# LUND UNIVERSITY

**An industrial case study of usability evaluation in market-driven packaged software development**

Natt och Dag, Johan; Regnell, Björn; Madsen, Ofelia S; Aurum, Aybüke

[Link to publication](#)

# An Industrial Case Study of Usability Evaluation in Market-Driven Packaged Software Development

**Johan Natt och Dag[1], Björn Regnell[1], Ofelia S. Madsen[2], Aybüke Aurum[3]**

[1]Department of Communication Systems, Lund University, Sweden, (johan.nattochdag, bjorn.regnell)@telecom.lth.se
[2]C-Technologies AB, Lund, Sweden, ofelia.madsen@cpen.se
[3]School of Information Systems, Technology and Management, University of New South Wales, Australia, aybuke@unsw.edu.au

**Abstract.** In market-driven software development it is crucial to produce the best product as quickly as possible in order to reach customer satisfaction. Requirements arrive at a high rate and the main focus tends to be on the functional requirements. The functional requirements are important, but their usefulness relies on their usability, which may be a rewarding competitive means on its own. Existing methods help software development companies to improve the usability of their product. However, companies that have little experience in usability still find them to be difficult to use, unreliable, and expensive. In this study we present results and experiences on conducting two known usability evaluations, using a questionnaire and a heuristic evaluation, at a large software development company. We have found that the two methods complement each other very well, the first giving scientific measures of usability attributes, and the second revealing actual usability deficiencies in the software. Although we did not use any usability experts, evaluations performed by company employees produced valuable results. The company, who had no prior experience in usability evaluation, found the results both useful and meaningful. We can conclude that the evaluators need a brief introduction on usability to receive even better results from the heuristic evaluation, but this may not be required in the initial stages. Much more essential is the support from every level of management. Usability engineering is cost effective and does not require many resources. However, without direct management support, usability engineering efforts will most likely be fruitless.

## 1. Introduction

When developing packaged software for a market place rather than bespoke software for a specific customer, short time-to-market is very important (Potts, 1995). Packaged software products are delivered in a succession of releases and there is a strong focus on user satisfaction and market share (Regnell, Beremark & Eklund, 1998). Thus, companies tend to put their primary effort into inventing and implementing new functional features that are expected to improve the product. Although developers rely heavily on the number and the existence of new features, usability is recognized as a competitive advantage on its own. Still after many years of usability engineering research, there are many companies that do not approach and explicitly improve usability. Although several methods and techniques exist (Nielsen, 1993) and studies show their cost-effectiveness (Bias & Mayhew, 1994), the seeming difficulty of approaching usability prevents the success of companies. Since usability evaluations of software products are necessary in order to increase user-friendliness (Nielsen 1993), there is a need to put even more focus on usability evaluation methods that are easy to use and adopt and that give fast and appropriate results.

In this paper we present an industrial case study that employs two known usability evaluation methods (Natt och Dag & Madsen, 2000). The study was conducted at Telelogic AB, a large software developing company in Sweden, and the methods were used to evaluate their main product, Telelogic Tau[1], a graphical software development tool aimed for the telecommunications industry. It is shown that the two methods may be used for continuous evaluation of usability without much effort or resources and without any particular experience in usability engineering.

## 2. Research methodology

Several factors have been taken into consideration when choosing evaluation methods. The methods must be easy to perform and give understandable results that can be utilized in daily work without extraordinary analysis. Furthermore, the methods need to be appropriate for use over and over again, and it must be possible to extend the proficiency in using these methods when experience in usability evaluation within the company increases. If usability experience is lacking, the methods must not be too sensitive to the evaluators' performances.

To obtain satisfactory results that fulfill these requirements, we carefully selected two known usability evaluation methods, a questionnaire and a heuristic evaluation, which give quantitative and qualitative results respectively.

---

1. Telelogic Tau is a registered trademark of Telelogic AB.

## 2.1 The SUMI Questionnaire

In order to obtain end users' opinions about the software we used a commercially available questionnaire, 'The Software Usability Measurement Inventory' (SUMI) (Kirakowski & Corbett, 1996). SUMI is a standard questionnaire specifically developed and validated to give an accurate indication of which areas of usability should be improved. It is tested in industry and mentioned in ISO-9241 as a method of measuring user satisfaction. The questionnaire consists of 50 statements to which each end-user answers whether he or she *agrees*, *disagrees* or is *undecided*. Only 12 end users are necessary to get adequate precision in the analysis, but it is possible to use fewer users and still obtain useful results.

