Spring 6-15-2016

OBSTACLES AND WORKAROUNDS IN USABILITY PRACTICES DURING TENDERING FROM IS VENDOR'S PERSPECTIVE

Kimmo Tarkkanen
University of Turku, kimmo.tarkkanen@utu.fi

Ville Harkke
Åbo Akademi University, Ville.Harkke@abo.fi

OBSTACLES AND WORKAROUNDS IN USABILITY PRACTICES DURING TENDERING FROM IS VENDOR'S PERSPECTIVE

Research

Tarkkanen, Kimmo, University of Turku, Turku, Finland, kimmo.tarkkanen@utu.fi

Harkke, Ville, Åbo Akademi, Turku, Finland, ville.harkke@abo.fi

Abstract

When organizations purchase information systems from IT vendors, they produce formal calls for tender (CFT) that specify, among other attributes, the required usability criteria of the systems. HCI research has concentrated on the IS purchaser organizations' view and studied usability practices and criteria embedded into CFTs. However, IS vendors' perspective is much less studied and especially the effects of characteristics of CFTs on IS vendors' usability practices remain unexplored. We participated in a large tendering and experienced how the CFT introduced obstacles in the vendor's usability work that needed workarounds, which may decrease system usability in context after the tendering and falsify the evaluation results of the purchaser. We identified obstacles arising from the tendering context, evaluation practices and usability measures. The joint effect of detailed usability test tasks and user performance measures published as well as a limited communication and time during the process were experienced as the most problematic and counterproductive in vendor's usability work.

Keywords: Usability evaluation, Usability requirements, Call for tender, Tendering.

1 Introduction

Public information system purchases are subject to tender, for example in the European Union and many other countries, if a certain monetary limit is exceeded. A call for tender (CFT), invitation for tender, request for tender or request for proposal is a document that defines a set of usability requirements and a selection criterion for the systems partaking in public or private tendering of information systems (IS). The submitted tenders are evaluated according to the defined criteria and the winner is the one who best meets all the requirements outlined in the CFT. Creating CFT is a critical phase for governmental organizations in ensuring that usability is taken into account during the IS development (Jokela et al., 2013).

HCI research has concentrated on the IS purchaser's view and studied mostly the creation of CFT and how to embed usability criteria in it in order to ensure the usability of the selected system (cf. Carey, 1991; Lauesen, 1998; Thorén, 2005; Lehtonen et al., 2010; Jokela et al., 2013; Jensen et al., 2013; Nieminen et al., 2014). While there still is a need for more research on usability practitioners' real activities (Følstad et al., 2012; Woolrych et al., 2011), the IS vendors' perspective on CFTs and their usability practices during the tendering process remain unexplored in the HCI literature. Vendors' perspective is highly relevant, because they have to act on CFTs, i.e. they need to create an offer and build an information system that satisfies purchaser's needs. A danger is that vendors spend resources to usability work only to the extent required by the CFT (Jokela et al., 2013). Thus, the quality of the CFT is measured in the IS vendors' practices, and by exploring these, we can further improve the creation of the CFT by the purchasing organizations. As tendering is a mainstream activity with large and complex information systems purchases in public and private sectors, the issue concerns not only the

HCI discipline but also the whole community of information systems research (see Hochstetter et al., 2012).

In this research, the question is how the usability work of IS vendor takes place during the tendering process and how the tendering context effects this work. To explore this issue, we participated in a large tendering case as usability evaluators and researchers at the IS vendor's site. We reflect our collective processes and experiences solving the practical usability problems of a complex information system prepared for the tendering. In the results' analysis of the case study, we identify characteristics of the CFT that emerged as obstacles in vendor's usability work. Therefore we refer also to the concept of workaround, which presupposes that there is some sort of obstacle in achieving the goal of work and that must be circumvented, and hereby the "standard operating procedure" is bypassed or overridden, i.e. the work is carried out some other way than usually (see e.g. Alter, 2014). Although workarounds are often associated with the usage of computer systems (Collins English Dictionary, 2009), there are many kinds of applications of the concept of workarounds (cf. Tushnet, 2009). As the result section describes, usability workarounds are about bypassing human-centered design principles and usability guidelines and carrying out usability work some other way than usually due to the confronted obstacles. Perhaps, in the engineering discipline these could be called "engineering shortcuts", which reach a goal in a way that leaves the system, as a whole, compromised. In general, workarounds are under investigated and undertheorized especially from the viewpoint of their origin and outcomes (Pollock, 2005). To our knowledge, this holds with the usability engineering practice as well. These usability workarounds are not always very self-satisfactory, positive or even ethical, yet it is worth to discuss their origins and possible ways to eliminate them in IS tendering.

2 Usability in Call for Tender

Procurement policies manifest themselves in a call for tender (CFT) published at the beginning of the bidding. These documents also define the tendering process, compulsory in procurement of large governmental IT systems in the European Union and a common practice in private companies, to arrange a competition between possible IT system vendors. In general, the CFT defines an initial set of system requirements and a selection criterion for the systems entering in the competition. The idea is to ensure open and corruption-free public procurement and to allow every possible vendor to participate in the tendering. As usability can dictate the success or failure of enterprise-wide information systems (Yu, 2005), it should naturally be included in CFTs. Depending on the size of the procurement, the CFT may include statements about (Carey, 1991): How usable the system must be? What are the measurement units and the specific levels at minimum? How is usability evaluated and measured? What and how are evaluation methods applied? How must usability be designed and evaluated by the vendor? What evaluation methods and how many design iterations are required? How much weight aforementioned criteria have in the selection?

