Analysis in usability evaluation

Findings from a questionnaire study with 155 usability practitioners

Authors
Asbjørn Følstad, SINTEF
Effie L.-C. Law, University of Leicester
Kasper Hornbæk, University of Copenhagen

SINTEF

KEYWORDS:
ICT, HCI, usability
evaluation, survey,
analysis practice

Report

Analysis in usability evaluation

Findings from a questionnaire study with 155 usability practitioners

VERSION 1		**DATE** 2012-04-16

AUTHORS
Asbjørn Følstad, SINTEF
Effie L.-C. Law, University of Leicester
Kasper Hornbæk, University of Copenhagen

CLIENTS
Norwegian research council, VERDIKT programme

CLIENT'S REF.
R2D2 Networks

PROJECT NO.
90B273

NUMBER OF PAGES/APPENDICES:
21pages + 3 appendices

ABSTRACT

Abstract heading

To improve current knowledge on analysis in usability evaluation, we present the findings of a survey study involving 155 usability practitioners who have reported on their latest usability evaluation. The study is based on a previous pilot interview study.

Analysis was found to be informal and pragmatic, conducted as part of fast-paced usability evaluations. Analysis structure and support from the usability research of the last 15 years hardly seem to have had any direct impact on analysis practices, even though general knowledge resources such as heuristics and guidelines are much used. Collaboration in analysis is common, and there is considerable awareness on collaboration as a means to improve reliability. Making redesign suggestions is a tightly integrated part of analysis, and also an aspect of evaluation that is seen as challenging. Severity classifications are commonly used, both for usability problems and redesign suggestions.

On basis of the findings we summarize six key implications for usability practitioners, to inspire future practice, and six implications for usability researchers, to guide future research.

The study is supported by the Norwegian research project R2D2 Networks (http://r2d2networks.origo.no) and the European COST project TwinTide (http://twintide.org).

PREPARED BY
Asbjørn Følstad

SIGNATURE

CHECKED BY
Jan Håvard Skjetne

SIGNATURE

APPROVED BY
Bjørn Skjellaug

SIGNATURE

REPORT NO.
SINTEF A22665

ISBN
978-82-14-05287-9

CLASSIFICATION
Unrestricted

CLASSIFICATION THIS PAGE
Unrestricted

Document history

VERSION	DATE	VERSION DESCRIPTION
1	2012-04-16	Final version

Table of contents

1 Introduction

Analysis is an important part of usability evaluation. Doing analysis is to turn the data from usability testing, or initial insights or hunches in usability inspections, into coherent descriptions of usability problems; descriptions that include assumed causes and implications and possibly also change suggestions (Følstad, Law & Hornbæk, 2010).

Analysis is challenging as it is likely to require interpretation of multiple quantitative and qualitative data sources. The data has to be interpreted relative both to the intended use of the system under evaluation, general usability knowledge, and the personal experience from previous usability evaluations. The challenging nature of analysis made by Rubin and Chisnell (2008) describe it as "the ultimate detective work" in usability testing. Similarly, Cockton, Lavery and Woolrych (2008) highlighted up-to-date knowledge and expert competency as critical for the successful application of usability inspection methods.

Even though the difficulties associated with analysis are acknowledged in the literature, we have little knowledge on how usability practitioners actually go about doing this part of usability evaluation. Introductory material on usability evaluation provides only high-level advice on analysis. For example, in two well known text-books on usability evaluation, Rubin and Chisnell's (2008) *Handbook of usability testing* and Dumas and Redish' (1999) *A practical guide to usability testing*, less than 8% is devoted to analysis (in comparison more than 40% of either book concerns planning and preparing the usability test). Also, previous research only provides indirect data on current practices of doing analysis in practical usability evaluation (Nørgaard & Hornbæk, 2008). This lack of knowledge has, however, not barred usability researchers from developing methods and tools to support such analysis, including process improvements (Andre, Hartson & Williges, 2001; Kjeldskov, Skov & Stage, 2004; Cockton & Lavery. 1999), problem description formats (Cockton, Woolrych & Hindmarch, 2004; Capra, 2006; Howarth, Smith-Jackson & Hartson, 2009), and tools for problem identification and consolidation (Andre, Hartson & Williges, 2003, Skov & Stage, 2003).

To provide knowledge on how usability practitioners do analysis in usability evaluation, we have conducted a survey study where 155 usability practitioners reported on their latest usability evaluation. Of these, 112 reported on a usability test, 43 reported on a usability inspection. Details on the respondents and the method are presented in section 9.

In this report we present the findings from the survey study. We believe that the findings are useful both for usability researchers, as they may inform future research on methods and tools to support analysis, and practitioners, as they provide insight in how other practitioners do analysis and thereby may serve as a source of inspiration.

The structure of the report is as follows. First we make a short recap of out findings in a previous pilot interview study (section 2). Then we present the main findings of the survey study (sections 3-7). We also suggest a set of implications for practitioners as well as future research (section 8). Finally, we provide a detailed presentation of the method used in the survey study, including background data on the survey participants (section 9). Appendix 1 includes detail presentations on the result of all quantitative questions, except items on background data and free-text items. The questionnaires are provided in Appendix 2 and 3.

2 A recap of the findings from a previous pilot interview study

Prior to the survey study, we conducted an interview study with 11 usability practitioners to get an initial understanding of how analysis is conducted. Also, the interview study informed us on which aspects of analysis to pursue in a survey study. The pilot study is presented in full in a previous report (Følstad, Law & Hornbæk, 2010). Below we summarize the main findings as background for the present report.

Analysis is informal and pragmatic
In the pilot interview study, the respondents mainly reported informal and pragmatic analysis of their usability evaluation data. The main analysis resource was the respondents' own professional experience. Most did not use structured forms or formats for note taking or problem description, hardly any conducted full analysis of usability testing videos, and few mentioned explicit use of general knowledge resources such as design patterns, guidelines or standards. In the survey study we found analysis to be equally informal and pragmatic as in the interview study. However, general knowledge resources were found to be more used than we concluded on basis of the interviews.

Collaboration in analysis is done to identify more problems and generate better redesign suggestions
Collaboration with colleagues during analysis was reported by most pilot interview respondents. The collaboration mainly was meant to identify more problems and generate better redesign suggestions or to serve as a quality assurance, that is, a check of the final analysis results. Collaboration was only to a smaller degree reported as a means to improve reliability, in particular by independent analyses of the same data set, which is somewhat at odds with the recommendations of usability research. The survey study sustained our conclusion on the frequent use of collaboration, but disconfirmed our conclusions on the motivation for collaboration. Quite to the contrary of our conclusions from the pilot interview, the survey respondents reported improved reliability as their main motivation for collaboration.

Redesign suggestions are an important outcome of analysis
Generating redesign suggestions were for most of the participants an integrated part of analysis. Typically the analyst both identified the usability problems and made change suggestions. Some reported to provide redesign suggestions in response to a complete list of usability problems, others reported to generate redesign suggestions immediately upon identifying a usability problem. These pilot interview conclusions were sustained by the survey finding.

A division concerning the use of severity vs. urgency classifications
The pilot interview participants were divided in their use of severity classifications. Some used severity classifications denoting the impact of the usability problem on the participants' experience or behaviour. Others classified the urgency with which they recommended the problems to be fixed. Severity or urgency classifications were provided for both usability problems and redesign suggestions. In the survey study we also found some use of urgency classifications, however not to a large degree.

Our aim for the survey study was to explore the generality of the conclusions of the interview study. In particular we wanted to make further exploration of the structure and support utilized during analysis, how collaboration in analysis is motivated and implemented, the integration of evaluation and redesign, as well as the use of severity ratings.

3 Evaluation methods and analysis contexts

Before going into details on our findings on how analysis is conducted in practical usability evaluation, we present some of the background data on the context of analysis and the methods used in the evaluation on which the respondents reported. We make this presentation separately for usability testing and usability inspection. As background we also include the respondents' perceptions of challenges and impact of their latest usability evaluation.

3.1 Usability testing

Most respondents reported on usability tests with between 5 and 12 user participants (25th percentile = 5; median = 8; 75th percentile = 12). This participant volume is in line with the recommendations of Dumas and Fox (2008) and suggest that the tests typically were formative; that is, that the tests conducted to identify usability problems as part of an iterative development process.

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

5 of 59

The reported time spent on the usability tests indicates a fast pace. Median working time for a complete usability test was 48 hours (25th percentile = 24 hours; 75th percentile = 80 hours). The vast majority of the respondents (81%) reported that these were typical time constraints.

The usability tests were distributed across the entire development process. However, the most frequent development phase was *Advanced prototypes* (43%), followed by *Running systems* (30%) and *Early prototypes* (23%). Only 5% of the usability test participants reported on a usability test in the *concept* phase.

We also asked the participants to report on their use of usability measures. Most of the usability tests included measures of *task completion* (84%) and *satisfaction* (80%). Less frequently used measures were *error rate* (45%) and *task time* (33%).

3.2 Usability inspection

Usability inspection was conducted by many different inspection methods. Less than half the respondents reported to use one of the classical inspection methods, *Heuristic evaluation* (30%) or *Cognitive walkthrough* (14%). 47% reported less formal methods, classified as *informal expert review*, *mix of methods* or *no particular method*; the remaining 9% reported to have used *other methods*.

Thirty-four of the respondents gave descriptions of their inspection method in free-text comment fields. The descriptions clearly indicate flexible use of the methods. In particular, heuristic evaluation was often combined with elements from other methods, or heuristics were integrated in other inspection methods. Method combinations included heuristics + walkthrough, heuristics + best practice, and heuristics + some form of expert review.

The reported time for the inspections indicates that speedy evaluations are the norm. Median working time for an inspection was 24 hours (25th percentile = 10; 75th percentile = 80).

Usability inspections, just as usability testing, were conducted in all development phases. However, to our surprise, about half the inspections (49%) were conducted on *running systems*. Aside from this, usability inspection was somewhat skewed towards the earlier development phases; 19% of the inspections were on *concepts*, 21% were on *early prototypes* and only 12% were on *advanced prototypes*. We assume that the high frequency of inspections on running systems is due to these being conducted as pilot evaluations, either at the start-up of a redesign process or to assess whether a more thorough usability evaluation is needed.

3.3 Perceptions of challenges and impact

3.3.1 Challenges in usability evaluation

We asked the respondents to assess five aspects of their latest usability evaluation with respect to how challenging they were perceived to be. The following scale was used: Very little challenging, little challenging, somewhat challenging, challenging, very challenging, and N/A - the latter recoded as missing.

