Mobile Questionnaires for User Experience Evaluation

Heli Väätäjä
Tampere University of Technology
Korkeakoulunkatu 6
33720 Tampere, Finland
heli.vaataja@tut.fi

Virpi Roto
Tampere University of Technology
AND Nokia Research Center
P.O.Box 407
00045 Nokia Group, Finland
virpi.roto@nokia.com

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
As user experience studies move from laboratories to mobile context, we need tools for collecting data in natural settings. Based on the results from a pilot study, we present early guidelines for designing mobile questionnaires to be filled in on handheld, palm-sized mobile devices. We found that special attention needs to be paid to the clarity and simplicity of the structure, layout and questionnaire content, including questions, visual icons, items and scales. In addition to the requirements set by the screen size, also data entry method, interaction style and mobile context related issues need to be taken into account when designing questionnaires for mobile devices.

Keywords
Mobile devices, questionnaires, user experience, field study, evaluation, experience sampling

ACM Classification Keywords
H5.2 [User Interfaces]: Evaluation/methodology

General Terms
Experimentation, Measurement, Human Factors

Introduction
Mobile questionnaires offer an interesting alternative for sampling [3] and measuring of user experience in

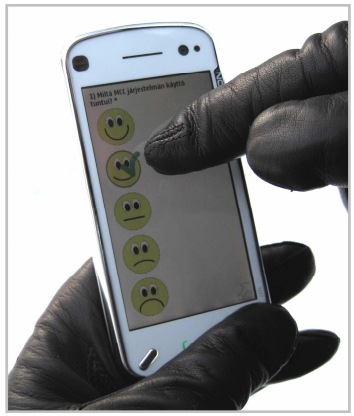

Figure 1. Scale for measurement of valence.

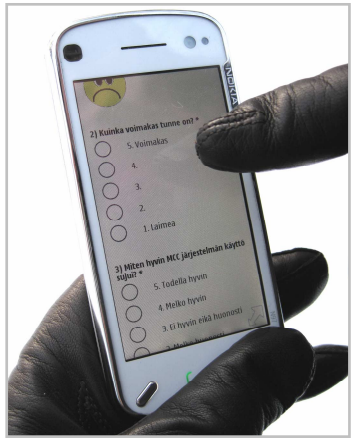

Figure 2. Scale for measurement of arousal.

field studies. Instead of having an experimenter present in usage situations or having questionnaires on paper, mobile questionnaires can be sent to participants to natural use contexts. By mobile questionnaires we refer to questionnaires, which respondents fill in on handheld, palm-sized mobile devices, such as mobile phones. Mobile questionnaires can be implemented by using SMS (text messages), MMS (multimedia message), online questionnaires aimed for mobile Web browsers, or customized mobile client applications. A number of solutions have been developed and used for user experience sampling (e.g. [5][8][9]), but guidance on developing mobile questionnaires is missing.

The characteristics of the mobile phones and mobile context pose special challenges to the design and usage of mobile questionnaires, especially when multiple questions or items are included. Special features of the mobile phones and wireless networks that need to be taken into account include 1) screen size, 2) data entry method 3) battery power, 4) connectivity issues 5) processing power as well as 6) mobile context [6][7][12]. In this paper, we discuss some of these issues in regard to mobile questionnaire development based on our experiences in the design phase and findings from a pilot study.

Our goal is to utilize mobile questionnaires in natural settings for user experience evaluation in our future field studies. In this paper, we first present two questionnaires to be used in mobile user experience studies. We then present findings from a pilot study with six usability experts as participants, who pre-tested mobile Internet based questionnaires. We conclude by presenting early guidelines for mobile questionnaire design based on the study and considerations made during the questionnaire development and present directions for future research.

Design of mobile questionnaires

Our field studies are conducted in the context of mobile news journalism where participants produce news stories and multimedia material using a mobile phone with a mobile client application for journalism. During the field studies, our aim is to repeatedly sample and measure user experience related aspects after carrying out a task (an entire assignment from receiving a mobile task to submission) or a sub-task (like receiving a task with the mobile application, photographing, or creating the story with the application). For this purpose, we included four questions in the first mobile questionnaire, inspired by [1][4][10]: 1) two questions on user's experienced emotions based on SAM with valence and arousal, see Figures 1 and 2, 2) satisfaction with the system and 3) satisfaction with the outcome of usage. We chose a 5-point scale for all four questions, to support fitting each question as an entity to the screen. In development phase, we tested paging of the four questions, each on separate page. We discarded this implementation, since even in a stable situation, there were considerable download delays, which are not feasible for users in mobile context [6].

