or clarity of error messages) are to be assessed at the same time.

## Acknowledgement

We want to thank our colleagues of the Institute for Information Processing and Microprocessor Technology of the Johannes Kepler University in Linz, Austria for their assistance and useful suggestions during this evaluation.

## References

- [1] L. S. van Velsen, T. M. van der Geest, R. F. Klaassen, and M. F. Steehouder, "User-centered evaluation of adaptive and adaptable systems: a literature review," University of Twente, 2007.
- [2] T. Tsandilas and M. C. Schraefel, "Usable adaptive hypermedia systems," *New review of hypermedia and multimedia*, vol. 10, pp. 5-29, 2004.
- [3] C. Gena, "Methods and techniques for the evaluation of user-adaptive systems," *The knowledge engineering review*, vol. 20, pp. 1-37, 2005.
- [4] R. D. Henderson, M. C. Smith, J. Podd, and H. Varela-Alvarez, "A comparison of the four prominent user-based methods for evaluating the usability of computer software," *Ergonomics*, vol. 38, pp. 2030-2044, 1995.
- [5] R. Benbunan-Fich, "Using protocol analysis to evaluate the usability of a commercial web site," *Information & management*, vol. 39, pp. 151-163, 2001.
- [6] E. Fossey, C. Harvey, F. McDermott, and L. Davidson, "Understanding and evaluating qualitative research," *Australian & New Zealand journal of psychiatry*, vol. 36, pp. 717-733, 2002.
- [7] A. W. Kushniruk and V. L. Patel, "Cognitive and usability engineering methods for the evaluation of clinical information systems," *Journal of biomedical informatics*, vol. 37, 2004.
- [8] M. R. Ebling and B. E. John, "On the contributions of different empirical data in usability testing," in *Designing interactive systems: processes, practices, methods, and techniques*. New York City, USA, 2000.
- [9] A. Jameson, "Adaptive interfaces and agents," in *Human-computer interaction handbook (2nd ed.)*, J. A. Jacko and A. Sears, Eds. Mahwah, NJ: Erlbaum, 2006.
- [10] K. Höök, "Evaluating the utility and usability of an adaptive hypermedia system," in *Intelligent user interfaces*. Orlando, Florida, 1997.
- [11] S. Weibelzahl, "Problems and pitfalls in evaluating adaptive systems," in *Adaptable and adaptive hypermedia systems*, S. Chen and G. Magoulas, Eds. Hershey, PA: IRM Press, 2005, pp. 285-299.
- [12] D. Nahl, "Ethnography of novices' first use of web search engines: affective control in cognitive processing," *Internet reference services quarterly*, vol. 3, pp. 51-72, 1998.
- [13] C. Schwendtner, F. König, and A. Paramythis, "Prospector: an adaptive front-end to the Google search engine," in *14th workshop on adaptivity and user modeling in interactive systems, held in conjunction with Lernen, Wissensentdeckung und Adaptivität*. Hildesheim, Germany, 2006.

## About the authors

Lex van Velsen is a Ph.D. candidate in the Technical and Professional Communication Department at the University of Twente (Enschede, the Netherlands). His research interests include personalized communication systems and user satisfaction with user support.

Thea van der Geest is an associate professor in the Communication Studies/Technical Communication Department of the University of Twente (Enschede, The Netherlands). She teaches courses in interface and interaction design, web design and research methodology. Her research focuses on the process of web design, methods for evaluating web sites, usability and accessibility, and information and document design for new media.

Rob Klaassen is an expert in the field of government communication. Between 1995 and 2007 he worked at the University of Twente (Enschede, The Netherlands). In 2004 he completed a PhD thesis on Mass media public information campaigns. Other research projects included studies into government webportals, citizens' information needs and search strategies, and evaluation methods for personalized systems. Klaassen currently works as a research consultant for the Netherlands Government Information Service.