The questionnaire was sent to 90 selected end users in Europe. Returned answers were sent to the questionnaire designers who statistically analyzed the answers and compared them with the results in a continuously updated standardization database. The database contains the answers from over 1000 SUMI questionnaires used in the evaluations of a wide range of software products. The comparison results in the measurement of five usability aspects:

- *Efficiency* - the degree to which users feel that the software assists them in their work.
- *Affect* - the user's general emotional reaction to the software.
- *Helpfulness* - the degree to which the software is self-explanatory. Adequacy of documentation.
- *Control* - the extent to which the users feel in control of the software.
- *Learnability* - the ease with which the users feel that they have been able to master the system.

Each aspect is represented by 10 statements in the questionnaire and the raw scores for each of the aspects are converted into scales with a mean value of 50 and a standard deviation of 10, with respect to the standardization database. Also, a *global* scale is calculated that is represented by the answers to 25 of the 50 statements that best reveal global usability.

To identify the statements to which the answers differ significantly from the average in the standardization database, *Item Consensual Analysis*, a feature developed by the questionnaire designers, is used. Through comparison between the observed and expected answer patterns, the individual usability problems may be more accurately determined.

## 2.2 The Heuristic Evaluation

To find usability problems specific to the evaluated software, we used a slightly extended version of a standard *heuristic evaluation* (Nielsen, 1994). In the heuristic evaluation, usability experts go through the interface and inspect the behavior of the software. The behavior is compared to the meaning and purpose of a set of ten guidelines called *heuristics* that focus on the central and most important aspects of usability, such as '*user control and freedom*' and '*flexibility and efficiency in use*'. This enables the evaluator to systematically check the software for usability problems against actual requirements in the specification and given features in the product. The result is a list of identified usability problems that may be used to improve the software.

In the market-driven development organization, there may be little experience in usability and usability experts may not be available. As this was the situation at Telelogic, we used experts on the software from within the company. The number of evaluators needed may then increase, as less usability problems may be found. Experts specializing only in usability tend to find problems mainly related to how easy the system is to operate, whereas domain experts rather find problems related to how well the system responds to its intended behavior. In this sense the usability and the domain experts complement each other, as domain experts, according to Muller, Matheson, Page and Gallup (1998), bring perspectives and knowledge not otherwise available. However, an evaluation is more likely to be conducted in the first place if we initially do not require the involvement of usability experts.

Twelve employees from within the company participated in the study. They were presented with a set of scenarios comprised of different tasks to perform. The method was extended in order to add increased structure to the evaluation and to help the evaluators to stay focused on usability issues. This was accomplished through the introduction of *Usability Breaks* (UB). A UB is a position in the detailed scenario where the evaluator is supposed to (1) stop the execution of the scenario and write down any problems found up to that point together with the associated heuristics, and (2) go through the ten heuristics to identify additional problems with the tasks just performed.

The evaluators were advised to spend 1 uninterrupted hour on finding relevant usability problems. In addition to generating a list of problems, the evaluators were asked to write down during which scenario identified problem were encountered, at what specific UB they were found, the heuristics that apply, the severity of each problem (high, medium or low), and a suggestion of a solution.

# 3.  Results

## 3.1  The SUMI Questionnaire

Of the 90 questionnaires sent to end users, 62 were properly filled out and returned. The analysis of the returned answers revealed that the evaluated software did not meet the appropriate standards on several aspects of usability. In Figure 1 the medians for each of the six SUMI scales are shown. The median corresponds to the middle value in a list where all the individual scores given by each evaluator have been numerically sorted. The figure also shows error bars for each scale, representing the upper and lower 95% confidence limits. As seen in the figure, all but two of the six SUMI scales are below average. The sub-scale *affect* is the only



**Figure 1.** The SUMI satisfaction profile.

one that lies above average, indicating that users feel slightly better about this product than they feel in general about software products. The *learnability* sub-scale indicates that the software may be regarded to be as easy or hard to learn as software products are in general.

