

End-Users' Voice in EHR Selection: Development of a Usability Questionnaire for Demonstrations in Procurement (DPUQ)

Mari TYLLINEN^{a,b,1}, Johanna KAIPIO^a, Tinja LÄÄVERI^{b,c} and Marko NIEMINEN^a

^aDepartment of Computer Science, Aalto University, Finland

^bOy Apotti Ab, Finland

^cUniversity of Helsinki and Inflammation Center, Helsinki University Hospital, Finland

Abstract. This paper describes the development of a questionnaire for evaluating usability during EHR system procurement (DPUQ). Established usability questionnaires can be used to gather user feedback after using the systems. However, during procurement, experimenting with real system use is practical only with a limited number of system candidates. There is a need for less resource-demanding usability evaluation in the early stages of procurement in cases with several vendors. DPUQ has been designed for usability evaluation by end-users during special scenario-based vendor demonstrations. The questionnaire includes three sets of questions to be used during and after the vendor demonstration. DPUQ delivers specific usability scores and can be used to compare system candidates in procurement complementing other evaluation methods.

Keywords. User questionnaire, electronic health record system, end-users, evaluation, procurement, satisfaction, usability

Introduction

Using standardized usability questionnaires (e.g. SUS [1], QUIS [2] and SUMI [3]) is an established practice for post-use usability evaluation, especially during system development. SUS has also been recommended for use in electronic health record (EHR) system procurement [4]. However, in the early stages of procurement the number of vendors can be quite large. Therefore, usability testing might not be a plausible method due to resource requirements. Including usability as criteria in early stages of system procurement needs a more cost-effective method. In this paper, we describe the development of a Usability Questionnaire for Demonstrations in Procurement (DPUQ) in a large-scale (40 000 users) public electronic health and social care record system procurement project "CAPIS". The evaluation is based on seeing the system during a demonstration following a user-scenario. Existing usability questionnaires are developed from the perspective of using the system. However, there is a difference between using a system and seeing it being used. This should be reflected in the usability questionnaire used when evaluating a demonstration.

¹ Corresponding Author, Mari Tyllinen, E-mail: mari.tyllinen@aalto.fi

1. Background: Questionnaires on Usability

In the field of human-computer interaction (HCI) research, several standardized usability questionnaires exist for the assessment of perceived usability or satisfaction. Many of those are to be used as post-study questionnaires after usability tests, e.g. QUIS (Questionnaire for Interaction Satisfaction) [2], SUMI (Software Usability Measurement Inventory) [3], CSUQ (Computer System Usability Questionnaire) [5] and SUS (System Usability Scale) [1].

QUIS [2] can be used as versions including 27, 41 or 122 items. It includes 9 themes: screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, teleconferencing, and software installation. QUIS uses a 9-point bipolar scale, from “confusing” to “clear”. SUMI [3] contains 50 items, including a global scale of 25 items and five subscales of efficiency, affect, helpfulness, control and learnability with 10 items each. SUMI uses a three point scale of “agree”, “undecided” and “disagree”. CSUQ [5] has 19 items, with questions related to ease of use, learnability, clear and relevant information and user interface. It uses a 7-point scale, from “strongly disagree” to “strongly agree”. SUS [1] has 10 items and produces a general summative score between 0-100. It uses a 5-point scale, of “strongly disagree” to “strongly agree”. The SUS statements alternate between negative and positive, however a version including only positive statements also exists [6].

These questionnaires include statements such as: “Terminology relates well to the work you are doing” (QUIS) [2], “Tasks can be performed in a straight forward manner using this software” (SUMI) [3], “The information is effective in helping me complete the tasks and scenarios” (CSUQ) [5] and “I think that I would like to use this system frequently” (SUS) [1]. These questionnaires have been applied in the field of health informatics research (e.g. [7]) in a standard way as part of usability testing.

Standardized usability questionnaires have also been criticized for not taking into account contextual aspects of work when usability is evaluated from the end-users', for example physicians', perspective [8]. These include simultaneous use of numerous IT systems, variety of work practices, various areas of expertise and places of work. This has led to the development of a tailored usability questionnaire for clinical context [8,9]. This questionnaire is targeted to physicians and it includes 38 core statements with a five-point Likert-scale as well as an overall rating of the EHR system in use [9]. In addition to usability, there are questions related to background, problems, successful functionalities, management issues, patient safety, occupational health and information systems development.

Different questionnaires have been applied in healthcare procurement both in relation to demonstrations [10] as well as after test use of the system [10,12]. HIMSS suggests using SUS as a post-test questionnaire when selecting and procuring an EHR [4]. However, questionnaires addressing usability during demonstrations in procurement seem to have reduced it to a subjective evaluation of the user interface or evaluation of specific functionalities [10].

During our review of existing literature, we did not find standard questionnaires developed to evaluate usability based on seeing the system being used. The established usability questionnaires address usability from several perspectives, but most of them require experience in using the system. Therefore, our aim in this paper is to construct and present a questionnaire that can be used to evaluate the usability of several candidates in scenario-based demonstration sessions in a procurement project.