In addition, at several points during the field studies (for example, before usage, after completing a specified number or type of tasks, and after the last task), we assess users' pragmatic and hedonic quality perceptions and judgment of appeal [2] with a mobile questionnaire based on the Attrak-Work questionnaire [11]. For the purposes of mobile questionnaire, we had to reduce the number of items from 36 to 14 and

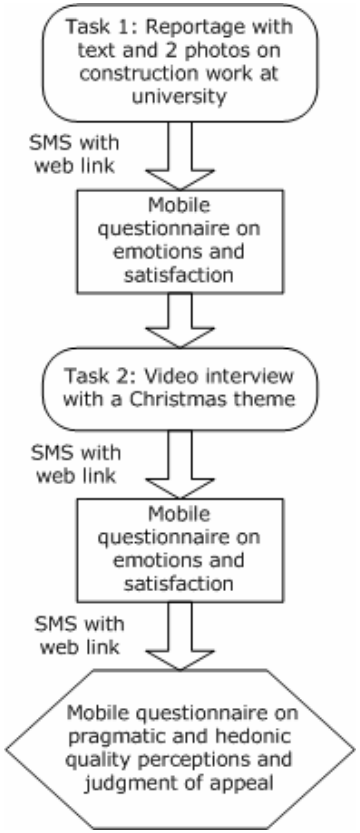

Figure 3. Procedure of pilot test tasks and mobile questionnaires.

change the scale from a 5-point semantic differential with two anchors to a 5-point Likert scale with one anchor. This was done, since even with a wide (3,5") screen, having two anchors and a 5-point scale was too wide to be viewed and answered in a convenient way for the user. For the mobile version, we selected either of the original anchors, so that also reversed anchors were chosen. As an example from item "Professional-Amateurish" we chose "Professional" for the mobile questionnaire. Furthermore, we shortened one attribute pair (item), so that it consisted of two words in Finnish. We grouped the 14 items into one question asking for rating of all items.

Procedure of the pilot tests

The purpose of the pilot tests was to pre-test and get feedback on the developed questionnaires to increase their validity and reliability as well as to test the procedure of administering the mobile questionnaires. In addition, we wanted to test how the two different types of questionnaires suit the mobile phones.

Pilot test participants were six usability researchers, who had used the test mobile phone model (Nokia N97 with 3,5" touch screen) and the mobile journalism application prototype previously in a heuristic walkthrough. The pilot test was conducted in laboratory context with an experimenter. The experimenter explained the test scenario where the participant acts as a journalist, and instructed the participant to think aloud during the test. The experimenter made notes on the participants' impressions and comments, focusing especially on the mobile questionnaire related findings. Immediately after the test, participant was interviewed on the procedure of using mobile questionnaires and for the used questionnaires in detail.

We planned the pilot test procedure with two realistic tasks. For practical reasons, the tasks were sent prior to the test to the test phone via the mobile journalism application. Similarly, three SMS messages with Internet links to the mobile questionnaires were sent to the test phone prior to the test. The participants were advised to utilize them at certain points of the study.

The first task asked for a 350-character-long reportage with two photos about the upcoming construction work at the university. The second task asked for a 10-30s long video interview with a Christmas related theme and a short descriptive text on the content of the video. Right after completing each task, the experimenter indicated that an SMS arrived. Participants then opened the SMS with a web link to the mobile questionnaire on emotions and satisfaction and answered to the questionnaire. At the end of the test, experimenter indicated again that an SMS arrived, with the link to the mobile questionnaire on users' quality perceptions and judgment of appeal. Pilot test procedure for tasks and mobile questionnaires is illustrated in Figure 3.

Results from the pilot tests

To gain a deeper understanding of the feasibility aspects of the mobile questionnaires for field studies, we analyzed the times to complete the two designed questionnaires. The questionnaire on emotions and satisfaction was filled in twice by each respondent. Figure 4 illustrates the completion times after tasks 1 and 2. Figure clearly shows the learning effect in filling in the questionnaire. When filling in the questionnaire for the second time the completion time was reduced by 1-2 minutes for four participants.