The reason for asking this question was to see if some aspects of evaluation are more in need of new support than others. The proportion of respondents reporting an aspect as *somewhat challenging* or more is presented in Figure 1.
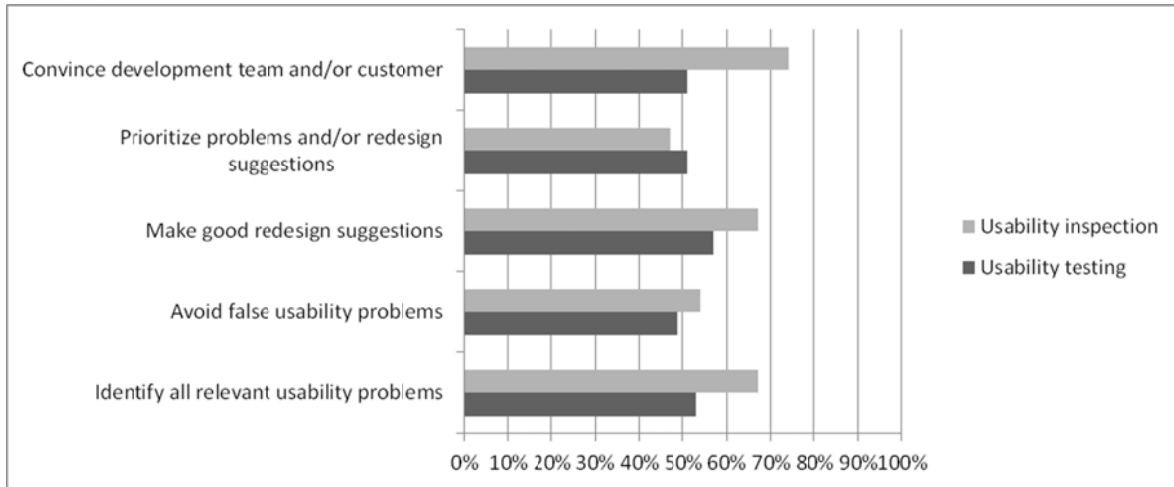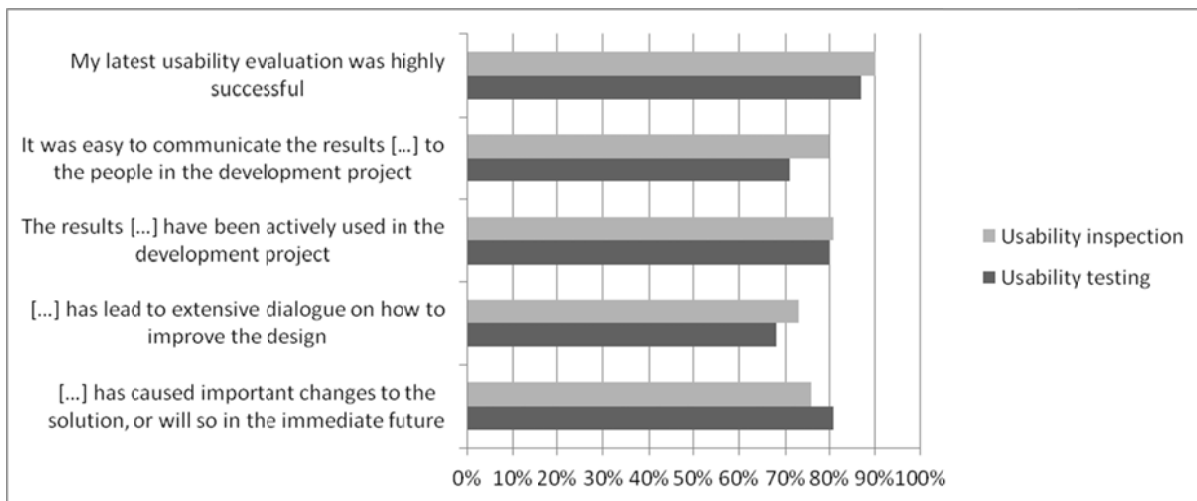
Figure 1: Proportion of respondents answering *somewhat challenging* or more on how challenging they found different aspects of their latest usability evaluation to be.

Interestingly, the aspect that was found to be challenging for the highest number of respondents was *to make good redesign suggestions* (60%). Also, we note that the usability inspection respondents found it far more challenging *to convince the team and/or customer* than did the usability testing respondents.

3.3.2 Perceptions of impact

We also asked the respondents to answer five questions thought to be related to the impact of their latest usability evaluation. The questions were answered on a five item scale from *strongly disagree* to *strongly agree*, with an option to respond N/A (coded as missing). The results for this question are presented in Figure 2.

Figure 2: Proportion of respondents answering *agree* or *strongly agree* on the five questions concerning the perceived impact of their latest usability evaluation.

We see that the vast majority of the respondents perceived their latest usability evaluation as a success, and also tended to be highly positive concerning the other four questions.

We intended to use the five questions as a scale to measure perceived impact of a usability evaluation by setting *strongly disagree* = 1 and *strongly agree* = 5. The scores on the five questions were found to load on one general factor (analysed on basis of a principal components analysis), and the inter-item reliability was found to be acceptable (Cronbachs α=0.74). However, due to the high levels of agreement observed for each item, the mean for the scale measure was fairly high and the variance fairly low (*mean* = 4, *SD* = 0.6). Low scale variance may be problematic, as this will impact its ability to discriminate between individuals (DeVellis, 2003); that is, the scale may be fairly insensitive.

3.4 Conclusion on evaluation methods and analysis context

The respondents typically reported their latest evaluation to have significant impact and to be highly successful. The usability evaluations included in the survey were typically fast-paced. Half the reported usability tests were conducted in 40 working hours or less. One fourth of the usability inspections were conducted in 10 hours or less.

This means that analysis needs to be highly time efficient to fit the typical time constraints for usability evaluations. By extension, structure and support for analysis will likewise have to be fast and efficient to use, to support usability evaluations as they are currently practiced. Making good redesign suggestions is perceived as challenging.

4 Analysis structure and support

How, then, is analysis conducted in the context of fast-paced usability evaluations? We approached this question from four angles: Which strategies are used to identify usability problems? How are note taking and usability problem description supported? What kind of special purpose tool support is used? And which knowledge resources are employed?

4.1 Strategies for usability problem identification

We asked the respondents to provide *a brief description of how you decided something being a usability problem.* In total 125 respondents gave such descriptions. These were itemized and coded in a thematic analysis (Ezzy, 2002). In total, 275 items were coded. Summaries of the findings are presented below, for usability testing and usability inspection separately.

4.1.1 Strategies in usability testing

The usability test respondents reported four high-level strategies for identifying usability problems: To look for (a) consequences for task performance, (b) user responses aside from task performance, (c) problem criticality on basis of severity and frequency, and (d) causal explanations. In addition to these four, some respondents mentioned the importance of usability expertise and professional experience. Other mentioned strategies were to use rating scales, web analytics, and predefined success criteria. Details are provided in Table 1.

Consequences for performance and users' responses: About one third (32%) of the items from the usability testing respondents concerned judgements on observable task performance, whereas one fifth (19%) concerned user participants' responses irrespective of task performance. The latter included emotional responses (such as boredom and frustration), behavioural responses (such as hesitation and seeking help), and verbal data (from the think aloud protocol). The distribution of these two high level strategies suggest that a substantial part of usability practitioners, but not all, see it as valuable to augment their analysis on basis of task performance with insights from users' responses. Possibly, the importance of users' responses depends on the application domain. For example, for entertainment applications such as games, where user experience is important, observations not directly linked to task performance may be seen as more relevant. For work support systems on the other hand, observations not directly related to task performance may be given less weight. This, however, we do not know as we did not collect data on application domain.

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

8 of 59

High level strategy	Strategy	Percent
Consequences for task performance	Task (in)completion	15%
	Task completion time	6%
	Other consequences	11%
Users' responses aside from of task performance	Emotional responses	8%
	Behavioral responses	4%
	Verbal responses	7%
Criticality	Frequency	13%
	Severity	5%
Causal explanations	Difficulty in finding	7%
	Difficulty in understanding	7%
	Mismatch with users' understanding	5%
Other	Usability expertise	8%
	Other (incl. rating scales, web-analytics, success criteria)	5%

Table 1: Usability problem identification strategies in usability testing

Usability problem criticality: Data supporting judgements on the criticality of usability problems were mentioned in 18% of the items. Interestingly, problem severity was mentioned far less than problem frequency. This may imply that the number of participants being observed to have a particular problem may be more important for determining something to be a usability problem than the severity of the observed problem instances. Observed severity, on the other hand, may be important for prioritizing once the observed incident(s) has been established as a usability problem.

Causal explanations: Finally, one fifth (19%) reported the identification of a causal explanation to be an important part of problem identification. Causal explanations typically were reported as something being difficult to find or understand, or a mismatch between the users understanding and the interactive system. The reported attentiveness to causal explanations is in line with Rubin and Chisnell's (2008) recommendation to conduct a *source of error analysis* during analysis.

4.1.2 Strategies in usability inspection

The strategies for problem identification employed in usability inspection differ markedly from the strategies employed in usability testing. As seen in Table 2, the most frequently reported strategy was to rely on professional experience and knowledge resources, reported in half the items. Predicting consequences for users were reported in one fifth (19%) of the items, whereas looking for causal explanations were reported in only 7% of the items.

High-level strategy	Strategy	Percent
Experience and knowledge resources	Professional experience/Expert knowledge and mindset	30%
	Usability resources	20%
Predicting consequences for users	Consequences for task performance	11%
	Consequences for users' responses	3%
	Taking the user's perspective	5%
Causal explanations	Causal explanations	7%
Other	Other approaches	18%
	Issue / incomprehensible	6%

Table 2: Usability problem identification strategies in usability inspection

Experience, expertise and knowledge resources: The importance of professional experience and expert knowledge and mind-set was far more prevalent for problem identification in usability inspection than in usability testing. Also, not surprising, usability resources, such as guidelines, heuristics, and design patterns, were more often mentioned by usability inspection respondents than usability testing respondents. The frequent reliance on experience and professional knowledge as a problem identification strategy is partially in line with Cockton, Lavery and Woolrych (2008) who argue that the successful use of usability inspection methods depends on analyst knowledge. However, these researchers also argue for the importance of structured processes for analysis. The prevalence of the reliance on professional knowledge – in particular that this is mentioned more often than general usability resources – may indicate that for some analysts professional experience, not general usability resources, is *the* key to successful usability problem identification. Possibly, the need for time-efficiency in usability inspection may be one reason for the reported reliance on experience and expert knowledge.

Predicting consequences for users: Predicting users' task performance and responses, or taking the user's perspective, was not frequently mentioned. This is not to say that predicting consequences for the users are not important in usability inspection, as usability problems identified by heuristics or guidelines will most likely have consequences for users. However, the little mention of predicting consequences for users as a strategy for identifying usability problems may indicated that problems more often are identified on basis of experience and knowledge resources rather than on basis of simulations of users doing tasks.

Causal explanations were mentioned less frequently for usability inspection respondents than for usability testing respondents. Possibly, references to violations of existing usability knowledge may serve as a sufficient explanation – so that providing additional causal explanations are seen as redundant.

4.2 Structure in note taking and usability problem description

We asked the respondents about their note taking practices, as well as their use of specific formats for usability problem description.

4.2.1 Note taking

All respondents reported to take notes during analysis. However, only about half reported to structured their notes by some kind of forms (usability testing: 40%; usability inspection: 57%). Less than one fifth reported to use *detailed forms* (usability testing: 18%; usability inspection: 15%). The remainder used *simple forms,* typically structured according to general topics/questions, tasks from scenarios of use, test script/protocol, or as checklists.

Although note taking is universal, the use of forms as a way of structuring notes is not. And in particular, the use of forms that provide a high degree of structure. Possibly, the variation in issues that needs to be noted down during usability evaluation is too rich as to be easily fitted to a rigid note taking structure.

4.2.2 Usability problem description formats

Usability problem description formats have been established in previous research, to facilitate usability problem merging and reporting. Lavery, Cockton, and Atkinson (1997) described a high-level template for reporting usability problems, Capra (2006) provided a set of guidelines for describing usability problems, and Howarth et al. (2009) presented problem descriptions formats integrated in usability evaluation analysis software.

Most of the respondents (59%) reported to use a structured format for usability problem description. The remainder reported that *the problems were described in plain prose*. However, of those using structured formats, nearly all reported to use *our own format* (55% of the total respondents); next to none (4% of the total respondents) reported to use *a structured format described in the standards or literature*. This finding is

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

10 of 59

quite astonishing as it indicates that there hardly exists a common practice for structured usability problem description – even though usability problems is a key outcomes of a usability evaluation.

The respondents also reported on a usability problem description practice that is not well covered in the literature: The use of *screen-shots to illustrate usability problem descriptions*. The vast majority of the respondents reported to use screen-shots for this purpose; only 24% reported not to use this. Possibly, the literature on usability problem description should be updated to reflect this practice.

4.3 Tool support in analysis

Usability research has also generated tools to support analysis in evaluation; both software tools, such as Andre et al.'s (2003) Usability problem inspector, and conceptual tools, such as Skov and Stage's (2005) tool for problem identification and prioritizing. Also, usability research has provided supporting models and frameworks for analysis, such as Cockton and Lavery's (1999) framework for structured problem extraction (SUPEX).

We asked the participants *which tools, if any, did [they] use for usability problem description and analysis* in their latest usability evaluation. 82 persons responded to this question with a free-text answer; 37 of these were disregarded as they either answered blank or explained that no tools had been used. Concerning the 73 respondents that skipped this question, we assume that their main reason for skipping was that they did not have any tools to report, or that they were uncertain on what was included in the term *tools*. In the analysis of the free-text answers we were looking for special purpose analysis tools, and consequently disregarded reports of general tools for text editing, spread sheets and presentation.