To cover all aspects of usability in the CFT, a mixture of requirement styles is needed (Carey, 1991). For example, the buyer may require vendors to use certain usability methods and number of evaluation iterations (i.e. a process-based requirement), or, define exactly how easily and quickly end-users should be able to accomplish a certain task (i.e. a performance-based requirement). In the mid-nineties, very few CFTs specified HCI issues or usability activities beyond general descriptions such as 'the system must be easy-to-use' (Winkler and Buie, 1994). In their empirical study, Lehtonen et al. (2010) found that CFTs involve four types of usability requirements: 1) General usability requirements ("must be easy to use"), 2) Usability process requirements ("must contain three evaluation iterations"), 3) Usability design requirements ("input format must be appropriate to the task") and 4) User performance requirements ("user must accomplish this in 10 seconds").

According to Jokela et al. (2013), only user performance measures can be verifiable, valid and comprehensive, and thus, although difficult to formulate, the most favorable usability requirements for CFTs. A performance measure is, for example, how easily and quickly the user should be able to ac-

complish a certain task. Performance can be measured in terms of effectiveness, efficiency and satisfaction i.e. the elements of the definition of usability (ISO, 1998). Effectiveness is often measured as the task completion rate, efficiency as the time spent on the execution, the number of interaction steps, mouse clicks and errors confronted, and satisfaction as subjective ratings of end-users. Despite well-defined measurements, overstating and understating requirements may rule out acceptable systems and qualify poor ones in the tendering (Carey, 1991). Systems may be built for overstated requirements and procedures, which score high in usability evaluation and competition, yet which as products at the markets, or tools at work, would not acquire acceptable levels of usability. On the contrary, understated usability requirements may lead to difficulties in comparison and selection process and danger the "openness" of the procurement. A recent study by Billestrup et al. (2015) questions the most formal requirements but also requirements that leave too much room for the vendors to innovate. Unless the purchaser finds appropriate usability requirements, an option is not to include usability requirements or criteria in the CFT at all (Jokela et al., 2013). One reason for this is that IS vendors may spend resources to usability work only to the extent required by the CFT (Jokela et al., 2013).

There are three types of evaluation methods commonly used by procuring organizations: expert-based, user-based or demonstration-based evaluations. These include techniques familiar from design context, for example, walkthroughs and heuristic analyses, usability tests with high-fidelity systems, and organizing and observing demonstration sessions of system use (Kushniruk et al., 2010; Kushniruk and Patel, 2004). The value of realistic usability testing in the real use context is emphasized (Jensen et al., 2013; Kushniruk et al., 2010). However, even the most realistic testing cannot overcome contradictions between usability practices and public procurement policies (Thorén, 2005). First, usability occurs during real use i.e. after the tendering is over (Thorén, 2005 p. 44). This imposes a problem for the purchaser: How to evaluate usability (objectively) during the procurement? The problem is inherent in any formative usability evaluation, because generalizing the results of an artefact test to a tool in use in the real context is not straight forward (Reijonen and Tarkkanen, 2015). Second, usability evaluations are subjective, even if the purchaser defines and publishes the qualities assessed in advance. The problem arises from the ever present phenomenon of the evaluator effect (Hertzum et al., 2014), which in the tendering context, means that the procurement may not be predictable by the vendors and that vendors may not be treated equally. For example, one can consider the great variety of results in CUE-studies despite jointly given goals and a client scenario that define most important tasks (Molich and Dumas, 2008). In order to have fair competition between the products and equal treatment of the vendors, Jensen et al. (2013) suggest describing the test scenarios and tasks of usability testing thoroughly and in detail. Users should not explore the systems as they like, but they have to imagine how the system supports their work while following the pre-defined task scenarios (Jensen et al., 2013). On the other hand, Kushniruk et al. (2010) express a concern that if the potential vendor gets a prearranged set of questions in advance, they may modify the demonstration system to appear to contain the desired functionality. Usability test tasks in the CFT can cover the most common (Jensen et al., 2013), the most frequent (Jokela, 2010), the most vital (Jensen et al., 2013) and the most common new tasks planned with the new system (Riihiaho et al., 2015). The definition and validation of these task scenarios are as important as the actual performance of the tests (Jensen et al., 2013).

The tendering processes and the subsequent development cycles are not uniform. Large-scale tendering may contain several complementary and interrelated information systems that either are standard systems on the market or developed from the very beginning for the organization. The iterative system development, i.e. configuring and tailoring, to organization-specific needs typically begins after the tendering process is over. In contrast, a small-scale tendering is targeted at standard off-the-shelf systems that are modified and configured only little for the implementing organization. In the tendering situation, there are weak possibilities for re-design and iterative evaluations of the HCI compared to standard usability engineering lifecycle (Beuscart-Zephir et al., 2002). However, vendors usually have to tailor their systems for the tendering purposes, even if systems are standard software packages. For example, the purchaser expects that systems will be supporting the scenarios used in usability tests and

demonstrations (see above), which are created from the premises of the purchaser with very limited access to candidate systems (cf. Kushniruk et al., 2009). From the vendors' point of view, we can recognize that the possible changes to existing and standardized production systems are somewhat limited. In order to meet the usability requirements of the CFT, vendors can deploy usability experts to evaluate the system against the requirements and introduce usability improvements that meet the requirements. This resembles the situation we discuss next in our case study.

