In: HCI Human Computer Interaction – Theory and Practice (Part I), Jacko, J., Stephanidis, C. (Ed.)
Lawrence Erlbaum Associates Publishers, London, GB, 2003, pp 474-478

Lightweight Usability Engineering
Scaling Usability-Evaluation to a Minimum?

Ronald Hartwig, Cristina Darolti, Michael Herczeg
Institute for Multimedia und Interactive Systems - University of Luebeck
Willy Brandt-Allee 31a; D-23554 Luebeck, Germany
Hartwig|Darolti|Herczeg@imis.uni-luebeck.de

Abstract

Today there are several approaches available to develop usable software, but many products still lack a minimum of usability. One reason may be the lack of usability in some "heavyweight" usability engineering and evaluation methods themselves. It seems that many development teams think such elaborate process models are too complex to be applied, so they just don't do usability engineering. This paper offers a scalable "lightweight" approach to usability engineering. It starts from the idea that an easy and imperfect but used method is better than a complex unused method. The paper gives a short overview on the intentions, the basic ideas, the implementation, as well as experiences in using the method and discusses their possible application in product certification.

1 Introduction

"Nothing will ever be attempted if all possible objections must first be overcome." Samuel Johnson (Gediga & Furness 1993)

The European directive for "Work with Video Display Units" (referenced as VDU-directive, EU/90/270) forces the employers to apply usability principles to software based working systems. Because of this, software companies are willing or impelled to ensure a certain grade of usability for their products. The following approach is a rethinking of some well known and often costly and difficult-to-apply usability engineering methods. (Dicks 2002) states, "most product managers are just not willing to pay the price in time and money" for supporting traditional usability engineering in their company, especially since the results are not necessarily immediate or direct. They are just unsure about the benefit of following in detail each process step as many elaborate process models (like the ISO 9000-x) propose; the standards are perceived as too general or abstract. Similar to their "quick and easy" programming, developers, who still hold the power, want a simple way for usability. The recent summary of practitioners' feedback (Rosenbaum et al. 2000) supports this affirmation. It can be said that there is a need for usability of usability methods, or to cite (Nielsen 1995): "User interface professionals ought to take their own medicine some more."

Before discussing the usability of small scale usability engineering methods it is necessary to define a common ground for the scope of usability itself. The international standard ISO 9241 (ISO 1996) and its software related parts 10 to 17 seem currently to be the commonly accepted base, but still there are known critics and disadvantages: Sometimes this standard is very detailed;

sometimes only principles and examples are given. Therefore we focused on the definitions from ISO 9241 part 11 (defining usability as the product of effectiveness, efficiency and satisfaction) and the principles of part 10 (dialogue principles) and part 12 (presentation of information).

2 The Method

The idea of scaling and combining different approaches was the conclusion of practical experiences with more elaborate methods. The first indicator was found in a master thesis (Hartwig 1997). The study showed that the use of static usability criteria lists did not justify the investment. In the study a software system was analyzed using heuristics, user tests and a criteria list derived from ISO 9241 called EVADIS (Oppermann et al. 1992). In this experiment the combination of heuristics and usability testing (4 tests) achieved an overall performance of detecting 90% of 48 problems (100% of all fatal problems, 89% severe, 71% medium, 100% simple) while the item list only helped finding 32% of all problems (20% fatal, 67% severe, 41% medium, 0% simple). Questionnaires were not used because the user sample was too small and assuming a return rate of 10 to 30% this would not have been reliable enough in terms of the statistical requirements of testing theory. The initial finding was that in spite of the low performance of the criteria list, the effort of using it was immense.

During the usability support of the multimedia production project "VFH" it got clear that the formative use of questionnaires did not support the constructive consulting work very much but again the effort to do it was high (Hartwig et al. 2002). No fatal or severe problem was detected using questionnaires, but the effort to organize and accomplish the inquiry took several days and involved many organizational resources. A third cause to doubt the usability of (too) heavyweight methods were experiences made in the development of a general standardized usability evaluation standard called "ErgoNorm" for the german accreditation office "DATech" (see (Dzida et al. 2001)) which tries to constitute a common ground for usability (product or process) certification. The authors think that the very elaborate way to conduct the engineering process as proposed by ISO 9000 derivates seems to deter many practitioners. But resources and the question at hand should match (Wixon et al. 2002). This impression is supported by the feedback from several ACM workshops described in (Rosenbaum et al. 2000).