2. Usability Questionnaire for Demonstrations in Procurement (DPUQ)

The development of DPUQ included four phases: literature review, preliminary version of questionnaire, piloting and final version of questionnaire. The final statements in the questionnaire are presented in Tables 1 and 2 with references to similar items in existing questionnaires. The statements were rated on a four-point Likert scale of "strongly agree" (3) to "strongly disagree" (0).

Based on the procurement setting as well as the user-scenario based demonstrations, considerations for the development were identified already in the beginning: The end-users would be evaluating also functional scope and quality with other questionnaires during the demonstrations. There were several different user-scenarios, which included a variety of different clinical settings, and the questions needed to be suitable for all of them. Certain usability themes were deemed as not possible to evaluate, such as recovering from errors and terminology (as not all systems were in the native language of participants).

The piloting was a significant step in the development process. The preliminary version of DPUQ included two parts: eight questions for a demonstration break and a standard SUS questionnaire for the end.

The following observations emerged from the piloting: The alternating of items between positive and negative in the traditional SUS was seen as error-prone and confusing. The SUS items needed wording changes to reflect the demonstration situation and evaluating a complex work system. Items relating to learnability, confidence, and efficiency were not suitable because end-users did not use the system themselves. Items relating to necessity of demonstrated functionalities were not suitable because the scenario was pre-written and not the object of evaluation. The demonstrations would be long so a third set of questions was needed.

Table 1. Parts 1 and 2 of DPUQ. These parts are answered during different breaks in the demonstration.

Item	Statement	References
1.1	The arrangement of fields and functions on the system screen is logical.	[2,9,8]
1.2	The system presents clearly what it is doing (e.g. when saving or deleting data, or loading data).	[2,9]
1.3	The system behaves in a very anticipated and logical way (the system behaves as I expect it to).	[3,9]
1.4	The various parts of the system have a lot of consistency.	[6,3]
1.5	The system supports collaboration between various parties.	[8,9]
1.6	The system supports information exchange between various parties.	[8,9]
2.1	Entering and documenting data is easy and smooth.	[9]
2.2	The necessary information is easily available in the system.	[(9,5)]
2.3	The necessary information provided by the system can be effectively utilize.	[5],[8]
2.4	The system screens are visually pleasant.	[5]
2.5	The system functionalities are well implemented.	
2.6	The way the system works is easy to understand and intuitive.	

Table 2. Summative part of DPUQ. These statements are answered after the demonstration.

Item	Statement	References
3.1	I would like to use this system in my work.	[1,6,3]
3.2	The system is not too complex.	[1]
3.3	Based on what I saw, using the system is fluent.	[1]
3.4	The various functions in the system were well integrated.	[1,6]
3.5	The various parts of the system have a lot of consistency.	1.4
3.6	The necessary information is easily available in the system and can be effectively utilized.	2.2, 2.3
3.7	The system supports collaboration and information exchange between various parties.	1.5, 1.6
3.8	I would imagine that most people would learn to use this system quickly.	[1,6]
3.9	Based on what I saw, the system is intuitive (it is easy to guess how the system is used).	[6]
3.10	The system is very suitable for completing my everyday work.	[5], ([8])

The final version of DPUQ was set to include themes of traditional usability issues (consistency, logic, status, complexity and visual appearance of the system) and EHR-related usability issues (compatibility of system and clinical tasks, support for collaboration in clinical work). The final version included three parts; parts 1 and 2 to be used during breaks in the demonstration and the summative part at the end of demonstration. Based on observations from piloting, the first eight statements were reduced to six, reordered, and the wordings were modified. This became part 2 of the final version of DPUQ. SUS was modified by replacing three statements with new ones, using only positive and reordered statements. This became the summative part of DPUQ which was designed to somewhat overlap with the previous parts.

3. Using DPUQ and Initial Experience

DPUQ was applied in the CAPIS project. The questionnaire was originally targeted for healthcare context, but once finished was deemed appropriate also for social welfare. Each of the four vendors was evaluated on nine different user scenarios addressing typical professional workflows in healthcare and social welfare (details in Table 3). The user scenario demonstrations had different lengths, ranging from 2.25 hours to 6 hours. The number of respondents per demonstration ranged from 9 to 31. The total number of filled DPUQ forms from the demonstrations of all four vendors was 771.

For each demonstration, two scores were calculated from DPUQ. First, an arithmetic average was calculated of all answers to part 1 and part 2. Then, for the summative part of DPUQ an arithmetic average was calculated of the sums of the answers of each user. Thus, the score for parts 1 and 2 could be between 0-3, and the score for the summative part could be between 0-30.

Initial correlation analysis of DPUQ and another usability evaluation method (HED) demonstrated that the evaluation results for usability from both methods seem to be well aligned ($\text{corr}=0.79$, $p<0.001$, $\text{df}=22$, $t=6.097$) [13]. Based on the analysis in [13] the scores given by end-users seemed to be slightly higher than those given by usability experts.

Table 3. User scenarios, their lengths and number of respondents.