3 Research Method

Methodologically the study is part of case study research tradition (Yin, 2003) although we as researchers intervened in actual design work and introduced our contribution to the setting as in action research (Baskerville, 1999). Case study as a research strategy allows an empirical inquiry on vendors' practices within real-life tendering context and investigating the currently unclear relation and boundaries between CFTs and vendors' usability work (cf. Yin, 2003 p. 13). We are interested in the causes and the consequences and in understanding how usability practices take place in this context (i.e. emergence of usability workarounds and the obstacles behind). These kind of questions fit well the case study research method (Eriksson and Kovalainen, 2008 pp. 39-40). In addition, the uncontrollable setting of tendering gives a distinct advantage to case study over experimental methods (Yin, 2003 p. 9). We collected empirical data from a large-scale IS procurement, in which we took part in the usability design and evaluation practices of the IS vendor. Our practical task as usability experts and researchers was to improve the usability of the candidate systems before the procuring organization began its own usability evaluation. The purchaser published a call for tender, which involved detailed descriptions of evaluation practices and formalized user performance measures. This allowed us to copy the forthcoming evaluation practices of the purchaser in detail and conduct our own evaluations based on these before the delivery of the system. Then, based on the results of the evaluation, the vendor was able to prepare the current version of the system for the purchaser's requirements. Table 1 represents the overall process where our responsibilities were on phase 2. The different usability measures had different weights and these were quantified to numeric values and total points given to a system in competition. Overall, usability had high significance for the purchaser decision.

1 The IS purchaser determines and represents formal usability criteria in a CFT based on the work of users and the context of use.
2 Potential IS vendor performs a usability evaluation for a candidate system based on the usability criteria and evaluation procedures represented in and copied from the CFT.
3 The vendor redesigns the candidate system based on the usability recommendations of the evaluations.
4 The purchaser performs usability evaluations for the redesigned candidate system of the IS vendor (among other candidate systems), in order to quantify usability and find the best system.

Table 1. The overall setting of the tendering and the phase of usability evaluations in the case.

The system we pre-evaluated consists of three different parts or modules. The system and its modules are each large, off-the-shelf type of products, implemented and sold usually by third party consultants and used by various kinds of organizations in a global scale, yet configurable to specific needs of each customer including the procuring organization of the case. Thus, the system is similar to any configurable ERP software at the market. When these systems are large and complex enough, there is a need for tailoring the system: First for the tendering and secondly for the actual use. That was also a case here as all the tested modules required some configuration to meet the specific requirements of the CFT. With more complex and large systems, the tendering process also grows in complexity, time and number of iterations of system evaluations. Our research took place during the second round of the tendering process, where the purchaser had made the initial selection of systems and vendors i.e. shortlisted these from the open call. During a period of two months, we participated in one of the tendering

phases, which preceded usability evaluations, expert reviews and laboratory-based usability tests, by the purchasing organization. We must emphasize that usability evaluations were not the only criteria in which the purchaser relied in their selection of the best system. There were many other different points of evaluation and comparison that are not discussed here.

Table 2 shows usability evaluations and data collected in this study (performed in order of appearance). First, the module 1 was evaluated by three usability experts acting in a role of prospective user (this was possible) and following the usability test tasks described in the CFT. After collecting individual observations, we performed data analysis together and invented ways to improve the system usability. A report was delivered over to the vendor that contained problem id, test task id, problem description and recommendation to fix it. The procedure of the second expert review of the module was similar, yet supplemented with Nielsen's (1993) heuristics and severity ratings in the report. The module 2 was first expert reviewed and then tested with four potential end-users in one-by-one test sessions each lasting about two hours. The primary content of the report was a list of usability problems and recommendations as with module 1. Last, the module 3 was expert reviewed during a 2 hours' usage demonstration given by the developers during which we identified and fixed on the fly about 10 problems. The first expert review of the module 1, usability tests of the module 2 and the review session of the module 3 were audio and video recorded.

	Expert reviews	Usability test
Module 1	3 experts, test task walkthrough, 32 problems 2 experts, test task walkthrough and a heuristic evaluation, 30 problems	---
Module 2	1 expert, 11 problems reported	4 users, 1,5-2 hours sessions, SUS results and 59 problems reported
Module 3	2 experts, expert review during 2h usage demonstration according to tasks, 10 problems fixed	---

Table 2. Usability evaluation methods applied to system modules and collected data.

The data collected and analyzed for this study contains material published in the CFT as well as 142 empirical observations and recommendations for redesigning the system modules. In the data analysis, the corresponding author revisited each problem-recommendation issue considering its nature, purpose and evidence. Those issues were selected for further analysis that were experienced as usability workarounds i.e. recommendations that somehow bypassed good HCI design principles (see definition of workarounds above). For example, a recommendation that did not strive for minimalist design (cf. Nielsen, 1993) but rather suggested information overload in order to shorten the navigation path and decrease time on task was selected. We identified workarounds in 61 (43%) of the reported issues, usability tests having slightly more workarounds (49%) than expert reviews (39%). We considered the reason why each recommendation exists and how it is justified from the vendor's (and ours) point of view. Thus, we aimed to identify an obstacle (e.g. decrease time on task, see above) that led to an atypical and "rule breaking" usability redesign recommendation. The obstacles then formed broader categories, namely the tendering context, usability evaluation practices and usability metrics as represented in the next section. It must be emphasized that the data presented in this paper is produced by us as usability evaluators observing prospective users and candidate systems, writing usability test reports of the sessions and retrospectively reflecting our own usability recommendations and practices in the situation. We have not interviewed IS vendor representatives for this study nor do these findings represent their opinions. We are also unknowledgeable about how (and how many of) found usability problems and recommendations represented here the system developers actually fixed although we know that the feedback from developers was positive in general. For example, we do not know whether the error messages were really eliminated or the features inactivated (see the results section later), but the point is that these *were possible* HCI actions and *represented the best usability decisions* at

hand in this tendering context. In this way, the paper highlights some possible usability practices of the IS vendors in relation to purchaser's decisions regarding the tendering and CFT.

