Evaluation for Evaluation: Usability Work during Tendering Process

Kimmo Tarkkanen
University of Turku
Turku School of Economics
Rehtorinpellonkatu 3
Turku, 20500 Finland
kimmo.tarkkanen@utu.fi

Ville Harkke
University of Turku
Turku School of Economics
Rehtorinpellonkatu 3
Turku, 20500 Finland
ville.harkke@utu.fi

Abstract
In large-scale IT implementations the iterative system development and system tailoring to organization-specific needs begins after a tendering process is over. In the tendering process a procuring organization defines usability requirements, evaluation procedures and selection criteria, while IT vendors try and build and propose solutions that best fulfill these requirements. Existing usability research has instructed the procuring organizations on how usability should be included in the request for proposals, in order to ensure the usability of the selected system. However, the IT vendor perspective on request for proposals and related usability work during the tendering process remain unexplored. This paper grasps the issue with an empirical study on large-scale IT procurement. The preliminary findings give practical advice to vendors planning usability work and buyers writing request for proposals.

Author Keywords
Usability evaluation; usability requirements; request for proposals; tendering; procurement

ACM Classification Keywords
H.5.m. Miscellaneous

Introduction
Request-for-proposals (RFP) are documents that start a tendering process, a compulsory way to procure large governmental IT systems in the European Union, and also a common practice in private companies to arrange a competition between possible IT system vendors. RFP defines an initial set of system requirements and a selection criterion for the systems entering in the competition. Usability, as part of non-functional system requirements, may be included in both sections: RFP defines how usable the system must be, and, how much weight usability will have in the decision. In addition, it defines which usability evaluation methods, measurement units and specific levels are used [2]. Three distinct evaluation methods are commonly used by procuring organizations: expert-based, user-based or demonstration-based evaluations. The value of realistic usability testing is emphasized [5, 7].

To cover all aspects of usability in the RFP, a mixture of requirement styles is needed [8]. For example, the buyer may require vendors to use certain usability methods and number of iterations, or, define exactly how easily and quickly users should be able to accomplish a certain task. Performance measures of the latter kind can be verifiable, valid and comprehensive; qualities that requirements in RFPs should meet [6]. However, overstated or understated requirements may rule out acceptable systems and qualify poor ones [2]. Unless appropriate usability requirements are found, an option is not to include usability requirements or criteria in the RFP at all [6]. A real danger is that IT vendors spend resources to usability work only to the extent required by the RFP [6]. In that regard, RFP becomes a critical phase for governmental organizations in ensuring usability in IS development [6].

More research on usability practitioners' activities is suggested [3, 12], also concerning practices during IT procurement and RFP planning [6]. However, we recognize that since the 80's usability research has instructed mainly IT procurement in the inclusion of usability in RFPs [2, 6, 7, 8, 9], but lacks still focus on RFPs and subsequent usability actions from the respondent, IT vendor, perspective. In the literature, vendors have been considered almost as opponents in the process. This situation evolves into a partnership first after the tendering is over, when iterative system development and tailoring to organization-specific needs can begin and traditional usability activities take place. However, are usability fundamentals, methods and techniques during the tendering similar with traditional development? Does the tendering effect on usability activities of the vendor? This paper grasps the vendor's perspective on RFPs as well as usability practitioners' work during large-scale IT procurement.

Case description

Our findings are based on our professional involvement in the pre-evaluation testing procedure of a system, following loosely the tradition of action research (see [1]). We co-operated with an IT vendor in a tendering process of large-scale IT procurement. Our task as usability experts and researchers was to improve the usability of three systems that were soon to be demonstrated to the procurement organization and subsequently tested by the users of the organization. In practice, we pre-evaluated usability of the demonstration versions of proposed systems according to the RFP provided by the procuring organization. Two of the systems were usability tested by prospective users (3+4 pers.) according to the RFP. Both systems were also expert reviewed by the authors. The third system was only expert reviewed during a demonstration.

The RFP defined relatively high and significant weight for usability on the purchasing decision. Usability requirements, the value of each requirement and methods for evaluating the fulfillment were carefully determined. Methods to be applied were expert reviews, demonstrations and usability tests. Testing procedures were also described in detail, including usability test tasks (as [5] suggests), which then allowed us to simulate the forthcoming evaluation by the procuring organization in advance. The usability measures were derived from definitions, yet described only conceptually, thus missing the units of measurement. For the competition and system comparison, the evaluation results and the overall system usability, were to be quantified to scores given to each system.

Results

Some interesting implications for both the procurement organization and the IT vendors were noted during the pre-evaluation process. These preliminary findings are represented in bulleted points below and discussed with examples after each node. (Note: the examples are fictive due to the confidentiality of the case).