Special purpose analysis tools that were reported by more than one respondent are presented in Table 3.

Tools	Usability testing	Usability inspection
Screen recording and analysis software, such as Morae and Silverback	11	0
Screen recording tools, such as Camtasia and SnagIt	5	0
Eye tracking tools	2	0
Drawing and prototyping tools, such as Balsamiq, Axure, Visio, and Photoshop	2	6
Web analytics solutions, such as Google analytics and Seevolution	0	5

Table 3: Special purpose analysis tools with associated frequency by which they were reported.

It is interesting to note that drawing and prototyping tools are used by some as part of analysis, in particular for usability inspection. It is also interesting to note that tools for web analytics are used to inform analysis. Tools and other analysis support from usability research were hardly mentioned.

Elsewhere in the questionnaire, we asked the usability testing respondents on their use of video recordings from the test sessions. 73% made such recordings, 34% reviewed these recordings in part, and 19% reviewed them in full. This finding is somewhat in contrast to only 11 respondents reporting tools for screen recording

and analysis. Possibly, this reflects that such tools are used for recording and viewing rather than facilities for analysis.

4.4 Knowledge resources in analysis

The respondents were asked which knowledge resources, from a list, they *did make explicit use of during analysis* in their latest usability evaluation. The knowledge resource *my professional experience* was reported by nearly all (91%). However, only 9% reported their professional experience as their only knowledge resource.

Aside from professional experience, *heuristics and guidelines* were reported by the majority of the participants. Interestingly, this knowledge resource was reported both by the majority of usability testing respondents (60%) and usability inspection respondents (76%). Possibly, heuristics and guidelines are used in analysis of usability testing results to support a source of error analysis; that is, the heuristic or guideline may provide explanations for observed usability problems.

Also *design patterns* and *standards* were reported as a knowledge resource used by almost half the respondents. As for heuristics and guidelines, these knowledge resources were frequently used both in usability testing (41% and 40%) and usability inspection (54% and 42%).

A less frequent knowledge resource was *Previously established personas*, used by 18% of usability testing respondents and 27% of usability inspection respondents. 4% of the respondents reported *Other* resources.

4.5 Conclusion on analysis structure and support

The fast-paced and flexible evaluation methods described in section 3 by necessity require fast-paced and flexible analysis. Probably, the respondents' use of simple forms and home-grown problem description formats to support analysis should be seen in this light.

Concerning structure and tools, two findings are particularly interesting. First, special purpose tools for analysis do not seem to be in very widespread use for analysis purposes. Even though most respondents reported to make video recordings of their sessions, relatively few reported to use software for screen recording and analysis as an analysis tool. It may be that this discrepancy is due to practitioners not using the analysis features in such tools, but rather use them as advanced video recorders and -editors.

Second, the respondents did not report any use of tools and other aids from usability research. This finding is surprising, given the effort that has gone into the research-based development of such. Possibly, this lack of use may reflect an awareness issue, that is, practitioners may just be unaware that such research results exist. However, it may also indicate that research-based tools and aid for analysis is difficult to introduce in fast-paced evaluations. If this latter explanation holds, usability research on analysis facilities may need to change direction in order to better serve the community of practitioners.

It is also noteworthy that even though strategies for usability problem identification differ between usability testing and usability inspection, there seem to be much overlap concerning the knowledge resources used in analysis. In particular, it was surprising that heuristics and guidelines, as well as design patterns and standards, were reported to be explicitly used to such a degree during analysis also in usability tests.

In section 3 we saw that practitioners seem to flexibly use evaluation methods as components to be combined in response to a given evaluation context rather than recipes to be strictly adhered to. This finding is in line with Woolrych, Hornbæk, Frøkjær, and Cockton (2011) who argue that usability research should *"not treat evaluation methods as indivisible wholes"* (ibid), but allow for method components, such as procedures for task walkthrough, principles for participant recruitment, and procedures for problem merging, to be flexibly

used across evaluation methods. Likewise, analysis resources and practices may be seen as components that may be shared across evaluation method instances.

It should also be noted that the frequent use of general knowledge resources represent a deviation from the findings in our pilot interview study; general knowledge resources seem to be more generally employed than out interview study indicated.

5 Collaboration in analysis

In our pilot interview study, we found that collaboration during analysis was frequent and mainly conducted to identify more usability problems and improve the quality of redesign suggestions. The survey study sustained our conclusion that collaboration is frequent, as only 26% reported *no collaboration*. However, the conclusion that collaboration is mainly conducted to identify more usability problems and improve the quality of redesign suggestions did not hold.

Respondents who reported to engage in collaboration during analysis were asked to state the main purpose of this collaboration. The reported purposes are presented in Figure 3. Collaboration was only infrequently motivated by an aim to identify more problems. Also, collaboration to generate better redesign suggestions was less frequent than suggested in the pilot interview study.
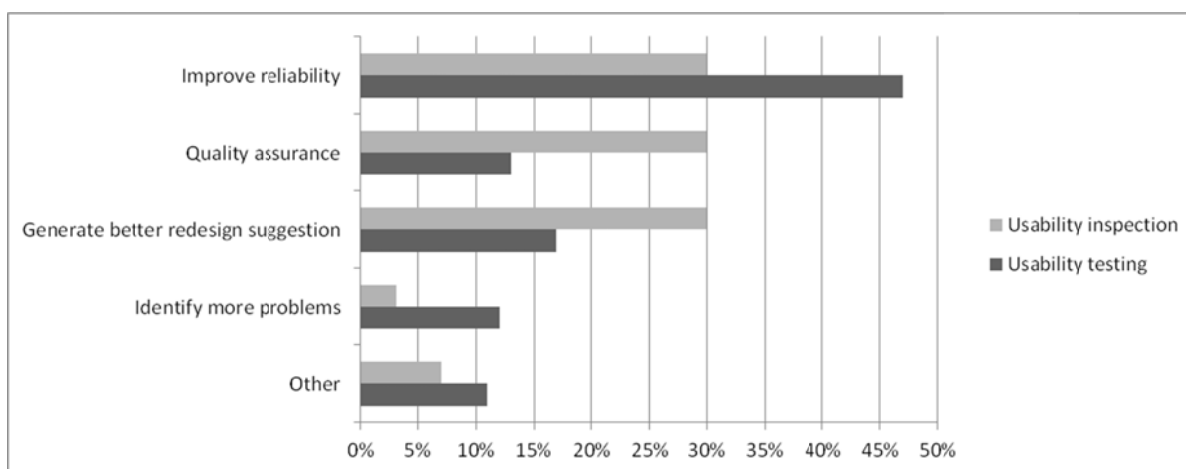
Figure 3: Collaboration purposes with percentages of respondents reporting each purpose.

The most important purpose of collaboration is to *improve reliability* / *avoid that the findings were biased by personal perspective*; about one third reported this as their main purpose. However, this purpose may not be in line with the actual collaboration that the same respondents reported.

We know from usability research that reliability in analysis is challenging. Hertzum and Jackobsen (2003) termed evaluators tendency to interpret the same usability data differently as an evaluator effect; an effect that is significant even for parallel analysis of the same set of recorded usability test sessions. The prescribed safeguard against the evaluator effect is to have several analysts do independent analyses of the same data set. However, when we asked the respondents *what kind of collaboration did you have with other usability professionals during the analysis of your latest usability test/inspection*, collaboration involving *independent analysis of data from the same user sessions / the same parts of the system* were reported by only 41% of those respondents that claimed *improved reliability* to be their main purpose of collaboration. This finding may indicate that some practitioners are not sufficiently aware of the challenges associated with reliability in analysis.

Another type of collaboration was analysis *conducted as **a group activity** by two or more usability professionals*. Analysis in groups was reported by 29% of the usability testing respondents and 30% of the usability inspection respondents. Analysis in groups is to be expected in usability inspection, as this is recommended in the literature on heuristic evaluation (Nielsen, 1994) and cognitive walkthrough (Wharton, Rieman, Lewis, & Polson, 1994). However, analysis in groups for usability testing is not equally well described in the literature. Consequently, there may be a need both for guidelines for analysis in groups in usability testing as well as research studies on the benefits and limitations of such analysis.

6   Redesign in analysis

Our pilot interviews indicated that redesign suggestions is an important and integrated part of analysis. This finding was confirmed by the survey results. Nearly all respondents (96%) reported to include redesign suggestions in their evaluation deliverables. And half (51%) characterized their deliverable as *a set of redesign suggestions* either *in response to a set of usability problems* or *in part motivated from usability problems*.

Redesign suggestions are provided in many different formats. We asked the respondents to tell us how they presented redesign suggestions in the deliverable. In Figure 4 we see that even through textual descriptions dominate, about half the respondents used annotated screen shots and more than one third provided digital mock-ups of alternative user interfaces.
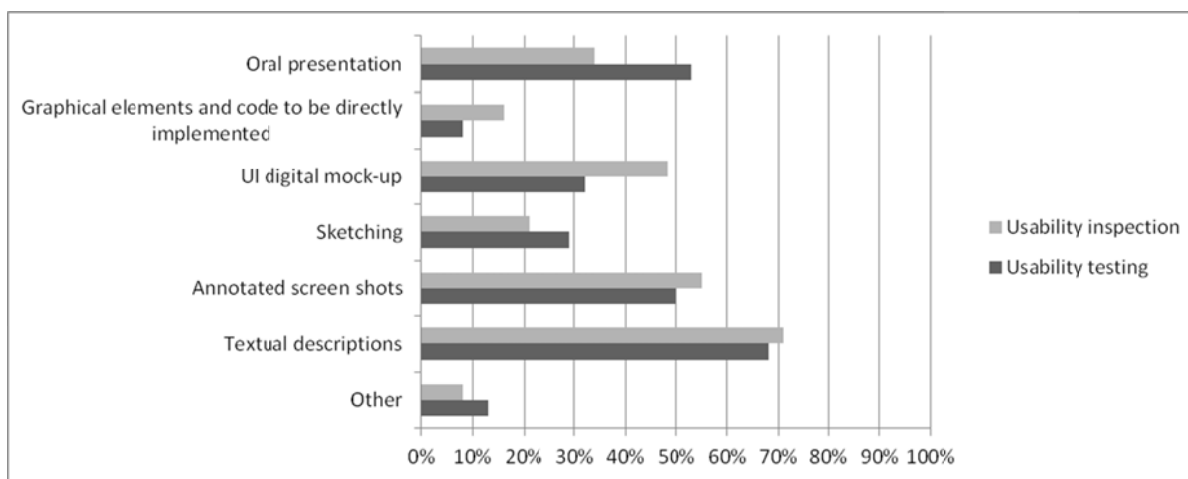
Figure 4: Formats for presenting redesign suggestions

But how do the practitioners reach redesign suggestions? In one part of the questionnaire we asked the usability testing respondents on whether they used their test participants as sources to usability problems and redesign proposals (Q15); 64% asked their test participants *for their opinion on possible usability problems*, and 48% asked them for *possible redesign suggestions*. (We also asked for the use of usability test observers for this purpose. The proportions of respondents who had asked their observers for opinions on problems and redesign suggestions were 48% and 53% respectively. However, as we did not ask the respondents on the presence of observers, we do not know if this percentage is lower than it would have been if we had asked only those that actually had an observer present.)