4 Results: Vendor's Usability Workarounds and Their Origins in the CFT

Next, we present how the characteristics of the tendering in general and the call for tender document and its content in particular provide a fertile ground for usability workarounds in the vendor's site. We consider these characteristics as obstacles confronted by the usability evaluators and designers. These obstacles caused usability workarounds and some further problems as responses and solutions to obstacles.

4.1 Obstacles Related to Tendering Context

First, vendor's usability design was limited by the tight schedule and some restrictions to modify the system set by the tendering situation (Table 3, row 1). Therefore, we were hardly able to organize and perform the evaluation in time, so that there would be time also for design changes before system delivery. Consequently, we could plan and perform only one evaluation and design iteration, which did not end "when the solution meets requirements", but when the time was up. The tight schedule even gave rise to other workarounds in that only minor design improvements could be suggested and implemented. Moreover, the tight schedule led us to seek the most obvious, easiest and fastest solutions in order to meet the requirements, which further implied other types of usability workarounds discussed in next sections. Many of our redesign suggestions thus related to consistency issues in a traditional manner (e.g. the colors of visited links), but also creating consistency between usage patterns (see the next chapter e.g. how users can find task solution following already learned pattern).

Characteristics of the CFT	**Effect on usability work**	**Implications and rule breaking**
Tendering schedule is tight	Vendors can perform only one evaluation-design iteration. They concentrate only on minor usability improvements that can be implemented in time.	Usability work focuses on the visible appearance and layout, color and naming consistencies etc. instead of overall usability in the context, which denotes a narrow view on the usability concept.
Communication between parties is limited	Usability requirements and descriptions of use scenarios are subject to many interpretations. Vendors need to interpret requirements so that their system appears as the most efficient and effective way.	Usability decisions are based on assumptions instead of understanding the context. Conflicts may raise between parties after the tendering (e.g. law suits), if requirements are subject to many interpretations.

Table 3. Obstacles in usability work related to the tendering context.

Second, the limited communication with the purchaser during the tendering was an obstacle to performing usability work at vendor's site (Table 3, row 2). We had to interpret the requirements that were ambiguously presented to our system's advantage instead of asking details from the purchaser or their end-users. Many times it was hard to interpret when a certain task is (successfully) finished and who actually defines this and how. For example, we pondered and recommended[1], "*Much depends on how the evaluators interpret task completion. For example, do they evaluate the task as unsuccessful,*

[1] The words in square brackets are modified from the original wording and added here for helping readability and retaining anonymity.

if the [user] uses [Notes] field to write [appointment details] instead of using the [available options]? Is the decision of task completion given to the users themselves? If so, [users] can write all the [appointment details] wholly to the [Notes] field and consider the task as successful. Perhaps you could favor that behavior." Similar questions must have puzzled the purchaser also, because usability requirements, evaluation procedures and criteria were very detailed and rigid in the call for tender. We must emphasize that communication was not totally prohibited, but taken the tight schedule, the formality of the communication patterns and our position as external usability evaluators, answers would have been very difficult to get in time. Moreover, the answers may depend on system implementations as well, which purchasers are not familiar with. Our solution was to use the interpretation where our system was at its most effective and efficient implementation (see the quote above). However, if the requirements are subject to many interpretations, this may raise (even legal) conflicts after the tendering between the competing parties and with the buyer.

4.2 Obstacles Related to Usability Evaluation Practices

The call for tender published use cases and scenarios as well as usability test tasks that formed part of functional and non-functional requirements that the system was to meet (Table 4, row 1). The obstacle here is the type of the requirements itself, because after all, the goal of vendors' usability work is to fulfill the requirements and to win the tendering. Especially, publishing usability test tasks in advance and without any randomized sampling of the tasks, enabled us to recommend redesigning the system to correspond with the test tasks. This is a harmful workaround, because a test task design should not give clues and hints to users how a certain task is performed, while in this case these hints could be implemented straight to the user interface of the system. For example, we could change the name of the function to correspond with the name given in the test task description, which obviously helps the user to navigate to that specific function. On the other hand, we were also forced to use these workarounds in some cases, because the test tasks were created without any understanding about the system(s), resulting in requirements very far from the current features of the system(s). With simple information systems that embed few well-definable functions and for which the use context can be thoroughly simulated, planning test tasks without knowing the system seems feasible. Similarly, when one uses test tasks that define only high-level goals of users' work tasks such predefinition of test tasks seems appropriate.

For example, end-users needed to create the meetings and appointments of different types (consider e.g. board meeting, financial status update meeting, customer meeting, lunch meeting with colleagues) that the system was to support. The tested system provided automatic guidance and default values in creating the appointment details based on the type of the appointment selected by the user. However, users were confused by the defaults and automation in the background, not only because they were novices with this kind of a feature, but also because they were asked in the test task to select the appointment time and invite certain people to join that were partly automated by the system. In these circumstances, our proposal was that automated selections and appointment defaults need to be "switched off" and users need to have more control although we could see the potential in the efficiency of automation. Regarding the use of different appointment types, we suggested renaming "*appointment types with terms related to test tasks"* and to "*formulate appointment types either for every task or use [types] general enough to make the selection easy."* As discussed above, this means that system functions could match test tasks to the extent that remarkable help for users' performance could be provided. Wordings used in test tasks affect also another way round as "*test tasks 5-7 are formulated with the word [create] that encourages using the [Create function]."* All our test users approached the solution this way. However, this particular function did not support the exact goal of the tasks, which then needed renaming: *"Users should be guided not to use [Create function]. It should be used only in the last task 15 (if they ever will arrive that far). Therefore, we suggest at least renaming [Create] to [Create Board Meeting] or similar that clearly indicates the purpose and keeps users away from it."* With these usability decisions, the system apparently became more usable for the

test purposes, but there is no way of knowing whether these usability decisions are improvements or impairments after the tendering and several weeks of usage.