The Item Consensual Analysis (see Section 2.1) revealed 9 specific statements that differed significantly from the standardization database (99.99% certain). In Figure 2, the statement that was most likely to differ from the expected is shown. This particular result reveals that the software may not be very helpful. Of the remaining 8 statements that most certainly differed from the standardization database, only 1 statement generated a more positive response than what was expected. As many as 74% of the end users disagreed with the statement '*if the software stops it is not easy to restart it*', indicating that the software *is* easy to restart. Further analysis of the statements revealed several reasons to the low scores in Figure 1.

*Statement 28*: *The software has helped me overcome any problems I have had using it.*

|           | Agree | Undecided | Disagree |
|-----------|-------|-----------|----------|
| Profile   | 12    | 19        | 31       |
| Expected  | 17.10 | 30.97     | 13.93    |
| Chi Square| 1.52  | 4.63      | 20.93    |

**Figure 2.** Results from the Item Consensual Analysis.

## 3.2  The Heuristic Evaluation

The heuristic evaluation revealed 72 unique usability problems directly related to the software application. A sample usability problem identified by an evaluator is shown in Figure 3. About 20% of the identified problems were considered highly severe, about 65% somewhat severe, and no more than 14% less severe. This indicates that usability needs attention in order to increase the usefulness of the software, which is confirmed by the results from the more reliable SUMI questionnaire evaluation.

Solutions were given to most of the problems but for the 18 that had none a solution may be inferred through the problem descriptions and through the particular UBs in the scenario. The problem descriptions had high enough quality to be used as input into the requirements process.

Figure 4 shows the number of times each of the ten heuristics were used to classify the identified problems. The high use of *flexibility and efficiency in use* indicates that users may get frustrated when using the software. Further analysis of the particular problems related to this heuristic and at which UBs
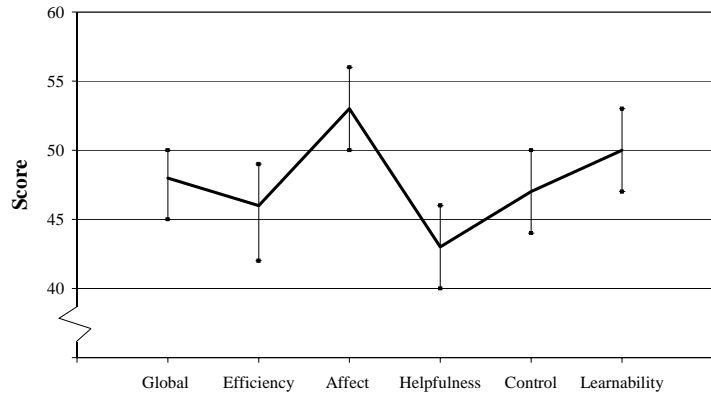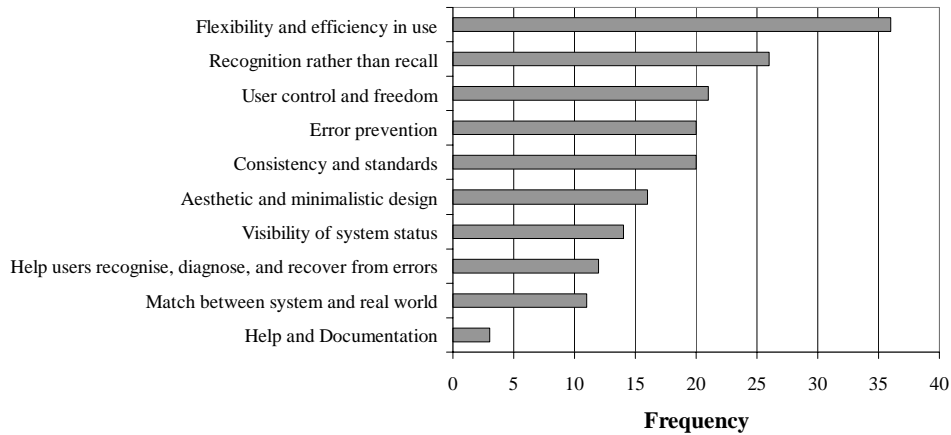
| Problem   | How to change a unidirectional channel to a bidirectional channel |
|-----------|-------------------------------------------------------------------|
| Scenario  | A                                                                 |
| UB        | 1.3                                                               |
| Heuristic | Flexibility and efficiency in use                                 |
| Severity  | High                                                              |
| Solution  | Add a channel symbol to the symbol menu and then add symbols for unidirectional (one in each direction) and bidirectional. |

**Figure 3.** Sample usability problem found by an evaluator.

**Figure 4.** The number of heuristics used to classify the 72 identified usability problems.

the problems were encountered, reveals that there are many tasks that are bothersome to complete due to non-intuitive functionality and a non-supportive graphical interface.