In another part of the questionnaire (Q27/Q24), we listed several possible avenues to redesign suggestions and asked the respondents to report how they did *reach the redesign suggestions that were included in the deliverable*. Details are provided in Figure 5.
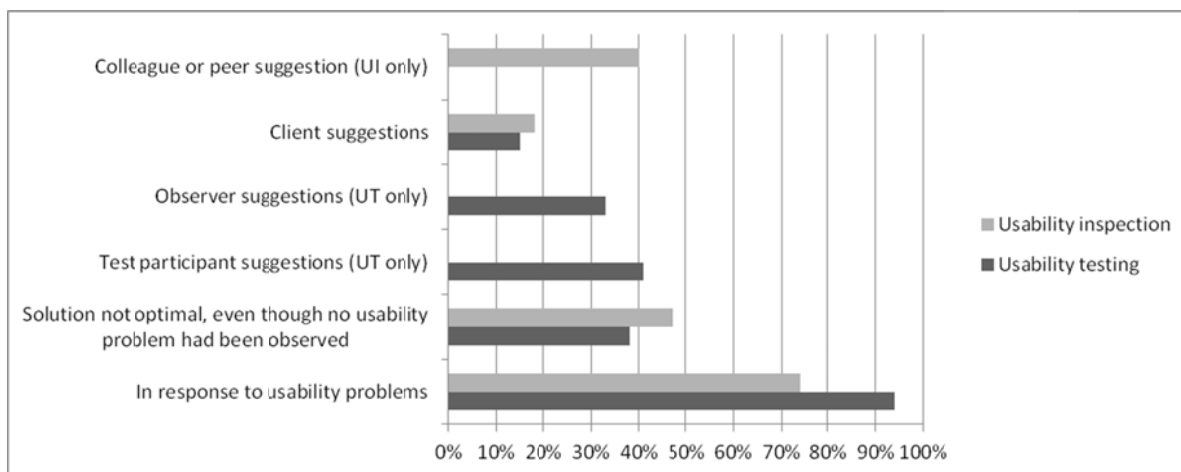
Figure 5: Responses to how the respondents had reached redesign suggestions (UI = usability inspection, UT = usability testing)

We see that the most important approach to redesign suggestions is to make these in response to observed usability problems. However, this is by no means the only road to redesign suggestions. More than one third reported to make redesign suggestions on basis of their finding that the target of the evaluation was not optimal – even though no usability problem had been observed. Further, test participants' and observers' opinion also seems to be important when making redesign suggestions.

We also asked the respondents when they did *make the redesign suggestions*. Interestingly, only about half the respondents (usability testing: 54%; usability inspection: 40%) reported to make redesign suggestions only after all usability problems had been identified. Most of the remaining respondents reported that some redesigns were made immediately and some were made after all problems had been identified.

Clearly redesign is a key part of analysis. Usability practitioners apply multiple strategies to generate redesign suggestions, and redesign suggestions can be made even though no usability problem has been observed – also in usability testing. Practitioners, during analysis in usability testing, use usability knowledge and design intuition to suggest improvements also for aspects of the user interface that has not been covered in usability test scenarios. Redesign suggestions may be made at any time in the analysis process. However, we assume that the timing of redesign suggestions depends on the usability problem in question. Immediate redesign suggestions should hardly be made for highly complex or deep-rooted usability problems, but rather for quick fixes.

7 Severity ratings

In our pilot interviews we were intrigued by some respondents reporting on urgency ratings rather than traditional severity ratings; the latter reflecting the estimated impact of a usability problem on a user. We followed this lead in the survey study and found that urgency ratings are indeed employed, but only by 20% of those that prioritizes their findings.

The majority of the respondents reported to prioritize their findings. In particular for usability testing prioritizing is abundant; 82% reported to prioritize their findings and half of these prioritize both usability problems and redesign suggestions. Among the usability inspection respondents, 67% reported to prioritize their findings; about two thirds of these prioritized both usability problems and redesign suggestions.

To conclude: The traditional approach to *severity ratings*, that is, to prioritize findings according to their predicted impact on user behaviour, is most common. Some practitioners, however, use *urgency ratings*, that

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

15 of 59

is, classifications reflecting how important it is to make a particular design change, but not as many as we suspected on basis of the pilot interview study. A more common practice identified in the survey study, however, was that practitioners tend to prioritize both usability problem descriptions as well as redesign suggestions. This, in turn, serves to underscore the integration of evaluation and design that we saw in section 6.

8 Implications

On basis of our findings we suggest implications for practitioners as well as researchers of usability evaluation. We intend the implications for practitioners to serve as suggestions and inspiration for analysis in practical usability evaluation. The implications for research are intended as guidance for future research on methods and tools to support analysis.

8.1 Implications for practitioners

The implications for practitioners are drawn directly from the findings. That is, we assume that emerging or common practices are beneficial – hence their popularity. Only at one point do we criticize current practice on basis of current research knowledge, and that is the last implication concerning collaboration and reliability.

Engage in master-apprentice relationships: Informal and fast-paced analysis clearly implies the importance of expert knowledge and skill. However, as the current literature only provides high-level descriptions of analysis, such expertise and skill will have to be transferred in master-apprentice relationships rather than through explicit knowledge resources such as text-books. By engaging in master-apprentice relationships, novices can to a greater degree avoid to learn by trial-and-error, something that in turn will improve usability evaluation practices.

Consider checklists: A particular threat of an informal analysis procedure is that it is easy to forget important parts of the analysis. A few of the respondents reported to use checklists during note taking (section 4.2.1). Checklists have been proven highly valuable to avoid costly mistakes in routine processes (Gawande, 2009). Quite possibly, analysis practices will benefit from the use of checklists – either home-grown on an individual basis or developed as part of the literature.

Share practices: As practitioners rarely utilize analysis support and tools from usability research, and many seem to rely on their own forms and formats to support analysis, there may be a potential benefit related to sharing. Practitioners do indeed share today, for example through conferences, seminars, and workshops. However it may be that such sharing practices could be extended even more - for example by way of social media, or just by lowering the threshold for presenting home-grown forms and formats for discussion outside ones own organisation.

Get inspired by others tool use: A varied selection of tools were reported to support analysis. In particular, we found the reported use of tools for web analytics and drawing to be inspiring. Web analytics may be useful both for testing the findings from usability evaluations, or to get information on what to test in the first place. The use of drawing tools is representative for the tight integration between evaluation and redesign, and it is likely that drawing skills will be an ever more useful in analysis in the future.

Do not be afraid to make redesign suggestions: Traditionally the literature has argued for a separation between analysis and design. However, current practitioners behave as if analysis and design are much more integrated than previously envisioned in the literature – for example do most evaluation deliverables include redesign suggestions. Such integration corresponds to Hornbæk and Frøkjær's (2005) finding that software developers appreciate redesign proposals, along with usability problems, as output from usability evaluation.

Know the reliability challenge: We found that practitioners see reliability as important in analysis, but that some seem to be unaware as to how reliability can be improved (or at least assessed). For collaboration to serve the purpose of improved reliability, individual analysts need to work independently on the same material. One way to achieve this is to include individual analysis as an introductory activity to collaboration; to allow all analysts to reach an opinion of current usability problems or solutions prior to collaboration.

8.2 Implications for research

Usability research may be highly useful to practitioners. However, this requires that the research is informed by knowledge of the practitioners' context. Our findings indicate that the last fifteen years of research on tools and methods to support analysts has had little direct impact on current evaluation practices. To guide future research on usability evaluation, we summarize six key implications of the survey findings.

Target method components: Our findings, both on the level of evaluation methods overall and on the level of analysis, indicate a need to refocus usability research towards method components that can be flexibly combined, rather than methods understood as fixed procedures. This implication resonates with recent research on method components (Woolrych, et al., 2011).

Support time-efficient analysis: To fit fast-paced usability evaluations analysis also needs to be time-efficient. Research is needed on methods and tools that support analysis without imposing time-demanding structure or process. The *instant data analysis* developed by Kjeldskov et al.(2005) may exemplify research-based analysis support that actually fit the time-requirements of usability evaluation.

Align research with commercial software: Only commercial tools, no tools from usability research, were reported to be used in analysis. This implies that analysis support from usability research should be compatible with the use of commercial tools. A relevant lead for future research is Howarth et al. (2009) who developed and studied a plug-in for the Morae analysis software.

Support home-growing: Forms and formats, such as formats for problem description, are were reported by the majority of the respondents. However, these were home-grown rather than taken from the literature. Because of this, we suggest that research should guide the development of home-grown forms and formats rather than aim to develop fixed general forms and formats. An example of such research is Capra (2006) who provided a set of guidelines for usability problem description which can be used as background for usability practitioners who make their own format for usability problem description.

Explore analysis in groups for usability testing: Group analysis of usability testing data is not well described in the literature. To support this practice, future research can establish guidelines for such analysis. One predecessor for such work may be the guidelines for analysis in groups for heuristic evaluation (Nielsen, 1994).

Facilitate redesign suggestions as part of evaluation: Making good redesign suggestions is challenging. At the same time, most evaluation deliverables include redesign suggestions. There is movement within usability research to look more into the interaction between evaluation and redesign (Hornbæk & Stage, 2006). However, it seems as the relationship between evaluation and redesign in practice, where the making of redesign suggestions to a significant degree is interwoven in the analysis of usability problems, is far more integrated than what is presupposed in current research.

9 Method

The study was conducted as a questionnaire survey. The reason for this was that we wanted to investigate the generality and extend findings from a pilot interview study where we had collected in-depth data from a small number of practitioners.

9.1 The respondents

The respondents were invited to the survey via a number of channels. We sent e-mail invitations to all local SIGCHI and UPA chapters, asking for these to be distributed to the chapter members. We distributed the survey via the UTEST mailing list and we sent email invitations to industry contacts of members of the European COST project TwinTide (a project on usability evaluation methods). We also set up an entry point to the survey at the SINTEF website, to recruit participants on basis of postings in social media. Finally, we distributed invitations as fliers at CHI 2011 - but this returned only 2 respondents. An overview of the recruitment channels of the 155 valid respondents is presented in Table 4.

Channel	Valid respondents
Local SIGCHI chapters	94
Local UPA chapters	24
UTEST mailing list	17
TwinTide project	9
SINTEF website	9
CHI 2011 flier	2
Total	**155**

Table 4: Distribution of valid respondents across recruitment channel

As incitements to participate, the respondents were promised a report of the findings (this report), and a ticket in a lottery for a 250 USD gift-card (the winner of which has been drawn).

In total 224 people responded to the invitation. Of these we included 155 in the subsequent analysis. The exclusion criteria were (a) not providing a response in any free text field (40 persons), (b) nonsense free text responses (3 persons), (c) no responses beyond the sixth question of the questionnaire (6 persons), and (d) not having conducted a usability evaluation the last six months (20 persons). The latter group of excluded respondents only answered the first question of the questionnaire.

To minimise effects of forgetfulness, we wanted the respondents to have conducted their latest usability evaluation fairly recently. The invitation was addressed to practitioners who had conducted a usability evaluation within the last six months. We asked the respondents about the start-up time for the evaluation they reported on. The majority had started their evaluation within the last two months. 10% reported to have started their latest evaluation more than six months ago. However, these had in the previous question stated that they had conducted a usability evaluation within the last six months. We therefore assume that also these had completed their evaluations within the last six months. Details on start-up times are provided in Table 5.

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

18 of 59

Startup of latest evaluation	Respondents
> 6 months ago	10%
3-6 months ago	28%
1-2 months ago	24%
< 1 month ago	38%
Total	**100%**

Table 5: Distribution of respondents across start-up times for their latest usability evaluation

The respondents worked in 21 countries across Europe (55%), the US (36%), Latin America (4.5%), and other parts of the world (4.5%). Three of the European countries had very high numbers of responses: Switzerland (15%), Czech Republic (15%) and Poland (10%). We checked the answers from respondents from these three countries with answers from the remaining sample. As no noteworthy differences were found between the respondents from these countries and other respondents, all were kept in the sample.

The vast majority of the respondents categorized themselves as usually doing *consultancy or development for external clients* (usability testing: 40%; usability inspection 63%) or *inhouse development projects* (usability testing 46%; usability inspection 23%). Less than 10% reported to usually work in *scientific research projects.*

The respondents were experienced usability practitioners, with a median of 5 years experience (25th percentile = 3 years, 75th percentile = 11 years). In the last 12 months they had conducted a median of 5 usability evaluations (25th percentile = 4, 75th percentile = 10).They had varied educational backgrounds spanning form computer science to media and communication. Education was reported as free text and then coded according to the categories presented in Table 6.