Characteristics of the CFT	Effect on usability work	Implications and rule breaking
Usability test tasks are published	Vendors have a possibility to design the system exactly match with the test tasks. They can follow the terminology etc. to improve system's usability and users' performance in these tasks.	Formulations of test tasks become guides for users how tasks are accomplished, which improves the performance drastically. Typically, test task design avoids giving clues to task completion.
Usability test tasks and use cases are low-level and strictly sequenced	Vendors need to design their systems to support the requested use cases and task sequences in detail.	Purchaser creates test tasks without familiarizing themselves with the systems in competition. This is not a common practice, yet applicable when systems are simple and tasks are high level. However, with low-level tasks, the system may be much more efficient and effective in another task sequence than requested. The buyer cannot assess whether other ways of accomplishment exists and whether these are even better than their current requirements and understanding of work. System usability becomes compromised.
Usability test tasks have interconnections and individual elements that recur within a task between tasks.	Vendors can analyze which test tasks are the most critical for other tasks to be completed efficiently and put design effort only on the critical ones. Usability design becomes influenced by the first test tasks: Users' learning curve needs to be maximized in the very beginning and the design needs to follow the created conventions in the rest of the tasks.	Usability design effort is misdirected from criticality in the real work to the relation of test tasks. These two rarely match when we consider the variation in tasks and human performances in real use.

Table 4. Obstacles in usability work related to the evaluation practices of the CFT.

The level of abstraction of the use cases and test tasks was rather low and included many details (Table 4, row 2). These requirements were hard to fulfil in a fairly short preparation time. In addition to the amount of details in usability evaluation tasks, these were to be performed in a numbered sequence. No workarounds could be used here, but the vendor had to resign and change the system to meet the buyer's requirements, which meant spending a lot of resources and scarce time. What usability practitioners at the vendor site would like to do in this situation is to show how much a certain use case or flow of work could be improved by innovation and innovative usability design (cf. Billestrup et al., 2015). For example, we noticed that the system could be much more effective, efficient and possibly also satisfactory in a different sequence of tasks than that requested by the purchaser. These possible innovations for the buyer's work and for their end-users are however dismissed with very formal usability criteria.

Moreover, due to the strict sequencing of test tasks those had interconnections and elements that recurred from task to task. This enabled us to analyze and determine which tasks are the most critical for other tasks to be completed efficiently (Table 4, row 3). Instead of putting the design effort where it should have been put - on the most critical tasks in the real life - this usability workaround misdirected the system design towards work tasks that are critical due to their exact place in the test. For example, *"The turn from task 7 to task 8 may induce several paths and navigation problems especially if [Appointment Details] field is not noticed by the user on the task 4."* Overall usability design becomes then influenced by the first test tasks: Users' learning curve needs to be maximized in the very begin-

ning and the design needs to follow the created conventions in the rest of the tasks. Task 1 needs to teach users the features and use practices that they need in the task 2 and so forth. For example, users may need to use a context menu in task 1, which they will become familiar with and search for solution also in other tasks. Therefore, it is the most efficient and effective to implement features, function use and navigation options etc. under as few techniques as possible and apply a tutorial kind of approach. These are indeed favoured in usability design in general. However, with complex systems that implement thousands of features that type of simplicity is not possible: The context menu would soon become useless if it contained access to all system features. Yet, in the context of tendering evaluation, training the user task-by-task and embedding as few different use practices as possible appears as the best option for both task completion and time and errors.

4.3 Obstacles Related to Usability Measures

While the above characteristics of the call for tender were kinds of requisites for applying user performance measures in tendering, also the measures themselves and their operationalization had an influence on applying usability workarounds. Users' performances are measured in terms of effectiveness, efficiency and satisfaction i.e. the classical elements of the definition of usability (ISO, 1998), in this case, as the task completion rate limited with time, the number of errors and performance steps and user satisfaction questionnaires. These variables had nearly equal weights in the overall system selection. However, the task completion rate was actually regarded as the most important goal in the vendor's usability design. If the users fail in the task, the efficiency and satisfaction scores will deteriorate significantly (in this particular case those were not scored at all), and vice versa, successful tasks are error-free and spend less time on task (Sauro and Lewis, 2009). Low SUS-questionnaire scores from the four test participants confirmed the need for designing for task completion first.

Characteristics of the CFT	**Effect on usability work**	**Implications and rule breaking**
Effectiveness is measured as task completion rate	Vendors need to overstate the relative weight of the task completion, because uncompleted tasks imply low scores also in efficiency and satisfaction measures in the tendering context. In this particular case, we were able to rename system elements (e.g. menu items, function names) and modify information appearance (dropdown lists) to correspond with the test tasks. We could also recommend certain elements and frames to be visible/open by default to help users in task completion.	Renamed features, modified data orders and visibility settings according to test tasks may serve only these, yet be inappropriate in other work tasks.
Efficiency is measured as execution time, number of errors and steps	Vendors can add shortcuts to shorten the navigation path despite of users getting lost in the system. This is applicable when users should have learned already where the next task begins. Another possibility is to inactivate features that are on the unrequested navigation path.	Users should be in control and informed about the system state and place (e.g. Nielsen (1993) heuristics).
User satisfaction is measured	Avoid showing error messages (e.g. by inactivation) not to give a feeling of erroneous system	In complex systems, error dialogs are inherent part of the systems.

Table 5. Obstacles in usability work related to the usability measures of the CFT.

