



SINTEF REPORT

SINTEF ICT

Address: NO-7465 Trondheim,
NORWAY
Location: Forskningsveien 1
Telephone: +47 22 06 73 00
Fax: +47 22 06 73 50

Enterprise No.: NO 948 007 029 MVA

TITLE

Analysis of usability evaluation data: An interview study with usability professionals.

AUTHOR(S)

Asbjørn Følstad, Effie L.-C. Law, Kasper Hornbæk

CLIENT(S)

Norwegian research council, VERDIKT programme

REPORT NO. SINTEF A13889	CLASSIFICATION Open	CLIENTS REF. P.nr. 180135/S10	
CLASS. THIS PAGE Open	ISBN 978-82-14-04461-4	PROJECT NO. 90B243.82	NO. OF PAGES/APPENDICES 14
ELECTRONIC FILE CODE NA	PROJECT MANAGER (NAME, SIGN.) Jan Heim	CHECKED BY (NAME, SIGN.) Erik G. Nilsson	
FILE CODE NA	DATE 2009-09-28	APPROVED BY (NAME, POSITION, SIGN.) Bjørn Skjellaug, Research director	

ABSTRACT

While the planning and implementation of usability evaluations are well described in the HCI literature, the analysis of the evaluation data is not. This implies that we know little of the current practice of usability professionals on this later phase of usability evaluation, and that we have an insufficient basis for research on methods and tools to support analysis.

We present interviews with 11 usability professionals on how they conduct analysis, including the resources, collaboration, creation of recommendations, and prioritization involved. The interviews provide three main findings:

- Analysis seem to be conducted according to pragmatic and informal approaches
- Collaboration among analysts seem to be fairly frequent. Two prevalent purposes of such collaboration was (a) identification of more problems and generation of better design suggestions, and (b) quality control.
- The analysis process seem to involve both problem identification redesign suggestions, even though the latter is mainly unsupported by existing usability evaluation methods.

We discuss how future research could be to better support analysis. Also, we present three questions that usability professionals may consider during analysis.

KEYWORDS	ENGLISH	NORWEGIAN
GROUP 1	ICT	IKT
GROUP 2	HCI	Menneske-maskin interaksjon
SELECTED BY AUTHOR	Usability evaluation	Evaluering av brukskvalitet
	Interview	Intervju
	Usability professionals	Usability praktikere

TABLE OF CONTENTS

1	Introduction	3
1.1	Analysis in usability evaluation	3
1.2	Interviews on analysis in usability evaluation	3
1.3	Motivation for the study.....	3
1.4	Structure of the report	4
2	Main findings	4
2.1	Informal and pragmatic analysis	4
2.2	Purposes of collaboration.....	4
2.3	The importance of redesign suggestions.....	5
2.4	Severity classifications.....	5
3	Implications.....	6
3.1	Implications for usability research.....	6
3.2	Implications for professional practice.....	6
4	Detailed results	7
4.1	Theme 0: The context of the reported evaluations.....	7
4.2	Theme 1: Formality and structure.....	7
4.3	Theme 2: Analysis resources	9
4.4	Theme 3: Collaboration	9
4.5	Theme 4: Evaluator bias and multiple analysts	10
4.6	Theme 5: Problem identification & redesign suggestions	10
4.7	Theme 6: Sources of redesign suggestions	11
4.8	Theme 7: Severity classifications	11
4.9	Theme 8: Recommendations, challenges, and needs.....	12
4.10	Summary of results	12
5	Method.....	12
5.1	Participants.....	12
5.2	Procedure	13
5.3	Analysis.....	13
5.4	Limitations and future work.....	13
	References	13

1 Introduction

1.1 Analysis in usability evaluation

Usability evaluation has become indispensable for HCI practice and research. The literature shows that evaluation is cost effective and useful across various methods and domains [1]. However, performing sound usability evaluations is difficult. In think aloud tests, evaluators need to plan tasks carefully, to recruit and handle participants, and to ask valid questions during the test [5, 15]. In usability inspection, evaluators need to apply the method validly and avoid false positives [2]. Finally, communicating findings to developers and designers, and getting them and other stakeholders to act on the findings can prove difficult [16].