Respondents' educational background	UT	UI
Computer science	19%	19%
HCI or Human factors	29%	14%
Psychology, behavioral science or cognitive science	13%	19%
Science and engineering	11%	7%
Media and communication	10%	9%
Other	18%	32%
Total	**100%**	**100%**

Table 6: The respondents' educational background across usability testing (UT) and usability inspection (UI)

9.2 The questionnaire

Two versions of the questionnaire were developed. One for usability testing respondents, containing 32 questions (presented in Appendix 2), and one for usability inspection respondents, containing 29 questions (presented in Appendix 3). Twenty-six of the questions overlapped between the two versions; this was done to enable comparisons between usability testing respondents and usability inspection respondents. Both versions of the questionnaire were piloted with usability practitioners other than the authors of the report.

The respondents were directed to the questionnaire version corresponding to their answer on what kind of usability evaluation that was their most recent. In this question, it was also possible to report that one had not conducted a usability evaluation the last six months and thereby terminating the questionnaire session.

For the answers to reflect actual analysis practice, the respondents were to consistently report on their latest usability evaluation - not their general usability practice. The respondents were therefore explicitly asked to

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

19 of 59

report on their latest usability evaluation at the beginning of the questionnaire, and were also repeatedly throughout the questionnaire reminded to report on their latest usability evaluation.

To avoid that the order of answer alternatives could affect the respondents answers, the order of fixed answer alternatives were randomized for all questions where this was assumed to be a potential problem.

To avoid effects of the respondents being concerned about recognition, as well as to protect the privacy of the respondents, the data collection was fully anonymous. The respondents were also clearly informed of their anonymity. At the end of the survey we asked for the respondents e-mail addresses, to use when we were to send out the report of the study findings and to participate in the respondent lottery, but these were collected in an independent form and could not be connected to the other answers given.

9.3 The analysis

The quantitative analyses included in this report are descriptive, reflecting the exploratory aim of the study. Analyses of free text answers were conducted as thematic analyses, following Ezzy (2002). For the rather complex analysis of usability problem identification strategies, the free-text answers to the question Q21(UT)/Q18(UI), we used two independent analysts to check reliability. Free marginal kappa coefficients were in the range of .70-.78, which indicates adequate agreement (Randolph, online).

References

Andre, T.S., Hartson, H.R., and Williges, R.C. (2001). The user action framework: a reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies, 54*(1), 107-136.

Andre, T.S., Hartson, H.R., and Williges, R.C. (2003). Determining the effectiveness of the usability problem inspector: a theory-based model and tool for finding usability problems. *Human Factors, 45*(3), 455-482.

Capra, M. (2006). Usability *Problem Description and the Evaluator Effect in Usability Testing.* PhD thesis. Virginia Tech, Blacksburg, VA.

Cockton, G., and Lavery, D.A (1999). Framework for Usability Problem Extraction. In *Proceedings of INTERACT '99* (pp. 344-352). London, UK: IOS Press.

Cockton, G., Lavery, D. and Woolrych, A. (2008). Inspection-based Evaluations. In A. Sears and J. Jacko (eds.), *The Human-Computer Interaction Handbook*, 2nd edition (pp. 1171-1190). New York, NY: Lawrence Erlbaum Associates.

Cockton, G., Woolrych, A., and Hindmarch, M. (2004). Reconditioned Merchandise: Extended Structured Report Formats in Usability Inspection. In *Proceedings of CHI 2004, Extended Abstracts* (pp. 1433-1436). New York, NY: ACM Press.

DeVellis, R.F. (2003). *Scale Development: Theory and Applications* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Dumas, J.S. and Fox, J.E. (2008). Usability testing: Current practices and future directions. In A. Sears and J. Jacko (eds.), *The Human-Computer Interaction Handbook*, 2nd edition (1129-1150). New York, NY: Lawrence Erlbaum Associates.

Dumas, J.S., and Redish, J. (2010). *A practical guide to usability testing*. Norwood, NJ: Ablex Publishing Corporation.

Ezzy, D. (2002). *Qualitative analysis: practice and innovation*. London, UK: Routledge.

Følstad, A., Law, E.L-C. and Hornbæk, K. (2010). Analysis in Usability Evaluations: an Exploratory Study. In *Proceedings of NordiCHI 2010* (pp. 647-650). New York, NY: ACM Press.

Gawande, A. (2009). *The Checklist Manifesto: How to Get Things Right*. New York, NY: Metropolitan books.

Hertzum, M., and Jacobsen, N.E. (2003). The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction 15*(1), 183-204.

Hornbæk, K., and Frøkjær, E. (2005). Comparing usability problems and redesign proposals as input to practical systems development. *In Proceedings of CHI 2005* (pp. 391-400). New York, NY: ACM Press.

Hornbæk, K., and Stage, J. (2006). The Interplay between Usability Evaluation and User Interaction Design - Introduction to Special Issue. *International Journal of Human-Computer Interaction, 22*(2), 117-123.

Howarth, J., Smith-Jackson, T., and Hartson, R. (2009). Supporting novice usability practitioners with usability engineering tools. *International Journal of Human–Computer Studies, 67*(6), 533-549.

Kjeldskov J., Skov M. B., and Stage J. (2004). Instant Data Analysis: Evaluating Usability in a Day. In *Proceedings of NordiCHI '04* (pp. 233-240). New York, NY: ACM Press.

Lavery, D., Cockton, G., and Atkinson, M.P. (1997). Comparison of Evaluation Methods Using Structured Usability Problem Reports. *Behaviour and Information Technology, 16*(4), 246-266.

Nielsen, J. (1994). Heuristic Evaluation. In J. Nielsen and R.L. Mack (eds.) *Usability Inspection Methods* (pp. 25-62). New York, NY: John Wiley and Sons.

Nørgaard, M., and Hornbæk, K. (2006). What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing. *In Proceedings of DIS 2006* (pp. 209-218). New York, NY: ACM Press.

Randolph, J.J. (online). *Online Kappa Calculator.* Retrieved from http://justus.randolph.name/kappa

Rubin, J. and Chisnell, D. (2008). *Handbook of usability testing* (2nd.edition). Wiley Publishing.

Skov, J., and Stage, J. (2005). *Supporting Problem Identification in Usability Evaluations*. Paper presented at OzCHI 2005.

Wharton, C., Rieman, J., Lewis, C., and Polson, P. (1994). The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen and R.L. Mack (eds.) *Usability Inspection Methods* (pp. 105-140). New York, NY: John Wiley and Sons.

Woolrych, A. Hornbæk K. Frøkjær, E. and Cockton, G. (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.

Appendix 1: Detailed results

In the following, detailed results for all quantitative questions on analysis in the respondents' latest usability evaluations are presented. We do not include a detailed overview neither of the qualitative data nor the background data, as we thought this would be too much. Also, summaries of qualitative and background data are given in the report. If you are interested in more details, however, please contact the first author of the report.

Question numbering are presented as Q?(UT)/Q?(UI). The first number refers to the usability testing (UT) questionnaire. The second refers to the usability inspection (UI) questionnaire.

Q12(UT)/Q13(UI), sub-question 1: How challenging did you consider the following when conducting this usability [test / inspection]: To identify all relevant usability problems.
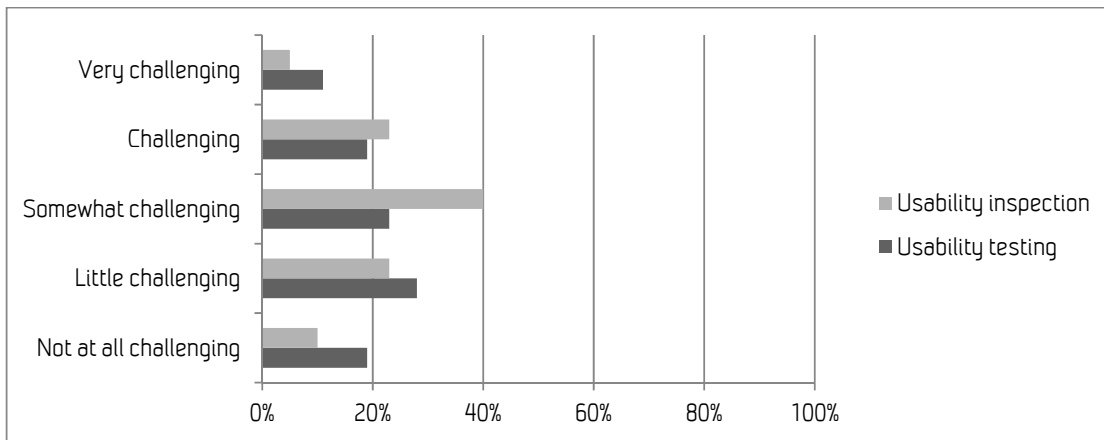
Q12(UT)/Q13(UI), sub-question 2: How challenging did you consider the following when conducting this usability [test / inspection]: To avoid false usability problems.
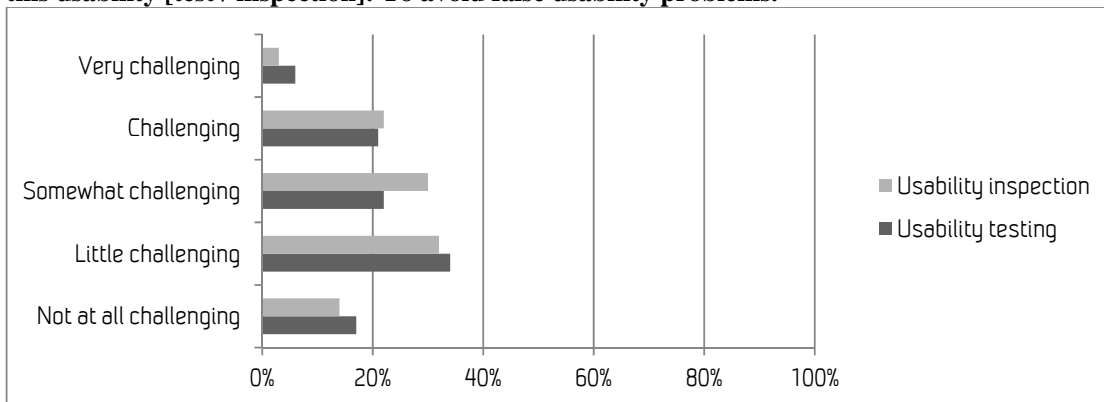
Q12(UT)/Q13(UI), sub-question 3: How challenging did you consider the following when conducting this usability [test / inspection]: To make good redesign suggestions.
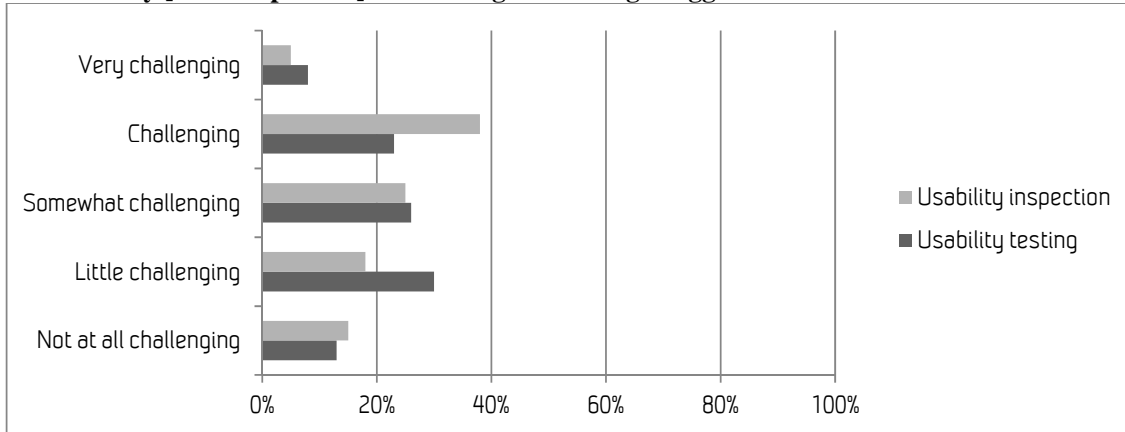
Q12(UT)/Q13(UI), sub-question 4: How challenging did you consider the following when conducting this usability [test / inspection]: To prioritize the problems and/or redesign suggestions.
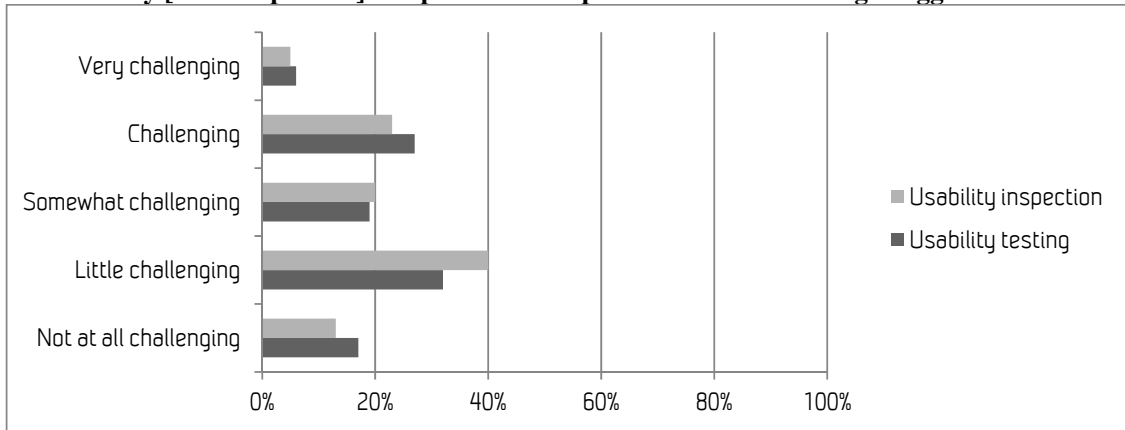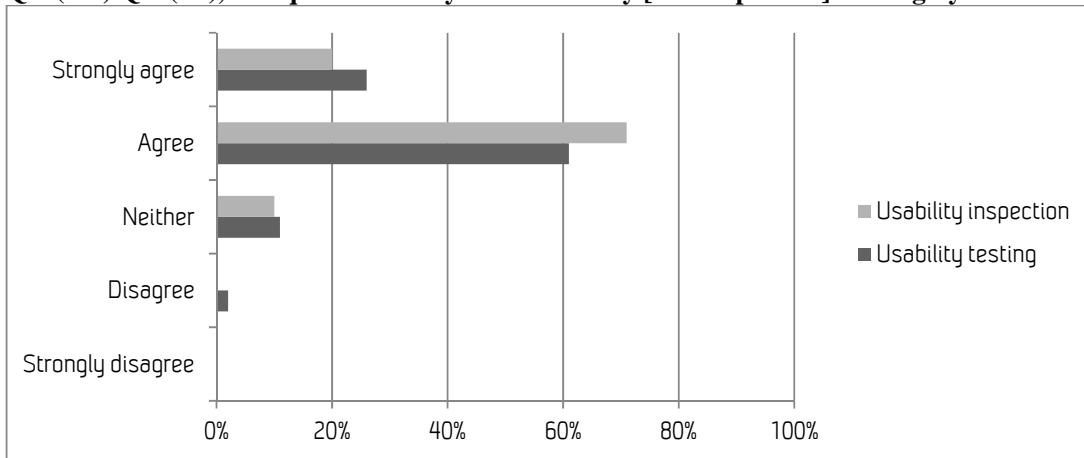

