In: HCI Human Computer Interaction – Theory and Practice (Part I), Jacko, J., Stephanidis, C. (Ed.), Lawrence Erlbaum Associates Publishers, London, GB, 2003, pp 946-950

Usability Engineering in Computer Aided Learning Contexts Results from usability tests and questionnaires

Ronald Hartwig, Inga Schön, Michael Herczeg
Institute for Multimedia and Interactive Systems, University of Luebeck
Willy-Brandt-Allee 31a, D-23554 Luebeck
Hartwig|Schoen|Herczeg@informatik.uni-luebeck.de

Abstract

This paper describes a process to consolidate observations and findings from usability evaluation phases into a comprehensible list of usability deficiencies as a basis for quality assurance methods. A so called "evidence test" process is transferred from the world of software usability engineering into the context of computer aided learning (CAL) and computer supported cooperative learning (CSCL). It is based on experiences from quality management and evaluation efforts in the German flagship project "VFH" (Virtual University of Applied Sciences in Engineering, Computer Science and Economics) and the project "medin". The paper is intended as a basis for continued research in the area of usability of computer aided learning, as well as a founded best practice description for practitioners who work in this context.

1 Introduction

The work and results we describe are taken from the context of two projects called "Virtual University of Applied Sciences" (Virtuelle Fachhochschule - VFH) and "medin", which intend to offer location independent learning. Both are sponsored by the German Federal Ministry of Education and Research (BMBF). The focus of interest lies in computer-supported multimedia-based teaching and the production of learning material which offers distance learning to students dispersed all over Germany. Quality management, especially in projects of this size, needs a well defined process as a guideline for all participating parties. It has been one major task to implement a procedure which is suitable to enhance or even enforce the usability of the products, namely multimedia learning modules for the use in a virtual university. The main interest is to closely connect usability tests and reviews to the original context of use by defining a comprehensible and reliable method on how to extract usability requirements from use scenarios and to test the compliance of the produced material with these requirements. This paper describes how results from the usability testing within the installed iterative process were analyzed and classified in order to validate their relevance for further improvement of the modules.

2 Didactical models

Unlike the working tasks known from usability engineering, the task "learning" is not a home homogenous process. Along with the external context of learning (e.g. the environment) the internal conditions (learning prerequisites and admission conditions) must be considered. The premises for learning change during the learning process; this causes the sequencing of the learning material to play an important role (Niegemann 1995). In traditional instructional design it is an important rationale to have processes that take care of learning premises and goals for the

learning content. For example (Gagne, Briggs & Wager 1992) differentiate five learning goals categories (knowledge represented by words, cognitive abilities, cognitive strategies, settings, motoric abilities). Instructional design is mainly based on the assumption that at the beginning of planning and developing of teaching instructions, the desired abilities should be analyzed on the basis of these categories. One example is the ARCS-Model (motivation as a task of instructional design). The model was extended with concrete recommendations for multimedia learning environments by (Keller & Suzuki 1988) and (Niegemann 2001). Different instructional design theories try to specialize on this theory to the field of computer supported learning. Examples are “Principles of Instructional Design” from Gagné, Briggs and Wager as mentioned above and, the more complete and less analytical, Instructional Transaction Theory (ITT) from (Merrill 1999). All theories have in common that students, through work with the computer program, should be stimulated to study the learning material intensively. The goal is not only to present the learning material in an attractive way, but also to support explorative and individualized learning. The connection to the usability engineering is that the above mentioned didactical theories imply certain learner activities. These activities correspond to tasks and sub-tasks known from usability engineering. Please note that an activity (“making an annotation”) is not the task (“learning algebra”) but a vital part of it. In reverse analyzing activities helps to identify tasks and subtasks.
Usability, as a central software quality, is seen as a prerequisite for giving the students the opportunity to learn efficiently. Typical requirements from learning contents are that users must always be able to keep an overview of learning paths in order to plan their proceeding. Students must be offered the possibility to handle parts of the learning material in the learning process by marking, excerpting and annotating. Explorative use of the acquired knowledge must be supported by communication possibilities (chat, whiteboard, newsgroups). In order to leave a maximum of mental resources for managing the complexity of the learning contents only an absolute minimum of elaborateness may be allowed for the handling of the tools. It is important to let the learner learn the contents and not the system usage. Therefore, the ergonomic design is a necessary but not sufficient precondition for the success of didactical models. Of course it must still be the major concern to plan and implement the appropriate didactical model in time.