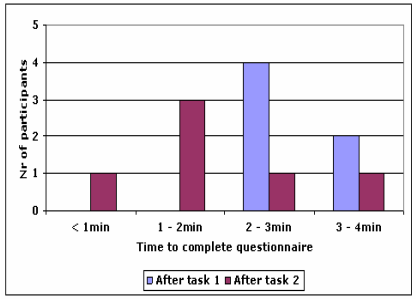

Figure 4. Time to complete questionnaires after task 1 (light blue) and task 2 (dark brown).

Figure 5. Example of scale and items from questionnaire on quality perceptions and overall judgment of appeal.

The first participant needed as much as 7 minutes to complete the first version of the questionnaire on quality perceptions and judgment of appeal, which had 14 items grouped into one question. She had to scroll up to check the scale from the top of the of item list several times, which indicated that there were too many items for a single question to fit the small screen. The respondent needs to easily see or check the scale fast. Therefore, after the first pilot test, we changed the questionnaire by regrouping the 14 items to three separate questions with 4 (see Figure 5), 5, and 5 items in each question asking for the rating in a similar manner. After the change in the structure, the completion times for four participants were 3, 4, 5 and 7 minutes. The participant with 7 minute answering time with second version had also otherwise trouble in interaction with the touch screen. One measurement was excluded from this analysis, since the answering process was not comparable with other tests.

All participants found the length of the questionnaire on emotions and satisfaction convenient. One of them described: "Questions are not too in-depth, they are easy and fast to answer". All respondents found the faces clear and easy to interpret as well as the selection of a face by touching easy and their size appropriate. For arousal all except one participant found the question easy to answer. However, two participants suggested that three levels instead of five would be sufficient as a scale for arousal. For the third and fourth questions on satisfaction with the system, and on the outcome of usage, one participant interpreted the questions to ask for an evaluation of her own success. Therefore, for future field studies we consider revising the wording of the questions on satisfaction and when possible use validated one-question measures as [10].

Although the time to fill in the questionnaire assessing pragmatic and hedonic quality perceptions and judgment of appeal was longer, all except the first participant, who filled in the first version with 14 items in one question, found the length of the questionnaire to be convenient. After revising the single question with 14 items to three questions, two of the participants commented on the same question being asked for three times. On the other hand, it was mentioned that the layout and structure motivated to answering. Furthermore, the used question type with 1-2 word attributes as items rated on a 5-point scale was found compact and focusing on central issues. However, from two to four participants interpreted six of the 14 attributes either in two ways or commented that attribute needs to be rated in relation to something. We therefore need to replace them with unambiguous attributes or have a dilemma of needing to make longer items, which is in contrast to needed compactness.

The procedure in which questionnaires were filled in right after using the system was commented to be easy by the participants since usage was fresh in their mind. Participants found the questions to be summative and giving an overall view to user experience related aspects. In relation to the screen size, participants found that the tilting screen was convenient, since it enabled more easily to view the question and the entire scale, which made answering easier to an item. In addition, when answering, it was found important to be able to zoom easily, so that one can see a larger entity from the whole screen. All in all, participants emphasized that it is important to gain directly an overview of an entire question, including the question or statement, items and the scale.

Early guidelines

We found, that the four most important things to consider in the development of the mobile questionnaires are 1) small screen size, 2) data entry method and interaction style, 3) mobile context and 4) chosen implementation for the questionnaire (online questionnaire, SMS/MMS, mobile client). The following early guidelines, which are based on the pilot study and our consideration for specific mobile system and context related issues in development phase of presented questionnaires, reflect these issues:

1. Minimize the total time to complete the questionnaire, including accessing, opening, reading, answering, and submitting. For a questionnaire that is filled in frequently and/or in mobile context, completion time is especially critical. For example, interruptions, multitasking, mobility of users and social situation call for minimizing the time to answer a questionnaire.
2. Limit the length of the questionnaire. Include a small number of short, well-focused, easily comprehensible questions and items to minimize cognitive load and needed concentration.
3. Fit one question on the screen with the entire scale and all its items. If this is not possible, check that the respondent can effectively scroll the question or zoom in and out to gain an overview of all options.
4. Consider the number of points in the scale carefully. We found 5-point scales to be convenient for a mobile phone screen (size 3,5"), since they fitted the screen easily with a readable font size.
5. Minimize extra activity to view the questions. Make sure that the size of the text is easily readable without zooming. In a mobile context, the user may only be able to glance at the questionnaire, and zooming may be difficult or impossible due to the device used or due to context related issues.
6. Make selecting an icon or item for answering easy. On touch screen devices, use large enough icons and adequate spacing between items, so that the neighboring items or icons are not selected by accident.
7. Consider whether to use paging of questions to separate pages or one page questionnaires. In the case of online questionnaires, paging causes undesirable download delays [6]. For touch screens we found it better to avoid paging due to these delays, since scrolling is easy. On the other hand, for conventional screens, the tradeoff between download delays and more cumbersome interaction needs to be considered.