Q12(UT)/Q13(UI), sub-question 5: How challenging did you consider the following when conducting this usability [test / inspection]: To convince the development team and/or customer.

PROJECT NO.
90B273

REPORT NO.
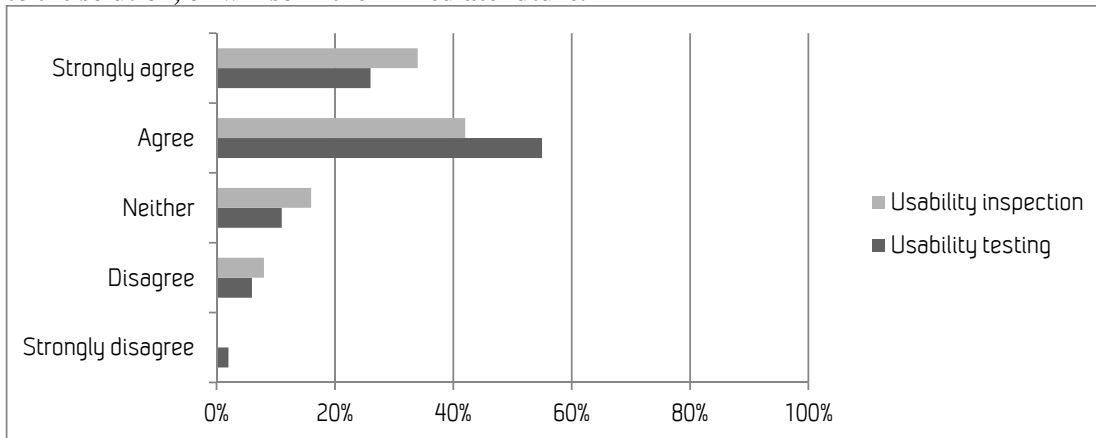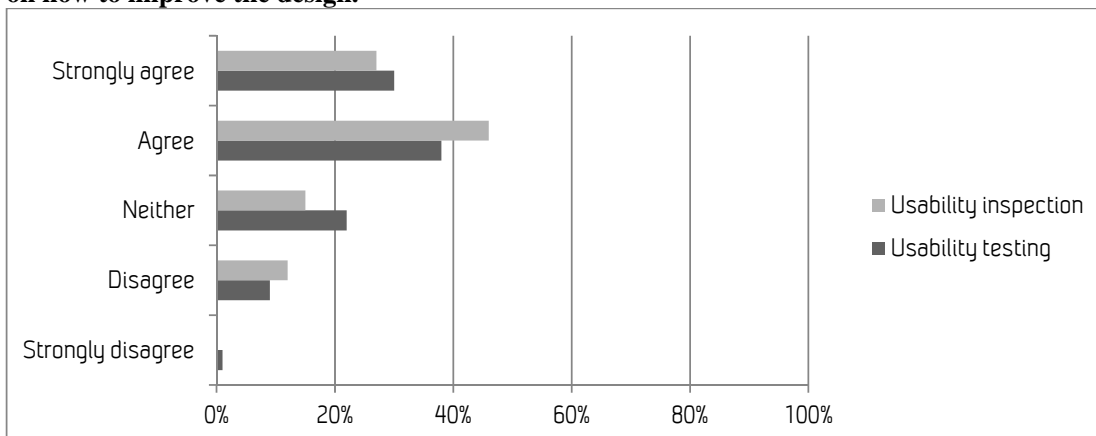SINTEF A22665

VERSION
1

23 of 59
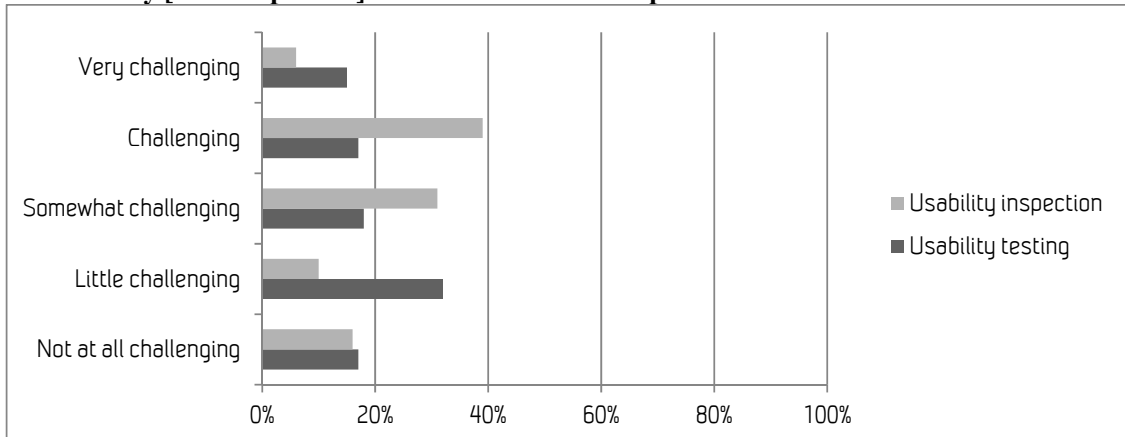
Q13(UT)/Q14(UI), sub-question 1: My latest usability [test/inspection] was highly successful.
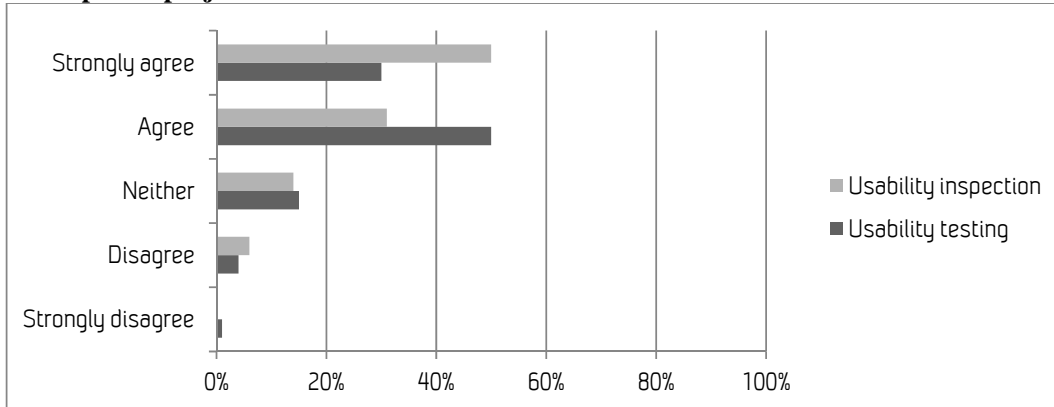

Q13(UT)/Q14(UI), sub-question 2: My latest usability [test/inspection] has caused important changes to the solution, or will so in the immediate future.


Q13(UT)/Q14(UI), sub-question 3: My latest usability [test/inspection] has lead to extensive dialogue on how to improve the design.

Q13(UT)/Q14(UI), sub-question 4: The results of my latest usability test has been actively used in the development project.


Q13(UT)/Q14(UI), sub-question 5: It was easy to communicate the results of my latest usability test to the people in the development project.
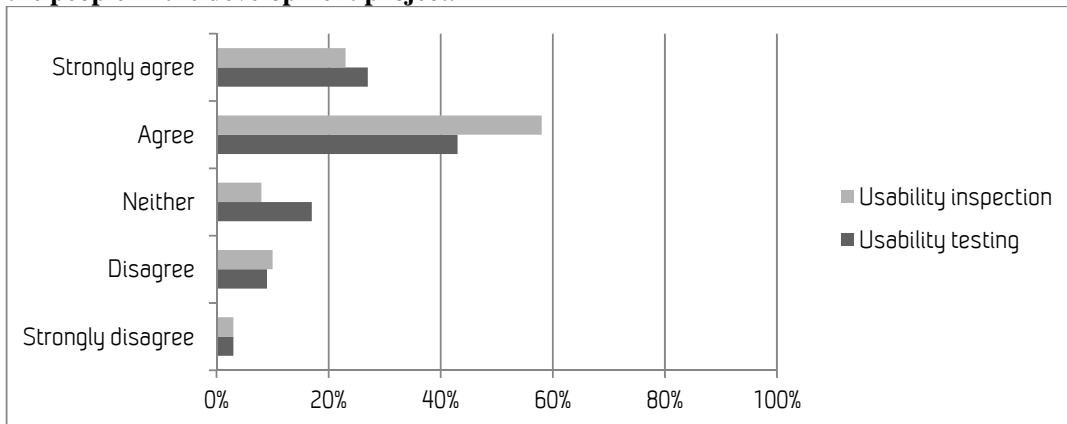

Q14(UT)/Q15(UI): What kind of notes was taken during the user sessions of your latest usability [test / inspection]? (Choose one or more)