The drawbacks described above determined the basic, maybe self evident, idea of this method: to allow downscaling to available resources, project size and required quality. Compared to the heavyweight approaches, where completeness is mandatory, we allow for scaling down the investment in usability engineering to whatever the project offers. The basic methods used for defining users, goals and usability requirements, for designing, testing and iterating are the classical, well known ones, (as described in approaches like "scenario based design" (Rosson, Carol 2002) or "contextual design" (Holtzblatt, Beyer, 1996)) but they are applied at a different level of detail and accuracy. The important difference to naively cutting down all efforts is keeping the central idea of quality management in mind: All findings must be comprehendible. If during the usability testing a potential problem appears it is checked against the context of use (Herczeg 1994) first. Only if the context of use and task and user attributes allow it, we call it a deficiency. Issues are justified using heuristic methods and expert usability knowledge. Only those issues that really affect the usability criteria and are relevant for the analysed task are rated as a problem. This ensures a minimum level of validity. Using the ideas of scenario based design (Rosson&Carroll 2002) ensures that at least the main tasks of the user are actually tested and reviewed. But in contrast to heavyweight approaches it is seen as sufficient to have one or two rough and general scenarios and to derive a small set of requirements instead of doing a full

featured version. The next subchapter gives an idea on how and where heavyweight and lightweight process versions differ.

Decisions about the level of accuracy i.e. the testing quality (lightweight vs. heavyweight) have to be a part of the general process and product quality goal definition process. Analogue to the development process this should take place as a "claims analysis" as described by (Rosson&Carroll, 2002). This allows for a legitimate declaration of quality and, thereby, for possible certification activities.

Table 1: Lightweight usability engineering in practice of the VFH-project ("lightweight") in comparison to the more elaborate DATech-Process ("Heavyweight")

Process activity	Heavyweight	Lightweight	Reasoning
Context of use documentation	Several use scenarios observed in real life context with iterative re-validation	Generic scenario derived from observations during first user tests	Being on site at the users workplaces for usability testing was a good time to document usage scenarios as well. For the first tests a hypothetical scenario was assumed and then adopted after testing.
Requirements engineering	The scenarios are analyzed and all requirements that evolve are documented. Test criteria are derived from these requirements	Critical incidents from user testing and reviews as well as questions from developers are used as indicators for clarification needs. Mainly these "open questions" are taken as starting points for scenario based requirements engineering.	Many requirements are obvious enough or fulfilled anyway so they don't have to be documented at all. Instead of a complete top-down approach, potential problems are used as actuators if they were confirmed to be relevant. Only these are target to requirements engineering. This allows shortening the documentation, communication and evaluation of requirements to an absolute minimum.
Evaluation	Expert reviews, usability tests and questionnaires are conducted based on the criteria derived from the scenarios.	Expert reviews are done based on the smaller set of criteria from above. Usability testing is combined with the context of use analysis. Questionnaires are only used on a more general level and mainly for summative purposes.	The main goal of the evaluation is formative, so it mainly relies on expertise and user testing. Both are loosely connected to the criteria developed so far. If critical incidents appear, the context of use documentation and the requirements engineering are refined for this part. The more summative questionnaires are used to back up the assumptions of the overall quality

Because of the limited space the implementation of the process is only denoted here (see Table 1) to give an idea on how this lightweight process looked like in practice of the VFH-project ("lightweight") in comparison to the more elaborate DATech-Process ("Heavyweight"). A more detailed description can be found in (Hartwig et al. 2002).

3 Results

It is difficult to quantify the success of the described lightweight method. In practice it was impossible to do the same project twice in order to compare the usability of this minimal approach to the complete heavyweight approach. But there are hints from practical use which support this more flexible way of applying usability engineering methods. The described method was used to assure the usability of 35 multimedia learning modules in a 5 year production process. About 120 persons were involved in the production process; these were spread over 5 major development laboratories all over northern Germany. About 7 man-years of usability engineering were available from two persons. 91 users were polled using standardized questionnaires after about 15 of the modules had been released and were used in daily business. Users were asked to answer how important a usability item is for their personal use and how satisfied they are with respect to this item. The answers were combined to a stress/relief-rating. Items which were rated as important influenced the rating more than those which users thought to be "nice to have" (see (Hartwig et al. 2002) for details). The analysis showed that those modules that were part of this lightweight usability engineering process were rated significantly better than those which where not affected. Also the overall average user satisfaction indicated a preference for those modules which had been produced using the described process model. Effectiveness and efficiency were rated based on the user tests where obstructions and impediments got obvious. Again, those modules which were part of the process had less fatal or severe problems. Compared, at least, to "no usability engineering at all" this method seems to support a better overall usability. The comparison to more elaborate methods seems self evident: It was simply impossible to allocate more resources to the usability engineering task so insisting on a possibly more exact and in terms of testing theory "better" engineering and evaluation approach would have stopped the project or at least all usability engineering activities.