In practice, designing a mobile questionnaire means balancing between presented guidelines, which may contradict each other.

Conclusions and future research

Mobile questionnaires, which are filled in on handheld, palm-sized mobile devices, offer new possibilities to study user experience in natural settings, especially when the experimenter is not present in usage situations. We presented early guidelines for designing multi-question questionnaires for handheld, palm-sized mobile devices based on a pilot study. We pre-tested two mobile questionnaires implemented with a mobile Internet browser on a touch screen mobile phone.

The special features of mobile phones, like small screen size, data entry method and usage in mobile context [6][7][12] need to be considered in the design phase of the mobile questionnaires. Attention needs to be paid to the questionnaire length, layout and structure, as well as to the simplicity and comprehensibility of the questions. Since participants were positive towards mobile questionnaires, we plan to utilize them in our

future field studies after revising some questions and items. In addition, we plan to update the guidelines based on our future studies and report experiences on applying one-question measures.

Further research is needed to identify, develop and test suitable one-question measures (see e.g. [10]) as well as short questionnaires for user experience evaluation with mobile devices. Validated and reliable measures could be utilized in convenient combinations when studying different aspects of user experience in varying research designs. Therefore, studies are needed to ensure the feasibility, validity and reliability aspects when measures and questionnaires are used on a variety of handheld, palm-sized mobile devices in mobile context.

Acknowledgements

We thank Tiina Koponen for her valuable assistance and Timo T. Koskinen from Nokia Research Center for providing the mobile system for the study. This research was funded by Ministry of Education, Finland, UCIT Graduate School and the Academy of Finland.

References

[1] Bradley, M., and Lang, P.J. Measuring emotion – the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry 25* (1994), 49-59.

[2] Hassenzahl, M. The thing and I: Understanding the relationship between users and product. In *Funology: From usability to enjoyment*, M.A. Blythe, K. Overbeeke, A.F. Monk, P.C. Wright, Eds. Kluwer, The Netherlands (2003), 31-42.

[3] Hektner, J. M., Schmidt, J.A. and Csikszentmihalyi, M. *Experience Sampling Method, Measuring the Quality of Everyday Life*. Sage Publications Inc., USA (2007).

[4] Hornbæk, K. Current practices in measuring usability: Challenges to usability studies and research. *Int. J. Hum.-Comput. Stud. 64* (2006), 79-102.

[5] Isomursu, M., Tähti, M., Väinämö, S. and Kuutti, K. Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *Int. J. Hum.-Comput. Stud. 65* (2007), 404-418.

[6] Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K. and Buchanan, G. 1999. Improving web interaction in small screen displays. *Proc. WWW8* (1999), 51-59.

[7] Jumisko-Pyykkö, S. and Vainio, T. Framing the Context of Use for Mobile HCI. *Int. J. of Mobile-Human-Computer-Interaction* (IJMHCI), 2, 1 (2010).

[8] Meschtscherjakov, A., Weiss, A. and Scherndl, T. Utilizing Emoticons on mobile devices within ESM studies to measure emotions in the field. *Proc. MME in conjunction with MobileHCI'09* (2009).

[9] Mulder, I., ter Hofte, G.H. and Kort, J. SocioXensor: Measuring user behaviour and user eXperience in conteXt with mobile devices. *Proc. Measuring Behavior 2005* (2005), 355-358.

[10] Sauro, J. and Dumas, J.S. Comparison of Three One-Question, Post-Task Usability Questionnaires. *Proc. CHI 2009*, ACM Press (2009), 1599-1608.

[11] Väätäjä, H., Koponen, T. and Roto, V. Developing practical tools for user experience evaluation – a case from mobile news journalism. *Proc. ECCE 2009*, ACM Press (2009), 240-247.

[12] Zhang, D. and Adipat, B. 2005. Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. *Int. J. Hum-Comput.Stud*. *18*, 3 (2005), 293-308.