- *Due to unspecified measures of usability in the RFP, we had to think like the usability evaluators of the procuring organization and please them, not improve the system usability as such.*

The first question when the project started was how usability will be measured by the procuring organization. At that time, the materials did not provide enough details for the vendor. For example, we realize that the execution time i.e. a commonly used measure for efficiency in summative tests was not appropriate due to inability to compare the result with other systems in the competition. Thus, the next big question was how the usability evaluators of the procuring organization would think. In fact, while this question solved the problem of unknown measures to some extent, it became the basis for all our usability work during the tendering process. In other words, pleasing the usability evaluators of the procuring organization would be the main goal of the usability evaluators of the vendor. Subsequent questions, like how to make usability look good in the eyes of the evaluator, and, what interactions they will like or dislike, followed us throughout the process. A well-known evaluator effect [4] is an evidence of the power of the evaluators of the procuring organization: Evaluators will guide sessions, convert test users' comments and usability observations to quantifiable scores to the vendor.

- *A tight tendering schedule defined in the RFP forced us to apply only one usability evaluation and development iteration, further leading to only minor changes to the system in a disadvantage to the buyer.*

The RFP defined tight deadlines for system demonstration and testing that hardly gave time for the vendor to implement a standard system that would fulfill all of the requirements. As a consequence, usability evaluation and improvements had to take place at the last minute allowing only minor usability changes at the system surface level (see [10]). The time to build demonstration versions was so tight that resources were mostly spent on configuring, assembling and coding activities despite the will to pay attention to usability at the same time. That resulted in unpolished UI transitions, complicated navigation paths and eventually poor efficiency.

For example, the pre-testing users, while trying to follow the tasks in the RFP, confronted many error dialogues and –pages due to the unimplemented features of the system. While such descriptive error dialogues are essential in actual use situations, these dialogues would have increased execution time and number of mouse clicks (i.e. resulting in poorer efficiency) in the actual test situation. In addition, users got an impression of an erroneous system leading to weaker satisfaction. One possible solution was to leave functions, links and buttons to inoperative states instead of presenting dialogues.

- *Usability test tasks and user stories described in the RFP were strictly sequenced (i.e. task A must be executed before B or within the task the phases were detailed), which was not always an optimal sequencing for the system logic, which, on the other hand, could have improved the sequencing of work tasks for the implementing organization.*

The user stories for the demonstration were not always very process-oriented. Those aimed at screening of system interfaces and different interactions. That resulted in unnecessary episodes and transitions between the actual tasks to be evaluated. For example, think a demonstration task (or task in a usability test as well) that requests to 'send a message to a colleague', while the subsequent task demands to 'show how the message appears for the colleague', after which the demonstration should continue from the original situation of the sender. Clearly, this kind of behavior is not very realistic, but requested to verify system behavior and appearance. However, in the vendors' interest is to avoid excess transitions, dialogues, interactions, mouse clicks etc., in order to maintain absolute efficiency and subjective expectations of the demonstration audience. Therefore, we ended up in identifying unnecessary transitions and replacing those with shortcuts to improve usability evaluation results during the demonstration. Unfortunately, the result may not have the best logic, for actual tool use, and may lead to an inferior user experience.

- *Despite carefully described test tasks, we had multiple interpretations of what is requested, and whether the system needs to be operable exactly as written.*

For example, a test scenario first describes how 'Smith and Johnson have written and published an article together'. A task related to the scenario is then to 'document this achievement to the library system of the university'. Should this documentation be done for both authors at one time, or only for 'the user', or twice for both separately, knowing that the system supports all options? Furthermore, we had difficulties in interpreting who should execute the subsequent task 'the other author finds a mistake in spelling his name, please correct it', because the system allows both to do that and because it was ambiguous whether 'finding' or 'correcting' is in the essence of the task evaluation.

On the other hand, some test tasks were not executable with the demonstration system exactly as written, which made it difficult for prospective users and pre-evaluators to rate whether a task is completed or not. For example, user's task is to find and send a message to a person called 'John Smith', which however cannot be found from the system but only 'John Smith Jr.', to whom the message is sent fine. Users will have doubts, take longer with the task or they may even halt the task execution, if test tasks in RFP are formulated emphasizing too many details. As a result a good system for searching for people and sending messages may get poor results for usability. For the vendors, the implication is that valid background data is critical for success of usability evaluation. In general, communication gaps between the buyer and vendors, including the authors, during the tendering made it difficult to design usability as desired by the procuring organization. This leads to our next point:

Practical conclusions

Vendor

• Try and think like the usability evaluators of the procuring organization.

• You may not have the time to improve the system usability as such. Your primary usability goal is what is asked for in the RFP.

• Interpret requirements and usability test tasks in the RFP to your advantage.

• Consider options to fixing a system which is highly usable in other task sequence than requested.

• Do not deviate from the required demonstration path, present superior features elsewhere.

Buyer

• Keep in mind that what You specify is what You get; the RFP and time to develop have a critical effect on the offers.

• Do not tie your hands by setting too detailed test tasks; the vendors are the experts in system design and should come with solutions.