Therefore, in addition to changes in terminology discussed above, changes in data appearance in different views, hierarchies and drop down lists could be considered, as well as setting the most critical views regarding the task completion to be open and visible by default (Table 5, row 1). The problem is

that these workarounds are again based only on the test tasks and not on reality. The outcomes of the workarounds may serve only the specific test task and not the other possible work tasks with the system. Many other similar hot fixes could be imagined outside the case study to improve effectiveness based on the published test tasks. On the other hand, these workarounds were not only "attractions" to success in tendering but also truly forced design tasks to meet the detailed and formal usability requirements.

Some further workarounds were applied with efficiency requirements although the workarounds above also improved this aspect (e.g. user can find the goal faster when task completion is helped). When the goal of usability design is to decrease the time on task or number of navigation steps per se, the best solution would be to design a shortcut into the users' first view that brings the user directly to the goal (Table 5, row 2). Furthermore, if the next test task is not related with the task at hand or it starts from a view that users should already be familiar with (from the previous tasks as discussed in the previous section), usability designers do not have to think whether users know where they end up or what happened after using the shortcut. This is clearly in conflict with the general guidelines of usability e.g. Nielsen's (1993) heuristics. Three related quotes from the evaluation report recommendations exemplify the situations above:

"No interaction steps are required in task 4, if the [Appointment Details] window is open by default, and if the information in the [Appointment Details] contains the data requested in the task"

"Furthermore, finding [Modify Appointment Details] path is too hard (in task 6)… Plan task 6 together with task 4: If the [Appointment Details]-window is open by default, ADD e.g. '[Modify Appointment Details]' option to this context menu (currently there is only a 'refresh' option), which would then open immediately the '[Modify Appointment Details]' -window."

"Note that the user does not have to know 'what happened' or 'where I am' (if they are suddenly in the [Modify Appointment Details] -window tab), because the next task (7) is to add new appointment, which users presumably know already at that phase of test."

These workarounds are applied at the expense of general user satisfaction. On the other hand, workarounds may also increase user satisfaction because test tasks are solvable more easily and faster without much cognitive effort needed. At the time of testing there were rather many unimplemented or untailored system features, that when accessed incurred an error message. This may not be the situation in other tendering contexts, but after eliminating possible errors in navigation and decreasing the number of execution steps, a one type of workaround was to avoid showing these error dialogs and messages (Table 5, row 3). The justification was that, if the user makes many slips and mistakes in navigation and sees many error dialogs, the user becomes annoyed. This would give the user a feeling of an erroneous system, and of course, increase the number of steps (due to closing the dialog and navigating back). Error messages are, however, inherent part of complex systems that can inform and train users for further actions.

5 Discussion

In the case presented here, we found that usability practices during a tendering process differ from a traditional user-centred design cycle in many more ways than limited redesign and iteration possibilities (cf. Beuscart-Zephir et al., 2002). The examples of different usability workarounds to improve task completion and performance efficiency were solely based on the call for tender documents. Usability work was aimed at winning the competition i.e. focused on acquiring better scores from usability. Workarounds were possible and had to take place for the system demonstration and testing purposes in order to satisfy the usability requirements although such practices were not favoured in the first place. In that sense, usability was over-emphasized rather than underachieved by the vendor (cf. Jokela et al., 2013). The possible implication is that while the system may score higher in the purchas-

er's evaluation the resulting system configuration may not be feasible at all for the real tool use situation in the future.

The most problematic characteristics in the call for tender were the published low-level formulations of test tasks (i.e. descriptions users' work) accompanied with objective, user performance based, usability measures (effectiveness and efficiency). A very short time to implement the requirements and a limited communication highlighted the problems, which are found to be obstacles in other software tendering cases too (Hochstetter et al., 2012). Indeed, many of the found obstacles are related and have a cause and effect relationship. For example, if there had been more time, more communication and less interpretations would have taken place regarding the usability decisions. Similarly, if test tasks had not been published, less sub-optimization regarding efficiency and effectiveness of the system would have been possible. It is therefore the joint effect of these several factors and characteristics of tendering that needs further discussion.

Our findings comply with Jensen et al. (2013) in that defining and validating scenarios are highly important (see Chapter 2). In the case, we could only wish that the formal usability criteria were created based on an in-depth understanding about the current and desired work practices and needs of the users. Purchasers definitely do possess and can produce this kind of knowledge. Yet, against Jensen et al. (2013), we claim that their articulation and operationalization of this knowledge in the call for tenders calls for alternatives. Publishing detailed test tasks may treat vendors equally, but may lead to sub-optimization as described, and as were previously concerned by Kushniruk et al. (2010). We hope that the "strictness" that procurement policies require would not override more explorative evaluation approaches that are commonly used in early design phases (cf. Jensen et al., 2013). Every time pre-defined test tasks are used in usability testing, those cause premature commitments (Diaper, 2002), implicitly assume relation and order of tasks, and rule out wide spectrum of possible requirements (Cordes, 2001). Such pre-defined tasks are not very well fit for testing usability of early designs (see e.g. Kankainen, 2002 p. 44) or complex systems (Redish, 2007).

The tendering context itself and the related workarounds are hard to influence due to legislation. However, purchasers are able to define the duration of the tendering process and phases. In our case, 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. Purchasers must carefully consider whether they want to arrange competition for systems that are "as-is" or "to-be" and give appropriate time to vendors to respond to usability requirements. As-is means that there is absolutely no time for modifications or that those are not allowed. In contrast, to-be asks for the best system to be delivered by the vendor where time and communication are not obstacles. Both types have their challenges in evaluation. Nevertheless, the decision of time relates to the above decision concerning task scenarios: How well the purchaser understands and communicates their desired uses and user needs and how detailed usability requirements and procedures they should write to the CFT.