It is difficult to engineer and evaluate a complex interactive system with respect to a complex task like “learning”, so we use the “divide et impera” method to cope with this complexity. In order to apply usability engineering methods it is essential to identify the subtasks from the didactical model and their desired concrete results. Then potential usage problems are identified and their severity is consolidated in an evidence testing process (as described in (Dzida&Freitag 2001)). Of course it is a necessary but not sufficient precondition to optimize the subtasks but it is a starting point at all. Still the whole task should be considered as well. Another important point while transferring usability engineering methods into the world of CAL is, that the shortest way to a solution is not always the preferable one in terms of learning support. One has to distinguish between desired slow-downs that support learning, because they force the learner to intensify his work with the contents and usability problems which make the user think about the system instead of the content. Efficiency has to be measured with this in mind and if there is a planned detour in the learning path it must be justified by the didactical model but not by technical insufficiency. Another aspect is that the task “learning” ideally has to be evaluated over a longer time frame (e.g. a whole study program) which is often hard to achieve in practice.

3 Transferring the usability quality concept to the world of CAL

The approach used in the project is based on well established usability evaluation methods (see (Herczeg 1994)) and engineering approaches ((Nielsen 1993), (Mayhew 1999), (ISO13407 2000)).

First, common usability issues were covered by expert walkthroughs (Reeves et al. 2002) mainly concerning the presentation of information and some consistency issues. As (Dimitrova 2001) showed, relying on such expert reviews is not sufficient, because experts will overlook surprises. Second, empirical data from questionnaires showed which usability issues were considered by users to be most important and how those issues were managed in the offered learning material. Questionnaires were analyzed to give an importance rating as an indicator of what users expect from computer aided learning support. The questionnaire items refer to issues which already had been identified as potential problems during the iterative process. The questionnaire yields how grave these problems are from the user's point of view. In addition, the questionnaire reveals how content the user are with the current offer (module). As the third and major step, usability tests were conducted on site with real students in their living rooms or studies (See (Hartwig, Triebe & Herczeg 2002) for a more detailed description). In usability engineering it is assumed that this so called "Triangulation" of three usability evaluation methods delivers the most promising and reliable results. Each method has its strengths and limitations but put together they cover the potential problem fields of interactive systems. The result of this evaluation is a list of potential problem items: objections from the expert walkthroughs, observations of users struggling with the system from the user testing and questionnaire items with low satisfaction values or a negative impact. All potential problems are analyzed using an evidence test described below. This process also yields information on how severe a problem is and how it can be overcome.

3.1 Evidence test

Usability engineering is defined, according to the international standard (ISO9241 1996-2000), as a product of the following three criteria:

1. Effectiveness: It is seen as the degree of completeness and correctness regarding the goal.
2. Efficiency: The effort that has to be made in order to reach the goal. Effort can be rated absolutely as time consumption or (mental) work load or relatively compared to alternative methods/tools .
3. Satisfaction: The degree of user contentment when using the system. It is normally calculated on an empirical basis.

In a first step it is analyzed if the potential problem deters the problem solving at all or leads to faulty task results (effectiveness "fail"). In this case the potential problem is a severe failure and no further analyzing has to take place. Presuming that the effectiveness of the observed part of the system did not fail, in a second step the efficiency is considered: the user effort for completing the subtask is compared to the available amount of time for this subtask. Finally, if the potential problem did pass the effectiveness and the efficiency testing, a traditional empirical enquiry is done. Using a scale from 1 (very good) to 5 (poor) and 1 (very important) to 4 (not important) and considering a slight positive bias, low satisfaction (<2.5) together with higher importance ratings (<3) are taken as evidence for user stress and make the satisfaction criteria fail (see (Hartwig et al. 2002) for a more detailed description of this questionnaire).

The support for "using individual learning strategies and working with the material" complex is taken as an example to clarify the process of consolidating an observed potential problem analytically into a deficiency: The students had problems in working with the offered contents instead of just reading them. Their subtasks included all kinds of reusing the contents such as summarizing, transcribing, annotating, linking them into a personal hypermedia network or discussing them with fellow students. It could be observed that many of the students referred to paper as their working media so that they could work in the manner they were used to instead of using the possibilities of computer supported techniques. The following example should help to understand, why students did this and how the decision was made that this "fallback" is considered to be a usability problem. "**Annotations**": If students used the offered annotation tools they only were able to annotate web pages as a whole, but not to annotate specific text pieces. Strictly speaking it did not affect the effectiveness, because at least theoretically it was possible to complete the desired subtask of annotating. They theoretically could write line numbers as references into their annotations in order to link to specific paragraph. But if you compare this to the efficiency of writing some short notes (e.g. containing some mathematical formulas or drawings) directly into a printed version, it gets obvious that the computer supported method implies a lot more effort for the user and therefore the efficiency is rated as "fail". Having said that, the new question is: Is it bad to have the students make their annotations on paper instead of the computer? The missing of a computer supported tool is rated as a failure as well, because the use of paper based annotations implied many new sources of error, problems with re-referencing to the online-contents and it implicitly lead to higher efforts when students try to put it back into the computer based media again if they wanted to communicate their personal annotations to fellow students. This lack of efficiency and the probability of new effectiveness problems lead to the consolidated decision that this complex really is a severe usability problem, because both alternatives (using the computer for the annotations vs. using pen and paper) imply severe usability problems.