Scenario	Length (hours)	Respondents per scenario *
<i>Healthcare</i>		
Emergency department (ED) – Intensive care unit (ICU) – Operating room (OR) **	6	22-23 (P,N)
Maternity clinic – Labor and delivery – Child health clinic	3	9-10 (P,N)
Inpatient ward	2.5	17 (P,N)
Ambulatory care	3	24-26 (P,N)
Home health and elderly services	2.25	11-12 (P,N,S)
E-services (for professional-patient communication)	2.25	16-17 (P,N,S)
<i>Social Welfare</i>		
Social assistance	2.5	29-31 (S)
Child welfare	2	29-31 (S)
Services for people with disabilities	2	29-31 (S)

* P= physicians, N=nurses, S=social workers

** Parts 1&2 of DPUQ were issued twice in this scenario because of the length of the scenario.

4. Discussion and Conclusions

In EHR procurement several system candidates should be evaluated thoroughly. Questionnaires have been recommended for including end-users' perception of system usability in procurement [4]. Additionally, using clinical information processing scenarios (CLIPS) are suggested for procurement [14]. Currently, no usability questionnaires enable evaluation without personal experience in using a system. DPUQ can be used to merge these recommendations when applying scenario-based demonstrations.

DPUQ produces numeric results for comparing system candidates. These results can be used to guide the selection process. However, based on our experience to get a more in-depth and accurate understanding of usability of system candidates DPUQ should be used concurrently with other evaluation methods, such as heuristic evaluation during demonstrations (HED) [13]. Future research will further analyze the validation of DPUQ also from psychometric point of view and the results of using it during procurement.

References

- [1] J. Brooke, SUS: a "quick and dirty" usability scale, In: Jordan P, Thomas B, Weerdmeester T, McClelland A (Eds.), *Usability Evaluation in Industry*, London: Taylor and Francis, 1996.
- [2] J.P. Chin, V.A. Diehl, K.L. Norman, Development of an instrument measuring user satisfaction of the human-computer interface, In: O'Hare JJ (Ed.), *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI '88)*, New York, NY: ACM; 1988: 213-218.
- [3] J. Kirakowski, The use of questionnaire methods for usability assessment, *Background Notes on the SUMI Questionnaire* [Internet], 1994, [cited 2016 Sep 20]. Available from: <http://sumi.ucc.ie/index.html>
- [4] HIMSS EHR Usability Task Force, Selecting an EHR for Your Practice: *Evaluating Usability*, [Internet], 2010, [cited 2016 Aug 16]. Available from:

- https://www.himss.org/sites/himssorg/files/HIMSSorg/Content/files/Code%20119_Selecting%20an%20EMR%20for%20your%20practice_Evaluating%20usability_HIMSS.pdf
- [5] J.R. Lewis, IBM Computer Usability Satisfaction Questionnaires, *Int J Hum Comput Interact* **7** (1995), 57-78.
- [6] Sauro J, Lewis JR, When Designing Usability Questionnaires, Does it Hurt to be Positive? In: *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, New York, NY: ACM; 2011: 2215-2224.
- [7] P.H. Lilholt, S. Heiden, O.K. Hejlesen, User satisfaction and experience with a telehealth system for the Danish TeleCare North trial: a think-aloud study, *Stud Health Technol Inform*, **205** (2014), 900-904.
- [8] J. Viitanen, H. Hyppönen, T. Lääveri, J. Vänskä, J. Reponen, I. Winblad, National questionnaire study on clinical ICT systems proofs: Physicians suffer from poor usability, *Int. J Med Inform* **80** (2011), 708-725.
- [9] J. Kaipio, T. Lääveri, H. Hyppönen, S. Vainiomäki, J. Reponen, A. Kushniruk, E. Borycki, J. Vänskä, Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland, *Int J Med Inform* [in press, accepted manuscript], [cited 2016 Nov 10]. Available from: <http://dx.doi.org/10.1016/j.ijmedinf.2016.10.010>
- [10] A. Holbrook, K. Keshavjee, S. Troyan, M. Pray, P.T. Ford, Applying Methodology to Electronic Medical Record Selection, *Int J Med Inf* **71** (2003), 43-50.
- [11] E. Liljegren, A.L. Osvalder, Cognitive Engineering Methods as Usability Evaluation Tools for Medical Equipment, *Int J Ind Ergon* **34** (2004), 49-62.
- [12] S. Jensen, S.L. Rasmussen, K.M. Lyng, Use of Clinical Simulation for Assessment in EHR-Procurement: Design of Method, *Stud Health Technol Inform* **192** (2013), 576-580.
- [13] M. Tyllinen, J. Kaipio, T. Lääveri, M. Nieminen, We Need Numbers! – Heuristic Evaluation during Demonstrations (HED) for Measuring Usability in IT System Procurement, *Proc. 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, New York, NY: ACM; 2016: 4129-4141.
- [14] A. Kushniruk, M-C. Beuscart-Zéphir, A. Grzes, E. Borycki, L. Watbled, J. Kannry, Increasing the Safety of Healthcare Information Systems through Improved Procurement: Toward a Framework for Selection of Safe Healthcare Systems, *Healthc Q* **13** (2010), 53-58.