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

25 of 59
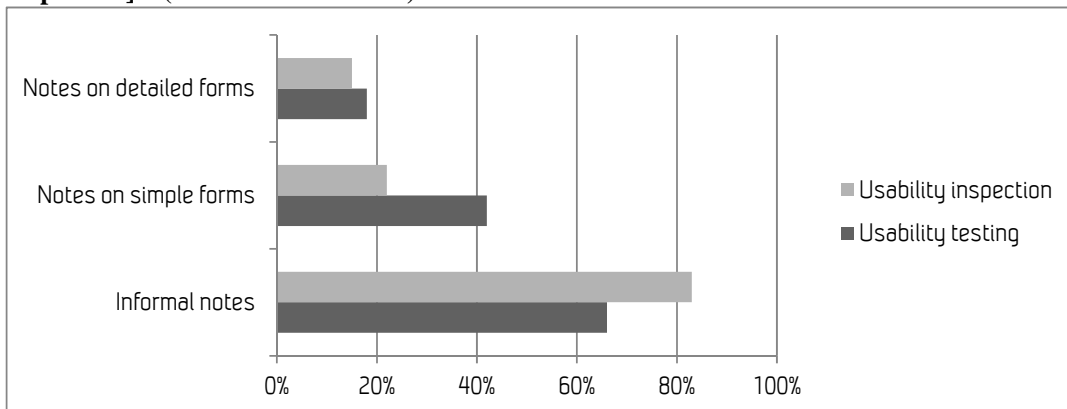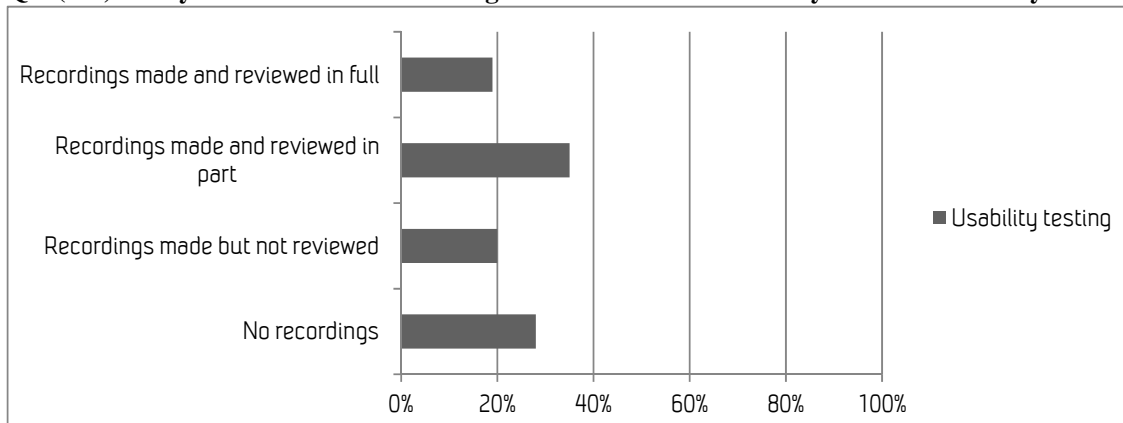
Q15(UT): Were the user participants or the observers asked for their opinion on … (choose one or more for each column)
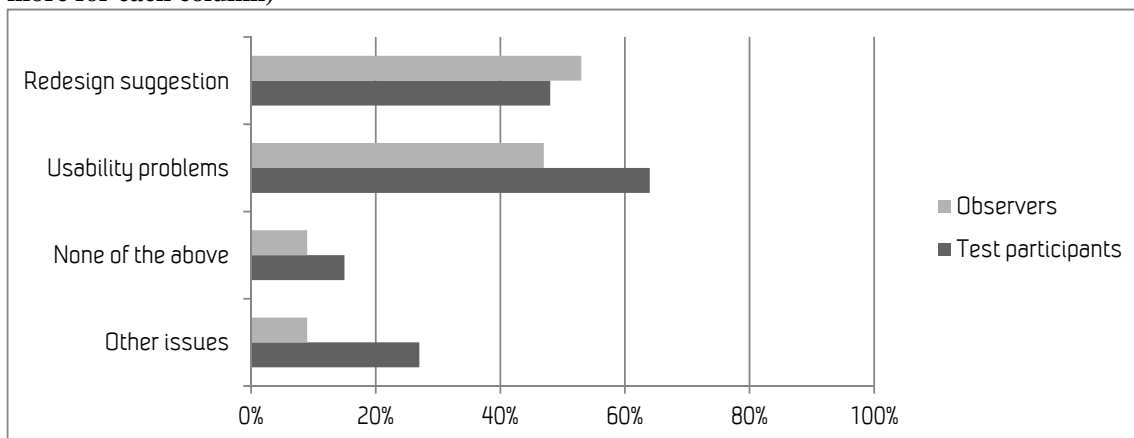

Q16(UT): Did you do a summary of key findings immediately following each test person?
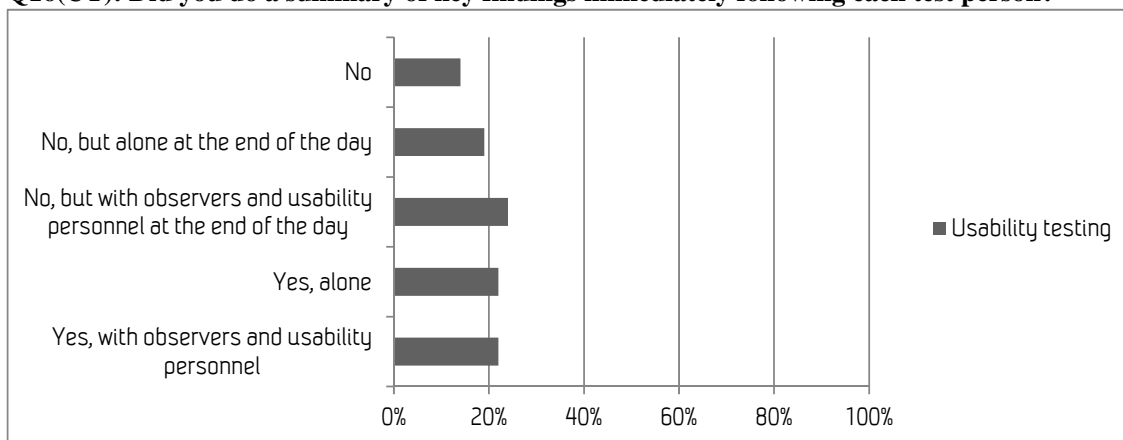
Q17(UT): Did you review video recordings from the user sessions of your latest usability test?

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

26 of 59

Q18(UT)/Q16(UI): Which of the following resources did you make explicit use of during analysis in your latest usability [test / inspection]? (Choose one or more)
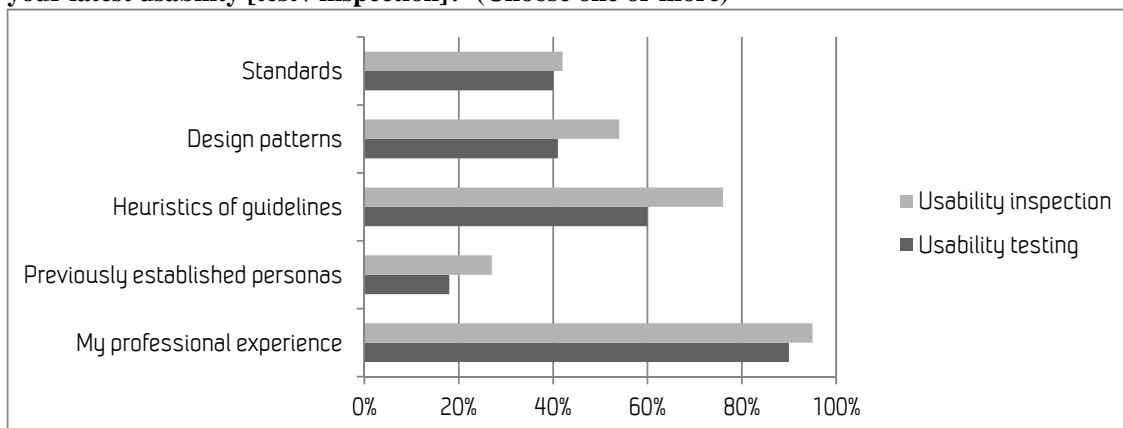
Q19: Which measures, if any, did you use during analysis in your latest usability test?
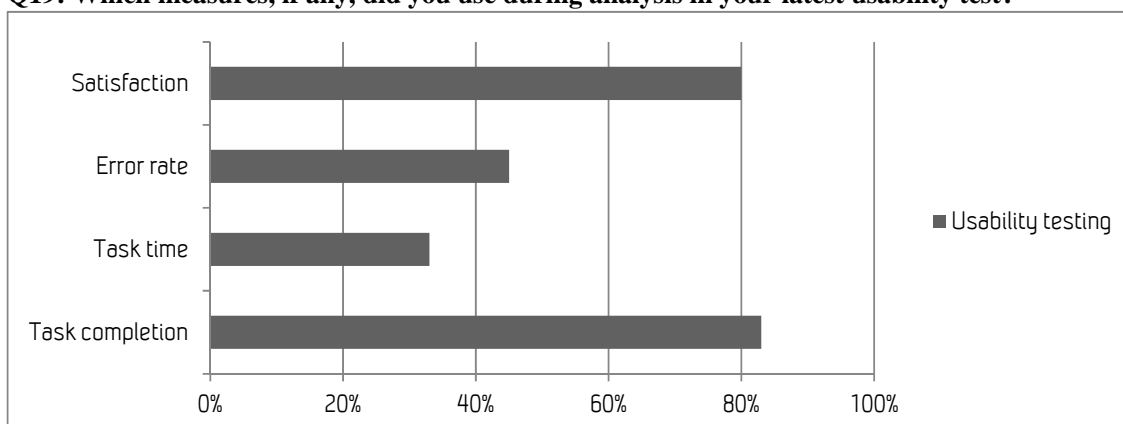
Q23(UT)/Q19(UI): What kind of collaboration did you have with other usability professionals during the analysis of your latest usability [test / inspection]? (Choose one or more)
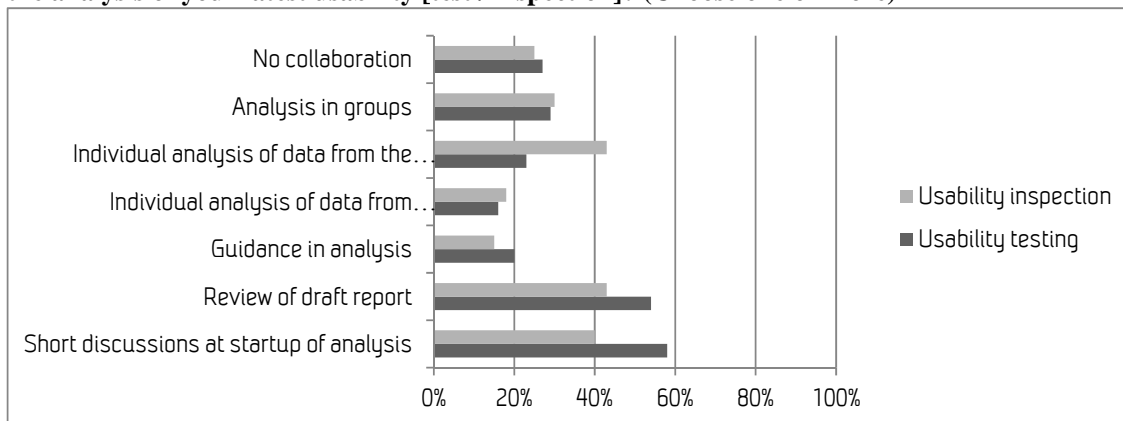
Q24(UT)/Q20(UI): If you were to choose one, what would you say was the main purpose of collaboration in the [period between user test sessions and the final report / later phases of your latest usability inspection]? (Choose one)