Task completion, time and errors are definitely appropriate measures for systems comparison in tendering. Measures are objective and beneficial for example when comparing the future costs of usage of different systems and legitimating possible higher investment costs (Riihiaho et al., 2015). These measures are also well justified in the real use situation, because in most environments, it is beneficial, for example, to use less time for a given task, and in some cases a justified range for values can be created. However, it is possible to improve the efficiency of the system and its usability scores independent of the context of use. In our case for example, users had to input an end date to an appointment under creation as a part of the test task. We observed that in the candidate system *"the end date - field requires filling a time which is mostly unnecessary. Set a default time to 23:59 or make the time field optional."* We did not know whether the field should be mandatory in real use context, but we knew that it would increase users' time on task and number of steps taken. Therefore, mandatory fields (that decrease usability scores) can be removed and responsibility let to users, if the work and the test task (or our knowledge of these) does not specifically ask for such limitation. We could assume that in

work contexts, users know how their work is done and that they make it more efficiently without too many limits - especially when the developer does not yet know the limits. Perhaps test users will enter the time value in the sessions and no difference in efficiency and usability score happens (or they will not enter). The point is that similar design decisions at this level of user interface are possible irrespective of the context. Thus, workarounds related to improving efficiency and effectiveness are applicable also in other type of tendering situations even when test tasks are not published in detail. On the other hand, the more context-dependent the found usability problem and its possible solution is, the more risk there is to do harm in usability design during tendering, which lacks decent knowledge about the context. Another simple example of a universal redesign recommendation from our case is when the "*list requires scrolling"* and we suggest to *"increase the window size."* This standpoint presupposes that artifacts have attributes that can be evaluated independently and without a direct connection to all aspects of its future use environment or its future usefulness in that environment. These attributes are, for example, design conventions of user interfaces and usage patterns that users know from other systems or their generic knowledge of IT (e.g. closing the window in MS Windows). By measuring these, purchasers dismiss "what is missing" and "what is wrong with the system" in the evaluation. The feature of a mandatory field can be an element missing from the system, yet not have an effect on evaluation results. Of course, test scenarios and tasks aim to answer even these questions and bring usability-in-context –view into play, but based on our findings, scenarios cannot be so comprehensive that optimization at the artefact level of usability would become impossible.

Usability evaluation in tendering seems to weave somewhere between being contextual and universal (artefactual), formative and summative, targeted at early and finished products. On the one hand, usability requirements and scenarios embed the view of organizational usability where usability stems from aligning IT and work (see Elliott and Kling, 1997). On the other hand, this is measured as efficiency and effectiveness as time and errors, which absolute values are not of main interest or even relevant in the context, but the differences in values between the systems. HCI research could contribute in seeking possible alternative usability practices and criteria for tendering situations. We consider that, for example, applying different perspectives on usability (see Elliott and Kling, 1997; Uldall-Espersen, 2007; Hertzum and Clemmensen, 2012), searching appropriate method ingredients and mixes for tendering (see Woolrych et al., 2011), and getting absorbed into each usability problem and their qualifiers as usability criteria (see Vilbergsdottir et al., 2014; Tarkkanen et al., 2015) have potential in removing the obstacles and improving CFTs.

6 Conclusions

In this study, we investigated how the usability work of IS vendor takes place during the tendering process and how the tendering context effects this work. We conducted a case study and participated in a large tendering project on the vendor's side, where we performed usability evaluations, in order to redesign their candidate systems to meet the purchaser's usability requirements. The premises and goals of usability work during the tendering can become very different from what is meant with good usability and interaction design. We found that very rigid and formative usability requirements in the CFT may force the vendor to use usability workarounds in order to satisfy the criteria before the deadline. Especially the low-level and detailed formulation of usability test tasks directed the usability design of the case towards these sub-optimal solutions. These design solutions may not satisfy the principles of usability and human-centered design, but may help in increasing the score of usability. To our knowledge, such possibilities and weaknesses of usability evaluations in the tendering context have not been identified and described empirically earlier. This is due to lack of research on IS vendors' usability practices during tendering, which calls for further research. In the case, workarounds were possible because the purchaser provided usability test tasks and scenarios for the vendors in advance. Researchers should help also purchasers to consider whether these are necessary acts in their situations.

We acknowledge that the research carried out and represented in this paper is not fully generalizable to all contexts and situations within IS tendering. Lack of details about the system and the project hinder the generalization of findings, theoretical considerations and practitioners' possible take away. In addition, the study lacks interviews of vendor representatives and exploits only data from our evaluations as part of vendor's practical activities in tendering. Moreover, the study would have benefited from more longitudinal data collection until system redesign and actual use at the purchaser's site. Despite these limitations, our case experiences and specific findings pinpoint the obstacles that vendors and purchasers definitely would like to overcome if confronted. The responsibility of HCI practitioners and researchers is to evaluate and reflect their practices during similar situations and develop means to overcome possibly harmful outcomes in the usability of the systems. Especially, the joint effects of the purchasing decisions on vendors' behavior and resulting systems call for further research. This means studying the combination of time given, communication allowed, evaluation procedures applied and measures established.

References

Alter, S. (2014). "Theory of workarounds." *Communication of AIS* 34, 55, 1041-1066.

Baskerville, R. (1999). "Investigating information systems with action research." *Communications of the AIS* 2 (19), 1-32.

Beuscart-Zéphir, M. C., L. Watbled, A. M. Carpentier, M. Degroisse and O. Alao. (2002). "A rapid usability assessment methodology to support the choice of clinical information systems: a case study." In: *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 46-50.