- *The rigid structure of the tendering and usability testing process defined in the RFP effectively hindered the vendor to fix the deficiencies of the test tasks and improve the underlying assumptions of the 'best practices' of the procuring organization.*

We noted that usability test tasks defined in RFP will have a very normative effect on the outcomes of the tendering, as described above. The deep structures of the proposed system (see [10]) may not support the prescribed test tasks, but questions arose whether the tasks were truly 'real and meaningful' for the users. How should the vendor respond when a task is irrelevant since the proposed system uses another type of logic, which could be perhaps more effective, efficient and satisfactory for the organization? Undoubtedly, every procuring organization aims to create usability requirements as well as descriptions of work processes that are as reliable and valid for them as possible. Yet considering these descriptions as the only reality and sticking to these with rigid evaluation procedures in the RFP may not be advantageous. As the test scenarios are built on a set of assumptions, the vendor could challenge these, but the rigid structure of the tendering and testing process effectively hinders this. In our case, the proposed solution could have been easily tailored (in its deep structures) to fit the requested tasks as well as other types of more complex user tasks, but this could not be demonstrated due to time and resource limitations and process restrictions. Efficiency of the system, as measured according to the RFP, would have deteriorated significantly if the system demonstration deviated from the requested path. Thus, RFP planners should pay attention to how these advanced features and functions, or improved sequencing of work tasks, can be effectively communicated to the procuring organization and whether these should be valued in the procurement. Alternative [8] and supplementary [2] requirements would benefit both parties.

Conclusions

Preliminary case findings suggest that usability evaluation and improvements during a tendering process differs from a traditional user-centered design cycle. Usability work on the tendering is aimed at winning the competition i.e. focused on acquiring better scores from usability, where "the customer, not the end user, is the target [9]". At the system level, that may mean usability decisions that are temporary or fake – just to make an impact. Systems built to match overstated requirements and procedures will indeed score high in usability evaluation, but as products on the markets, or tools at work those systems may not have acceptable levels of usability. Procuring decisions, materialized in RFPs, will have effect on vendors' emphasis and approach on usability issues. Based on the case, the vendor's usability work was most influenced by the buyer's decisions like:

- How well the buyer understands users, uses and needs (i.e. how detailed usability requirements and rigid procedures should be written to the RFP)?
- How much the buyer gives time to vendors to respond to usability requirements (i.e. is the competition arranged for systems that are *as-is*, *to-be*, or *fake-to-make impact)*?

Usability work during a tendering process seems understudied in usability literature. Research from IT vendors' point of view could improve not only their own skills as "chefs cooking culinary meals", but also provide ingredients and

resources (see [11]) for procuring organizations to help and assess the work of vendors. The expertise of the vendors could be used in advantage for the buyer when aimed correctly.

References

[1] Baskerville, Richard L. Investigating information systems with action research. *Communications of the AIS* 2.3es (1999) 4.

[2] Carey, T. A usability requirements model for procurement life cycles. In Carey, J. M. (Ed.) *Human factors in information systems: An organizational perspective*, 2, Intellect Books (1991), 89-104.

[3] Følstad, A., Law, E., and Hornbæk, K. Analysis in practical usability evaluation: a survey study. *Proc. CHI 2012, ACM Press (2012),* 2127-2136.

[4] Hertzum, M., Molich, R., and Jacobsen, N.E. What you get is what you see: revisiting the evaluator effect in usability tests. *Behaviour & Information Technology*, *33*, 2 (2014), 144-162.

[5] Jensen S., Rasmussen S.L., and Lyng K.M. Evaluation of a clinical simulation-based assessment method for EHR-platforms. S*tudies in Health Technology and Informatics 205* (2014), 925-929.

[6] Jokela, T., Laine, J., and Nieminen, M. Usability in RFP's: The Current Practice and Outline for the Future. *Human-Computer Interaction. Applications and Services*. Springer Berlin Heidelberg (2013) 101-106.

[7] Kushniruk, A., Beuscart-Zéphir, M.C., Grzes, A., Borycki, E., Watbled, L.,and Kannry, J. Increasing the safety of healthcare information systems through improved procurement: toward a framework for selection of safe healthcare systems. *Healthcare quarterly (Toronto, Ont.)*, *13 (2010)*, 53-58.

[8] Lauesen, S. Usability requirements in a tender process. *Proc. Computer Human Interaction 1998.* IEEE (1998), 114-121.

[9] Thorén, C. Approaches for inclusion of usability and accessibility in ICT procurements. *Proc. UITQ 2005*. (2005), 39-42.

[10] Wand, Y., and Weber, R. An ontological model of an information system. *Software Engineering, IEEE Transactions on*, *16*, 11 (1990), 1282-1292.

[11] Woolrych, A., Hornbæk, K., Frøkjær, E., and Cockton, G. Ingredients and meals rather than recipes: a proposal for research that does not treat usability evaluation methods as indivisible wholes. *International Journal of Human-Computer Interaction*, *27*, 10 (2011), 940-970.