Among these difficulties, this report focuses on the analysis of evaluation data. By *analysis* we mean the process of turning the initial observations and hunches about difficulties into prioritized, coherent descriptions of usability problems, including a description of causes, implications and potential solutions to the problems.

1.2 Interviews on analysis in usability evaluation

We have conducted an interview study of how usability professionals think about and perform analysis of the evaluations that form part of their job. The study is based on phone interviews with 11 usability professionals in the industry; all from different companies and spanning six different countries. Each interview lasted 30-60 minutes, and was related to the participant's latest usability evaluation.

Five of the participants reported on analysis in usability inspection, and six on analysis in usability testing. We postulated that analysis may be similar across these two categories of usability evaluation, even though the type of data from which analysis proceeds clearly differs.

1.3 Motivation for the study

Little knowledge of current practice: The aim is to better understand what usability professionals do when analyzing usability evaluations. This understanding is lacking in the literature. Hardly any studies have been conducted on how usability professionals perform analysis. We know of only one study of think aloud in practice that explored how usability professionals perform analysis [15].

Also, the literature seems to be lacking in guidance on how to conduct analysis in usability evaluation. Where such analysis is described, it typically goes into detail only on the establishment of performance measurements such as task completion, task time, and satisfaction. Well known text books such as those by Dumas and Reddish [6] or Rubin and Chisnell [17] only make superficial descriptions of the analysis process required when going from the observation of individual participants in user tests to the presentation of a coherent set of usability problems to the client. Similarly, the Usability Body of Knowledge from Usability Professionals' Association (UPA) contains no entries on analysis (www.usabilitybok.org).

Basis for future research: Because we lack knowledge on current practice in analysis of usability evaluation data, we do not know whether current usability research serves the needs of usability professionals in this area. On the basis of our main findings, we therefore will suggest topics for future usability research.

1.4 Structure of the report

The report is structured as to make the most important findings easily accessible. First we present and discuss the main findings (Section 2). Second we present possible implications of our findings (Section 3). Third, a detailed overview of the results is provided (Section 4). Finally, the details of the method are presented (Section 5).

2 Main findings

2.1 Informal and pragmatic analysis

The participants reported to mainly conduct informal and pragmatic analysis of their usability evaluation data. The informality is reflected both in the note-taking, where structured forms and formal procedures for note-taking was practically absent, and in the subsequent analysis, where only two reported on means for structuring the analysis, such as the use of spreadsheet or tables.

Even though all reported user tests were video recorded, only one reported to have conducted a full analysis of these recordings. In the usability inspections, none of the participants followed any of the most well known standard methods for inspection.

The key analysis resource was the participants' professional experience. Explicated or general knowledge resources, such as established personas, design patterns, guidelines or standards, were hardly mentioned. Some mentioned dialogue with the designer or development team during analysis, which could be considered as an analysis resource; e.g. for prioritizing their findings or investigating the feasibility of redesign suggestions.

The majority of the participants did not report on using structured formats for problem descriptions and none reported on structured formats for redesign suggestions. Nor did any participants report on the use of specialized tools for analysis.

This lack of formality and structure may be interpreted as a pragmatic approach to analysis, where the aim is to identify relevant usability problems rather than establishing an as objective and correct set of usability problems as possible.

The pragmatic approach to analysis reported by the participants did not seem to be a consequence of atypical evaluation contexts given that only two reported that their evaluation was conducted under less than adequate time and resource constraints.

The detailed results used as basis for the above findings are provided in Sections 4.2 - 4.3.

2.2 Purposes of collaboration

The majority of the participants reported to be collaborating with colleagues or clients during analysis. Collaboration seemed to serve three purposes:

1. Identify more problems and generate better redesign suggestions
2. Quality control of the final analysis results
3. Improve reliability of the findings.

However, the first two purposes seemed more prevalent than the third. This is somewhat at odds with the recommendations of usability research, where the collaboration of analysts doing their analysis individually before any collaboration is recommended as a means to reduce evaluator idiosyncrasies and thereby improve reliability [9].

In the user tests, collaboration during analysis was either conducted initially, as a discussion between the test leader and the observers, or at the end of analysis, by having colleagues looking over the output of the analysis for commenting.