Billestrup, J., A. Bruun and J. Stage. (2015). "UX requirements to public systems for all: Formalisation or innovation." In: *INTERACT 2015 Adjunct Proceedings*, 407-427.

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

Collins English Dictionary. (2009). HarperCollins Publishers.

Cordes, R. E. (2001). "Task-Selection Bias: A Case for User Defined Tasks." *International Journal of Human-Computer Interaction* 13(4), 411-419.

Diaper, D. (2002). "Scenarios and task analysis." *Interacting with Computers* 14, 379-395.

Elliott, M. and R. Kling. (1997). "Organizational Usability of Digital Libraries: Case Study of Legal Research in Civil and Criminal Courts." *Journal of the American Society for Information Science* 48 (11), 1023-1035.

Eriksson, P. and A. Kovalainen. (2008). *Qualitative Methods in Business Research.* Sage Publications.

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

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

Hertzum, M. and T. Clemmensen. (2012). "How do usability professionals construe usability?" *International Journal of Human-Computer Studies* 70 (1), 26-42.

Hochstetter, J., C. Cachero, C. Cares, and S. Sepúlveda. (2012). Call for Tenders Challenges in Practice: A Field Study. In: *Proceedings of the XV Congreso Iberoamericano en Software Engineering (CIbSE*12), Buenos Aires, Argentina.

ISO. (1998). "*International Organization for Standardization 9241-11:1998 Guidance on usability*." URL: http://www.iso.org (visited on 08/14/2015).

Jensen S., S. L. Rasmussen and K. M. Lyng. (2013) "Evaluation of a clinical simulation-based assessment method for EHR-platforms." S*tudies in Health Technology and Informatics* 205, 925-929.

Jokela, T., J. Laine and M. Nieminen. (2013). "Usability in RFP's: The current practice and outline for the future." In: *Human-Computer Interaction. Applications and Services*. Ed. by M. Kurosu. Springer Berlin Heidelberg, 101-106.

Jokela, T. (2010). "Determining Usability Requirements into a Call-for-Tenders. A Case Study on the Development of a Healthcare System." In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction.* ACM, 256-265.

Kankainen A. (2002). "*Thinking Model and Tools for Understanding User Experience Related to Information Appliance Product Concepts*." PhD Thesis. Helsinki University of Technology.

Kushniruk, A.W., and V.L. Patel. (2004). "Cognitive and usability engineering methods for the evaluation of clinical information systems." *Journal of biomedical informatics*, *37*(1), 56-76.

Kushniruk, A.W., E.M. Borycki, K. Myers and J. Kannry. (2009). "Selecting Electronic Health Record Systems: Development of a Framework for Testing Candidate Systems." In: *Advances in Information Technology and Communication in Health*. Ed. by J.G. McDaniel. Fairfax, VA: IOSPress, 376-379.

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

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

Lehtonen, T., J. Kumpulainen, T. Jokela and T. N. Liukkonen. (2010). "To what extent usability truly matters? A study on usability requirements in call-for-tenders of software systems issued by public authorities." In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction.* ACM, 719-722.

Nielsen, J. (1993). *Usability Engineering*. Academic Press.

Nieminen, M., J. Laine, S. Teräs, S., M. Runonen, V. Kalakoski, T. Valtonen, ... and E. Boryzki. (2014). "How to involve users in government system procurement?" In: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 805-808.

Molich, R and J.S. Dumas. (2008). "Comparative usability evaluation (CUE-4)." *Behaviour & Information Technology* 27, 263-281.

Pollock, N. (2005). "When is a work-around? Conflict & negotiation in computer systems development." *Science, Technology, & Human Values* 30 (4), 1-19.

Riihiaho, S., M. Nieminen, S. Westman, R. Addams-Moring and J. Katainen. (2015). "Procuring Usability: Experiences of Usability Testing in Tender Evaluation." In: *Nordic Contributions in IS Research*. Springer International Publishing, 108-120.

Redish, J. (2007). "Expanding Usability Testing to Evaluate Complex Systems." *Journal of Usability Studies* 2, 102-111.

Reijonen, P. and K. Tarkkanen. (2015). "Artifacts, Tools and Generalizing Usability Test Results." In: *Nordic Contributions in IS Research*. Springer International Publishing, 121-134.

Sauro, J. and J. R. Lewis. (2009). "Correlations among prototypical usability metrics: evidence for the construct of usability." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1609-1618.

Tarkkanen, K., V. Harkke and P. Reijonen. (2015). "Are we testing utility? Analysis of usability problem types." In: *Proceedings of the 4th International Conference on Design, User Experience, and Usability (DUXU)*, 269–280.

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

Tushnet, M. (2009). "Constitutional Workarounds". *Texas Law Review* 87 (7), 1499-1516

Uldall-Espersen, T. (2007). "Exploring multiple usability perspectives." In: *Proc of INTERACT 2007*. Springer Berlin Heidelberg, 628-632.

Vilbergsdottir, S. G., E. T. Hvannberg and E. L-C. Law. (2014). "Assessing the reliability, validity and acceptance of a classification scheme of usability problems (CUP)." *The Journal of Systems and Software* 87, 18-37.

Winkler, I. S. and E. A. Buie. (1994). "HCI challenges in government contracting." *SIGCHI Bulletin* 26 (4), 49-50.

Woolrych, A., K. Hornbæk, E. Frøkjær and G. Cockton. (2011) "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), 940-970.

Yin, R.K. (2003). *Case study research: design and methods*. 3rd edition. Sage Publications, Inc.

Yu, C-S. (2005), "Causes influencing the effectiveness of the post-implementation ERP system." *Industrial Management & Data Systems* 105 (1), 115-32.