4 Discussion

First of all the described results are specific to the project VFH and may not directly be taken as evidence for other contexts. But the indicators show that our minimal approach is appropriate for many other development projects as well, although a number of problems arise: Scalability, which gives way for customization, also gives way for possible poor application of the method; so the success of the method depends on the way it is applied to the context. The authors think that in practice an easy method which introduces errors is better than no usability engineering at all. Problems with small test samples may by overcome using qualitative usability methods which are independent of the number of test-users. Concentrating on the main goal to make tools helpful for accomplishing tasks is the key to sensibly allocate scarce resources: "A lot can be accomplished quickly, easily, with little time, effort, or expense. The main secret is to observe real people doing real tasks." (Norman 2000). Relying on user testing has shortcomings because (especially laboratory) testing is always an artificial situation, participants are rarely representative of the whole population and the results, even if positive, cannot ensure that the product works. Testing focuses on tasks and therefore may not uncover larger problems like problems in overall system conceptual model (Dicks 2002). But both, "heavyweight" and "lightweight" methods suffer of this. In the end we mention the idea of "lightweight" quality certification by this method. It may be doubtful if it is fair to give away certificates based on such a method. Since quality may also be seen as scalable and controllable, the customers must be told about the limitations of the certificates. Any certificate must not be used as a 100% usability guarantee. Usability certificate should be seen as one minimal criterion among others (functionality, price, strategic decisions)

when buying or planning software and its usage. It shows that at least the main tasks can be done with this software without too much harm.

References

Dicks, R. S. (2002). Mis-Usability: On the Uses and Misuses of Usability Testing. Proceedings of the 20th annual international conference on Computer documentation**,** ACM Press.

Dzida, W., & Freitag, R. (2001). Usability Testing - The DATech Standard. In Wieczorek, Meyerhoff (Editor): Software Quality - State of the Art in Management, Testing And Tools. Springer, pp. 160-177.

Gediga, & Furness, T. (1993). Virtual Worlds – Why and When? User Interface Strategies a live satellite TV Broadcast, University of Maryland, College Park

Hartwig, R. (1997). Gestaltung und Bewertung der Gebrauchstauglichkeit von Software - Überprüfung der Machbarkeit am Beispiel der Benutzungsoberfläche eines Moduls eines Schichtplangestaltungssystems (Thesis); Universität Oldenburg FB Informatik, Oldenburg

Hartwig, R.; Triebe, J.K.; Herczeg, M. (2002): Usability Engineering as an Important Part of Quality Management for a Virtual University. In: Proceedings of Networked Learning 2002 - Challenges and Solutions for Virtual Education. Technical University of Berlin, Germany, ICSC-NAISO Academic Press, Canada.

Herczeg, M. (1994). Software-Ergonomie – Grundlagen der Mensch-Computer-Kommunikation, Addison-Wesley, Bonn; Paris ; Reading (Mass.)

Holtzblatt, K., & Beyer, H. (1996). Contextual Design: Principles and Practice. Field Methods for Software and Systems Design. D. Wixon and J. Ramey (Eds.), New York: John Wiley & Sons, Inc..

ISO (1996): ISO 9241 – Ergonomic requirements for office work with visual display terminals (VDTS) – International Standard

Nielsen, J. (1995): Getting usability used, in Human-Computer Interaction – Interact 95, Chapman&Hall, London

Norman, D. (2000). Usability Is Not A Luxury. Retrieved November, 2002, from http://www.jnd.org/dn.mss/UsabilityIsNotLuxury.html

Oppermann, R., Murchner, B., Reiterer, H., Koch, M. (1992): Softwareergonomische Evaluation - Der Leifaden EVADIS II (2. Auflage) --Berlin; New York: Walter de Gruyter

Rosenbaum, S., Rohn, J.A., & Humburg, J. (2000). A toolkit for strategic usability: results from workshops, panels, and surveys. Proceedings of the CHI 2000, ACM Press, New York.

Rosson, M. B., & Carroll, J.M. (2002) Usability Engineering – Scenario based development of human-computer interaction, Morgan Kaufmann Pub.; San Francisco.

Wixon, D. R., Ramey, J., Holtzblatt, K., Beyer, H., Hackos, J., Rosenbaum, S., Page, C., Laakso, S., & Laakso, K. (2002). Usability in practice: field methods evolution and revolution , in Proceedings of the CHI 2002, ACM Press, New York, 2002