In the usability inspections, two reported on the collaboration being conducted as a group process; allowing for identifying more problems and generating better redesign suggestions.

The detailed results used as basis for the above findings are provided in Sections 4.4 - 4.5.

2.3 The importance of redesign suggestions

In most of the reported cases, the establishment of redesign suggestions was seen as an integrated part of the analysis. When encountering a usability problem, the analysts reported to try to identify and describe possible changes to the design that would alleviate the problem. At times, the redesign suggestion could be a suggested local fix to a problem, but the redesign suggestions could also imply larger changes.

In all the reported cases, the same analysts conducted both problem identification and redesign suggestions, and the redesign suggestions were typically established within the same analysis process as the usability problems, described as being developed in an iterative process of problem identification and redesign suggestions.

Some of the participants reported that redesign suggestions are developed in response to a coherent set of usability problems, others that these are developed as an immediate response to individual problems.

Some of the participants reported that an important motivation for them to develop redesign suggestion was their clients' responses. Redesign suggestions could make it easier to facilitate actual change in the product, and redesign suggestions could be the client's preferred outcome on an evaluation.

The detailed results used as basis for the above findings are provided in Sections 4.6 - 4.7.

2.4 Severity classifications

Severity classifications seem to represent a minimum structured format for problem description used by most of the analysts. However, important variation was found in the use of severity classifications. In particular, some participants classified their results according to urgency, that is, the degree to which the results requires immediate change. Others classified according to severity, reflecting identified usability problems' impact on the users' experience or behavior. In the research literature, usability is typically classified according to severity rather than urgency. However, as stated by Hertzum [9], severity assessments are also "recommendations about the urgency of fixing problems".

The use of urgency vs. severity classifications seems to illustrate different attitudes with respect to the analysts' role in the development process – as an active part of the redesign process or a passive provider of evaluations. An increased focus on redesign in the process of usability evaluation analysis may trigger the need to classifying according to urgency (related to the development process) rather than severity (related to the impact on user behavior/experience).

Severity classifications also seem to illustrate the interwoven relationship between problems and redesign suggestions. Both problems and redesign suggestions were classified in the majority of

the reported cases, and in about half the cases problems and suggestions were given the same classification.

The detailed results used as basis for the above findings are provided in Section 4.8.

3 Implications

3.1 Implications for usability research

Current usability research is to a great extent focuses on evaluation performance; that is, the degree to which usability evaluation methods, and by extension analysis, generate *valid*, and *reliable* output [8]. Improvements to evaluation validity and reliability are sought e.g. through structured analysis processes [2], the use of structured formats [4], and means to minimize the evaluator effect through the use of multiple independent analysts [9].

Also, a mainstream perspective within usability research is to view evaluation and redesign as two separate activities, where usability evaluation is not seen as an activity that should produce redesign suggestions [2]

Consequently, key themes of usability research seem not to be aligned with the way in which analysis is performed by usability professionals. In particular, research on how to support analysis as a pragmatic process returning a set of redesign suggestions is missing. This is critical, as usability research may therefore be seen as being irrelevant for professional practice.

Our study indicates that we need to develop research on three topics.

1. We see a need for supporting evaluation and redesign as an *integrated process*, possibly targeting evaluation as a first step in a creative process.
2. Researchers need to provide a better answer to the simple question of *what is analysis*. As noted at the beginning of this report, the literature is nearly silent on this issue. In particular, we need research-based development of analysis resources sufficiently effective and flexible so as to be useful for usability professionals.
3. *Collaboration* is confirmed as a key resource in both analyzing and generating recommendations from usability evaluations. However, collaboration may serve multiple purposes, not only minimizing the evaluator effect [9], and guidelines should be developed for collaboration to serve these multiple purposes.

3.2 Implications for professional practice

For professionals, we wish to summarize the implications of our study as three questions or areas of concern. The results indicate that industry usability evaluations may be prone to threats of thoroughness, validity, and reliability. In evaluations in which it is important to establish a comprehensive and valid set of usability problems, e.g. when developing security critical systems, usability professionals may be well advised to ask the following:

- A. How can I improve my analysis support to ensure that valid usability problems are reliably identified?
- B. Can fellow usability professionals be involved in the analysis so as to improve (a) validity and reliability (e.g. through individual analysis) and (b) creative exchange (e.g. through discussion)?
- C. Can it be useful to use structured formats for problem description and redesign suggestions, even though this may impose some limitation on analysis flexibility?