Tabular 1: How typical problems affected effectiveness, efficiency and satisfaction

	Usability problems while...	Effectiveness	Efficiency	Satisfaction
1	using individual learning strategies and working with the material	OK	Fail	Fail
2	being informed about news and changes	Fail	Fail	-
3	orienting and navigating within the presented material	OK	Fail	Fail
4	using a chat	OK	Fail	-
5	trying to work parallel with different information sources on one screen	OK	Fail	-
6	detecting and recovering interactive possibilities of the media	Fail	Fail	-

4 Results

In order to come to general results the faults are categorized in more general classes. These classes reflect problem complexes encountered during the project. The results are taken from testing 12 different modules in 15 on-site user tests and 91 questionnaires from about 50 users (from about 150 students). They are specific for the described project but we expect similar findings for other modules and contexts. Therefore table 1 is intended as a warning, how bad problems in those areas *can* affect the usability of the product with respect to the learning task. Many of problems already had been foreseen or did already appear in earlier iterations of the development process. The results of the questionnaire, which mainly contains questions related to these already anticipated problem areas, were predominantly positively rated by the students. This shows that the use of a project guideline was effective at least to avoid well known problems. But the fact that in spite of this still severe faults persisted shows that relying only on questionnaires and expert reviews would be careless. This would be a systematical mistake because grave problems could be overseen this way.

5 Summary

While the three evaluation methods contribute different views on the usability of computer aided (cooperative) learning, the combination of these three approaches is a promising starting point on what a general guideline for usability and a user- and task-centered process should cover. The results offer clues where, in our opinion, usability experts should be present in the design and development process in order to prevent technical constraints from interfering with the success of implementing the didactical models. Currently a semantic web with specific usability knowledge data (observations, guidelines, test results) as well as learning content and development instructions is developed in order to support the knowledge transfer throughout the whole project and the participating developers. They should be enabled to comprehend all decisions and corroboration processes in order to improve the overall usability of the product.

References

Dimitrova, M. (2001): "Are Experts Able to Predict Learner Problems During Usability Evaluations?" In: Proceedings of ED-MEDIA 2001. AACE: Finland, pp.1023 -1028

Dzida, W. and 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. ISBN 3-540-41441-X

Gagné, R. M., Briggs, L. J., Wager, W. W. (1992). Principles of Instructional Design. 4. Ed., Fort Worth: Harcourt Brace Jovanovich.

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/The Netherlands.

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

International Organization for Standardization (1996-2000): ISO 9241 - Ergonomic requirements for office work with visual display terminals, Parts 1-17. International Standard

International Organization for Standardization (1999): ISO 13407 - Human-centred design processes for interactive systems. International Standard

Keller, J.M., & Suzuki, K. (1988): Use of the ARCS motivation model in courseware design. In D. H. Jonassen (Ed.), Instructional designs for microcomputer courseware (pp. 401-434). Hillsdale, NJ: Erlbaum

Mayhew, D. (1999): The Usability Engineering Lifecycle. Morgan Kaufmann Publishers. (1999)

Merrill, M.D. (1999): Instructional transaction theory (ITT): Instructional design based on knowledge objects. In C.C. Reigeluth (Ed.), Instructional Design – Theories and models. A new paradigm of instructional theory (pp. 397-424) Mahwah, NF: Erlbaum

Niegemann, H. M. (1995): Computergestützte Instruktion in Schule, Aus- und Weiterbildung. Theoretische Grundlagen, empirische Befunde und Probleme der Entwicklung von Lehrprogrammen. Frankfurt am Main: Peter Lang

Nielsen, J. (1993): Usability Engineering, AP Professional, Boston

Reeves, T.C.; Benson, L.; Elliott, D.; Grant, M.; Holshuh, D.; Kim, B.; Kim, H.; Lauber, E.; Loh, S.: "Usability and Instructional Design Heursitics for E-Learning-Evaluation"; in Proceedings of ED-Media 2002, AACE, p. 1615-1622