The results in Figure 4 are confirmed by the results from the SUMI questionnaire evaluation. *Efficiency* is identified as a problematic area by both methods and *helpfulness* is related to *recognition rather than recall* and *user control and freedom*. This and the fact that experts on the software were used as evaluators indicate that the heuristic evaluation pinpoints relevant usability problems.

# 4. Conclusions

In this paper we have presented the results of using two known usability evaluation methods at a market-driven software development company inexperienced in usability. We have found that although experience on usability is lacking, the two methods are easy to use and do not require many resources. The questionnaire is available at a low cost and system experts can easily develop the heuristic evaluation scenarios. Furthermore, the two methods complement each other very well. Both kinds of result were found to be usable and meaningful by the developing company and the generated problem list was particularly welcomed. The results gave them insight into the specific areas that needed improvement and helped them to appreciate the issues to put their effort into.

The selection of evaluators was the most time-consuming task. Mainly, this was because there was little support for usability issues. There was a noteworthy interest from the development department, but we have found that management support on every level in the organization is crucial to effectively get results. Without management support it is not very likely that the results will be used in further development at all. Also, a short 30-minute introduction to the concept of usability will most likely motivate the evaluators to perform even better and be more focused on usability issues in particular.

The initial drawbacks were nevertheless highly compensated by the low cost and the quick and useful results. The estimated costs of applying the methods are shown in Figure 5 (Melchior, Bösser, Meder, Koch & Schnitzler, 1995).

| | **Cost (figures in man-days)** | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Small** | **Medium** | **Extensive** | **Training** | **Material** | **Reliability** |
| Heuristic Evaluation | 2 | 4 | 4 | 1 | None | Medium |
| SUMI Questionnaire | 1 | 3 | 3 | 2 | US$500 | High |

**Figure 5.** Estimation of cost of applying the two usability evaluation methods.

Currently, we are applying a follow-up study to investigate precisely how the generated problem lists have been used in succeeding releases, what impact on software usability they have had, and to what extent the usability has increased.

## Acknowledgements

## References

Bias, R. G., & Mayhew, D. J. (1994). Cost-justifying usability. Boston, MA: Academic Press.

Kirakowski, J., & Corbett M. (1996). The software usability measurement inventory: Background and usage. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability Evaluation in Industry (pp. 169-177). London, UK: Taylor & Francis.

Melchior, E.-M., Bösser, T., Meder, S., Koch, A., & Schnitzler, F. (1995). Usability Study: Handbook for practical usability engineering in IE projects. ELPUB 105 10107, Telematics application programme, Information engineering section.

Muller, M. J., Matheson, L. Page, C., & Gallup, R. (1998). Participatory Heuristic Evaluation. Interactions, 5(5), 13-18.

Natt och Dag, J., & Madsen, O. S. (2000). An Industrial Case Study of Usability Evaluation. Lund: Lund University. (ERIC Document Reproduction Service No. LUTEDX (TETS-5390)/1-190/(2000)&local 8)

Nielsen, J. (1993). Usability Engineering. San Francisco, CA: Morgan Kaufmann.

Nielsen, J. (1994). Heuristic Evaluation. J. Nielsen, & R. L. Mack (Eds.), Usability Inspection Methods (pp. 25-61). New York, NY: John Wiley & Sons.

Potts, C. (1995). Invented requirements and imagined customers: Requirements engineering for off-the-shelf software. Proceedings of the Second IEEE International Symposium on Requirements Engineering, pp. 128-130.

Regnell, B., Beremark, P., & Eklundh, O. (1998). A market-driven requirements engineering process: Results from an industrial process improvement programme. Requirements Engineering, 3(2), 121-129.