4 Detailed results

On basis of a thematic analysis, described in the method section, the interview data was studied according to nine themes. The themes are presented in the subsections below.

Themes	Corresponding high level issues (from Main findings)
0. The context of the reported evaluations	-
1. Formality and structure	Informal and pragmatic analysis
2. Analysis resources	
3. Collaboration	Purposes of collaboration
4. Evaluator bias and multiple analysts	
5. Problem identification and redesign suggestions	The importance of redesign suggestions
6. Sources of redesign suggestions	
7. Severity classifications	Severity classifications
8. Recommendations, challenges and needs	-

Table 1. The nine themes of the analysis

4.1 Theme 0: The context of the reported evaluations

The target evaluation method (user test or usability inspection) was often part of larger evaluation setup. Six reported combining the reported evaluation with evaluations done by other methods; typically combining user testing and usability inspection. One of these also reported on including user workshops. One reported on visits to customers to investigate the environment of use prior to the user tests. Four made no mention of associated usability activities (two of these reported on user testing only and two on usability inspection only).

The reported cases mainly seemed to be examples of evaluations conducted under typical resource and time constraints. Only two participants reported time and resource constraints as negatively affecting their choice of methods. Median time spent on analysis and report writing was 25 hours (range: 2-72 hours)

4.2 Theme 1: Formality and structure

The most prominent finding associated with this theme was a lack of formality and structure, both for the inspections and for the analyses of user test data.

Formality and structure in note-taking

Note-taking during user tests or inspections is an early opportunity to introduce formality and structure into the analysis. However, the participants' described note taking as mainly informal or pragmatic.

“We don't have a specific procedure. One of the persons has the task of taking notes so he writes down the most important things that happen.” (P2)

“Whenever the participant says something interesting I note it down, if I see a problem I remember to note it down and I also take down quotes” (P8)

A structured format for note-taking was used in only one of the user tests.

“Before the tests we made a protocol for each user where we printed the questions and then we had a table where we filled in certain criteria, if they managed to solve the problem, if they managed to solve the problem after some aid, and also [...] a special space to record which path they took, which clicks they made” (P4)

Formal or structured note-taking was not reported for the usability inspections, even though all but one took notes during the inspections.

Formality and structure in subsequent analysis

For most of the 11 participants, the subsequent analysis seemed equally informal or pragmatic. Only two participants reported using spreadsheets or tables to provide structure for their analysis. Although all reported user tests seem to have been video recorded, only one participant did a full analysis of these recordings. The others relied mainly on their notes and memory as background for their analysis.

“Often I just take one at a time, one set of a notes at a time, look at my notes and I can see this guy has a problem here, then I look at all the other ones and see who has the same problem, and if they have then I remember in my head, I can often remember (who had what problem), then I look into the notes and decide how serious this problem should be rated” (P5)

For six of the respondents, the analysis and report writing seemed to be intertwined; that is, the report was being written as part of the analysis process rather than after the establishment of a coherent problem set.

“I went through the notes one by one and whenever I found a new usability problem in the notes I wrote it into the final report.” (P3)

The pragmatic nature of the analysis is also illustrated by the participants’ choice of inspection methods, as well as in the absence of the use of structured formats and analysis tools. None of the usability inspections followed well known standard methods. Three had conducted an informal inspection based on the professional knowledge of the inspector; the two other combined scenario walkthroughs and heuristics, similar to heuristic walkthroughs [18].

The use of structured formats and specialized analysis tools / methods

Only one of the participants reported using a structured format for problem description. In the research literature, such formats for problem descriptions include e.g. that of Cockton, Woolrych, Hall, and Hindmarch [3] who presented a format including *“problem descriptions, the likely/actual difficulties, the context of a problem, and assumed causes.”* [11].

Seven participants did nothing that even resembled structured problem description. No one reported using structured formats for redesign suggestions. One of the participants even voiced a concern regarding the use of structured formats:

“[Structured formats] takes the flow out of performing the tasks, so it’s artificial, but it gets sort of like a checklist” (P10)

Apart from two participants, who reported using spreadsheets or tables as analysis tools, the only analysis tool reported was a template report used in a combined analysis and report writing process. However, seven participants reported screen shots as being important for their note-taking and/or presentation.

4.3 Theme 2: Analysis resources

By analysis resources, we mean knowledge on different areas of interest when conducting analysis, e.g. knowledge on users, the work-domain, technical environment, and design solutions. Analysis resources could e.g. include a set of personas (containing user knowledge), or a design pattern collection (containing design knowledge).

Apart from reports of the use of heuristics in usability inspections, the participants provided limited information about such analysis resources. It seemed that they mainly relied on their “*experience*” (P11) or “*professional background*” (P7).

Three reported the use of established heuristics, particularly Nielsen’s [14]. One reported the use of standards and best practices, and one reported the use of guidelines. Three also mentioned the benefits of having a dialogue with the designer or development team during analysis, which could be considered as an analysis resource. Participants considered this particularly relevant for prioritizing their findings or investigating the feasibility of redesign suggestions.

4.4 Theme 3: Collaboration

Collaboration in analyzing and reporting usability problems promises to increase reliability [9] and improving thoroughness [18] in usability evaluation, and is therefore recommended in HCI research. The background for the attention *reliability* has had within usability research is the findings that different evaluation teams investigating the same user interface find widely different usability problems [13]. Similarly, different analysts analyzing the same set of user test data also tend to identify different problem sets [12]. Such lack of reliability is problematic, if one perceives the aim of a usability evaluation to establish a problem set which is as close as possible to an ideal set of all existing usability problems for the given user interface.

In the user testing cases, collaboration was prevalent in the test sessions. However, collaboration during analysis was less common. None collaborated on a thorough analysis of the same set of observation data. However, four reported some kind of collaboration during analysis: two reported collaboration on the consolidation of findings after multiple analysts had analyzed separate parts of the dataset, one reported discussing the findings with a colleague, and one reported collaboration between test leader and observers.

In the usability inspection cases, three participants report collaborating on analysis. In only one case, however, did analysts work independently of each other.

4.5 Theme 4: Evaluator bias and multiple analysts

Theme 3 targeted the actual manifestations of collaboration on data analysis. Theme 4 targets the participants' motivations for such collaboration. As we saw above, seven participants reported some kind of collaboration on analysis. However, the motivation for conducting such collaborations seemed to vary. Three participants reported that collaboration was conducted in order to achieve different perspectives, identify more problems, and better redesign suggestions.

"It's normally an advantage to have several people look at it because you'll get a wider spread in the findings." (P9)

"You find more problems and good solutions when you are two or three people working together rather than by yourself" (6)

Two participants reported that collaborating helped consolidate the findings in an almost completed analysis.

"[I] asked my colleague to read this report and see if I had missed anything, and if I should add anything." (P4)

Finally, in two of the cases, multiple analysts seemed to have been involved to reduce evaluator bias. In one case, two independent analysts conducted usability inspection. In the other, collaboration between test leader and observers were reported to improve confidence in the findings, as stated below:

"After the tests we sit together and discuss what we have seen, so there is not only myself saying that problem occurred." (P2)

4.6 Theme 5: Problem identification & redesign suggestions

All participants except one reported that redesign suggestions are seen as an integrated part of the analysis. In all these cases, the same analysts did both problem identification and redesign suggestions. Redesign suggestions were typically reported together with their associated problems.

Disagreement existed on whether redesign suggestions are developed in response to a coherent set of usability problems (exemplified by P9, below), or as an immediate response to individual problems (exemplified by P7).

“We’ve focused on the usability problems first, then afterwards going over to redesign suggestions. In practice it’s hard to talk about problems without discussing solutions.” (P9)

“Every time I found a problem, I made a proposal how to solve it. This was a kind of iterative thing ...” (P7)

The prevalence of redesign suggestions seem, at least in part, to be motivated from the clients’ needs or preferences.

“[...] the customer liked that format where they had concrete advice on how to improve the user interface and also some positive comments about the expert evaluation, that led to us changing the format for presenting usability reports so we made a shorter report with giving greater advice for improving” (P4)

“It’s so much better if you give redesign solutions. If you don’t give the designers redesign solutions then nothing will happen.” (P6)

4.7 Theme 6: Sources of redesign suggestions

All participants reported relying on their own professional experience as a main source for redesign suggestions.

“[To reach redesign suggestions] we use our experience as a whole, from previous projects and solutions.” (P8)

Five participants also reported the use of general knowledge, including guidelines and standards. Three participants reported that the test participants’ comments could be important sources of redesign. The participants were not clear with respect to when in the analysis process redesign suggestions were made, but it seems likely (as was expressed by some) that this depends on the nature of the identified problem.

4.8 Theme 7: Severity classifications

All but two participants reported to have severity classifications. The remaining two made informal indications of problem severity.

There was variation with respect to the classifications used. In particular there was a spread across (A) severity categories and (B) urgency categories.

- A. Five used *severity categories*, that is, categories classifying according to the estimated impact on users behavior or experience. Three of these saw the severity classification as related both to the problem and the associated redesign suggestion, whereas two provided additional descriptions of the relative importance of the redesign suggestions.

B. Four used *urgency categories*, that is, categories classifying according to the analyst's recommendations on what to fix first. Two of these saw the urgency classification as related to both the problem and the associated redesign suggestion. One classified the redesign suggestions, rather than the problems.

4.9 Theme 8: Recommendations, challenges, and needs

The participants differed on their reported recommendations and challenges.

The only *recommendation* made by more than one was related to the use of screen shots and redesign visualizations in the report (stated by three).

The only *challenge* to be mentioned by more than one participant was related to documentation of the sources for the evaluation results; which basis the problems and recommendations had. It is interesting to note that problem merging and appropriate use of heuristics is only mentioned by one participant, in spite of this being a challenge receiving a lot of attention among HCI researchers (see e.g. [2]).

The needs reported, reflect two main issues. First, a need to bridge the gap between problems and design, e.g. as in finding good solutions for the problems, or as in identifying design issues being the root of many problems (serving as symptoms on the main design issue). Second, three of the participants voiced a wish for tools to speed up the analysis process, or make it more efficient.

4.10 Summary of results

In order to summarize some of the main findings in Chapter 4, we have summarized some of the salient features of the Themes 1-7 in Table 2. The purpose of the table is to provide a quick overview. We have extracted one feature from each of these seven themes, and we have enumerated the number of interview cases that do (not) correspond to the feature - or to a certain extent (somewhat).

Feature	No	Somewhat	Yes
High levels of formality and structure	9	2	0
Established references and stakeholders as analysis resources	3	4	3
Collaborative analysis with peers or client	4	4	3
Attention towards evaluator bias	7	2	2
Redesign proposals as an integral part of evaluation deliverable	1	0	10
Discussions with peers or clients as important sources of redesign	5	2	4
Severity classifications implemented	2	4	5

Table 2. Overview of salient features of the themes 1-7

5 Method

Due to the lack of previous research, the method needed to be exploratory and provide in-depth information. Thus, semi-structured interviews were chosen.

5.1 Participants

To avoid limiting the results to a particular industrial context, we sought to recruit participants from different companies and countries. Ninety industrial usability professionals were invited based on peer recommendations, additional invitations were sent to the mailing list of the UPA.

Twenty-three persons responded. Eleven of these fulfilled our requirement of having conducted a usability evaluation within the last month (designated P1 through P11 throughout the rest of the report).

Nine of the participants were consultants working for external clients in different companies. The participants were residing in six European countries: Denmark (3), Sweden (2), Finland (2), the Netherlands (2), Norway (1), Austria (1). The median work experience as usability professionals was 8 years (ranging from 1.5 to 35 years).

5.2 Procedure

The interviews were conducted by phone and concerned the latest usability evaluation the participant had conducted. Before the interview, participants identified their latest usability evaluation; during the interview, they were requested to have the report from that evaluation available as reference. Six of the evaluations were user tests, five were usability inspections. Median interview duration was 39.5 minutes (range: 30-57 minutes). The interviews were recorded and transcribed.

In order to counter possible biases related to interviewer expectations, we used an interviewer that was skilled in qualitative interviewing but only moderately knowledgeable in usability evaluation. The interviewer was not involved neither in the planning of the study nor in the analysis of the interview data.

The interview guide included questions and prompts so as to obtain information on the evaluation context, data collection and analysis, and analyst challenges and needs. Specific questions targeted structured formats, analysis methods and tools, collaboration, and redesign suggestions. The interview guide is available at http://docs.google.com/View?id=dhbq645m_0drskdf6.

5.3 Analysis

A thematic analysis [7] involving three analysts (the authors) was conducted. Specifically, the interview transcripts were segmented into units of meaning, each corresponding to a single idea and varying from a few words to an entire paragraph. A coding scheme consisting of nine themes was developed, following the analysts' individual reading of the transcripts. Both the subsequent thematic coding of the transcripts and the final analysis within each coded theme were done by one analyst, and reviewed in detail by the other two.

5.4 Limitations and future work

This study is a first attempt at collecting data on how professionals conduct analysis. Given the sample size and the nature of data collection, these results are preliminary, and large-scale surveys are important future research.

In conclusion, analysis of usability evaluations seems to be a process that we understand too little about, and for which research could work to provide better support.

References

1. Bias, R. and Mayhew, D. *Cost-Justifying Usability: An Update for the Internet Age*. Morgan Kaufman, San Fransisco, CA, 2005.
2. Cockton, G., Lavery, D. and Woolrych, A. Inspection-based Evaluations. In A. Sears and J. Jacko (eds), *The Human-Computer Interaction Handbook, 2nd edition*. CRC, 2008.

3. Cockton, G., Woolrych, A., Hall, L. and M. Hindmarch. Changing analysts' tunes: the surprising impact of a new instrument for usability inspection method assessment, *Proceedings of People and Computers XVII: Designing for Society*, Springer (2003), 145–162.
4. Cockton, G., Woolrych, A. and Hindmarch, M. Reconditioned merchandise: extended structured report formats in usability inspection. In *Proc. CHI 2004* (2004), 1433-1436.
5. Dumas, J. S. and Loring, B. A. *Moderating Usability Tests: Principles and Practices for Interacting*. Morgan Kaufman, 2008.
6. Dumas, J., and Redish, J. *A practical guide to usability testing*. intellect, 1999.
7. Ezzy, D. *Qualitative Analysis: Practice and Innovation*. Routledge, London, 2002.
8. Hartson, H. R., Andre, T. S. and Williges, R. C. Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, 13, 4 (2001), 373-410.
9. Hertzum, M. Problem Prioritization in Usability Evaluation: From Severity Assessments toward Impact on Design. *International Journal of Human-Computer Interaction*, 21, 2 (2006), 125-146.
10. Hertzum, M. and Jacobsen, N. E. The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, 13, 1 (2001), 421-443.
11. Hornbæk, K. and Frøkjær, E. Comparison of techniques for matching of usability problem descriptions. *Interacting with computers*, 20, 6 (2008), 505-514.
12. Jacobsen, N. E., Hertzum, M., & John, B. E. (1998). The evaluator effect in usability studies: Problem detection and severity judgments. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 1336–1340). Santa Monica, CA: Human Factors and Ergonomics Society.
13. Molich, R., Ede, M.R., Kaasgaard, K., Karyukin, B. (2004) Comparative usability Evaluation, *Behaviour & Information Technology*, 23(1), 65-74.
14. Nielsen, J. Heuristic evaluation. In J. Nielsen, & R.L. Mack (Eds.) *Usability Inspection Methods*, John Wiley & Sons, New York, NY, 1994.
15. Nørgaard, M. and Hornbæk, K. What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing. In *Proc. DIS 2006*. ACM Press (2006), 209-218.
16. Redish, J., Bias, R., Bailey, R., Molich, R., Dumas, J. and Spool, J. Usability in Practice: Formative Usability Evaluations - Evolution and Revolution. In *Proc. CHI 2002, ext. abstracts* (2002), 885-890.
17. Rubin, J. and Chisnell, D. *Handbook of usability testing (2nd.edition)*. Wiley Publishing, Indianapolis, IN, 2008.
18. Sears, A. Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human-Computer Interaction*, 9, 3 (1997), 213-234.