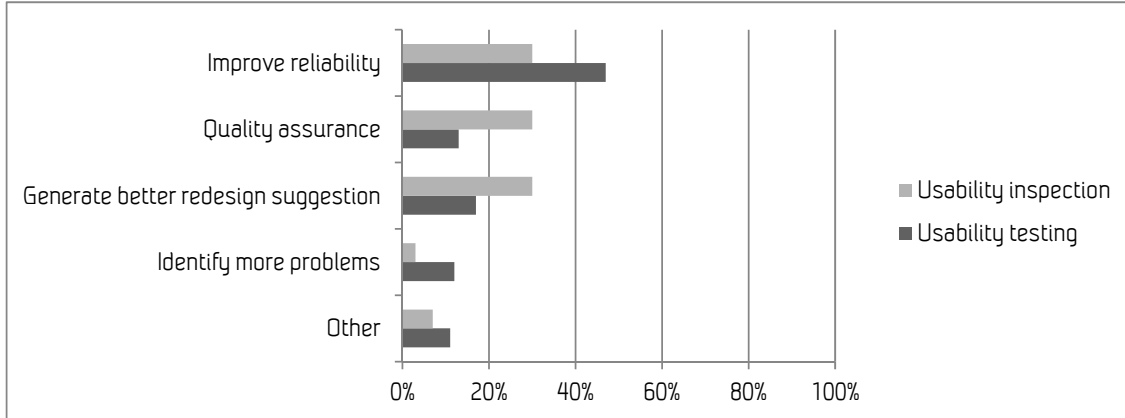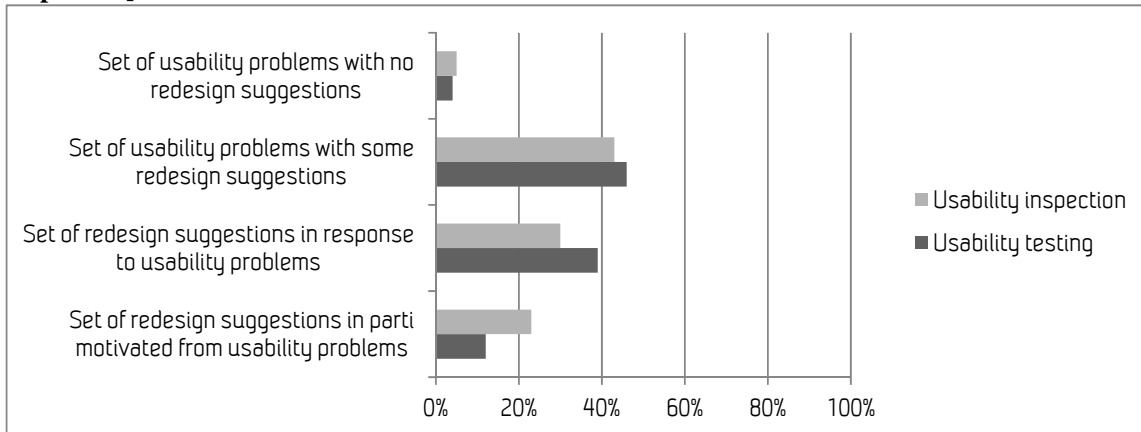Categories (top to bottom): Improve reliability, Quality assurance, Generate better redesign suggestion, Identify more problems, Other
Legend: Usability inspection, Usability testing

Q25(UT)/Q22(UI): How would you characterize the deliverable from your latest usability [test / inspection]?

Categories (top to bottom): Set of usability problems with no redesign suggestions; Set of usability problems with some redesign suggestions; Set of redesign suggestions in response to usability problems; Set of redesign suggestions in parti motivated from usability problems
Legend: Usability inspection, Usability testing

Q26(UT)/QQ23(UI): When did you make the redesign suggestions?

Categories (top to bottom): Some redesigns immediately, others after all usability problems had been identified; Whenever a usability problem was identified; First all usability problems, then redesign suggestions; Other
Legend: Usability inspection, Usability testing
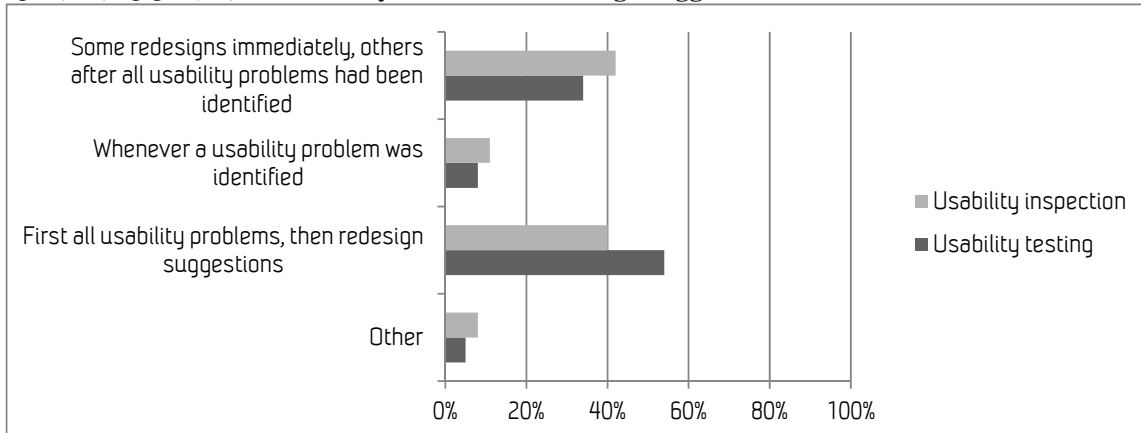
Q27(UT)/Q24(UI): How did you reach the redesign suggestions that were included in the deliverable? (Chose one or more)
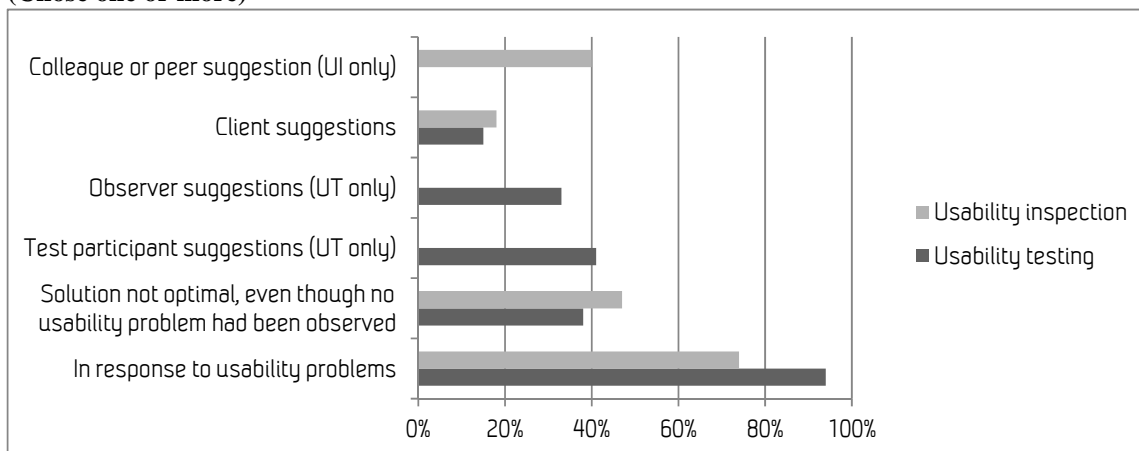
Q28(UT)/Q25(UI): How did you present redesign suggestions in the deliverable? (Chose one or more)
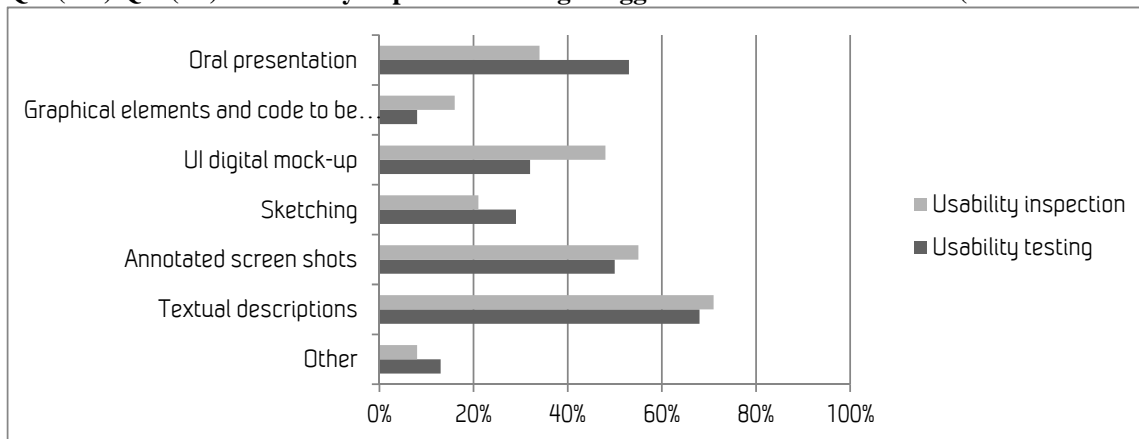
Q29(UT)/Q26(UI): Did you use screen-shots to illustrate usability problem descriptions?

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

29 of 59

Q30(UT)/Q27(UI): Did you use a structured format for usability problem description?

- Yes, structured format from standards or literature
- Yes, my/our own format
- No, problems described in plain prose

■ Usability inspection
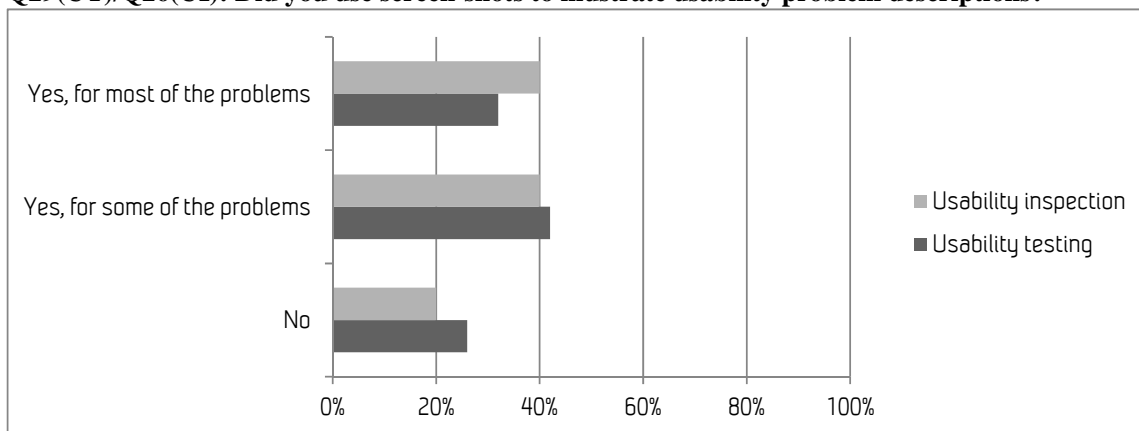■ Usability testing

0% 20% 40% 60% 80% 100%
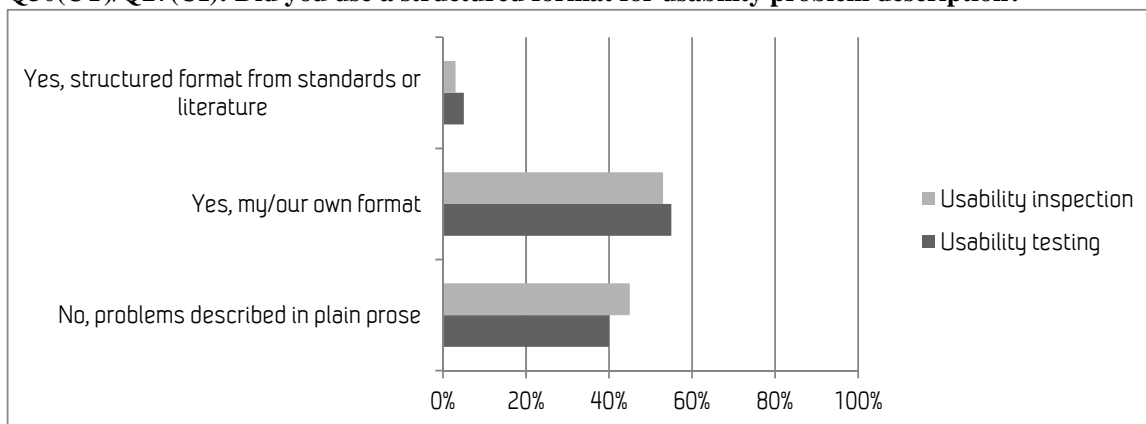
Q31(UT)/Q28(UI): Did you prioritize the usability problems or redesign suggestions?

- No
- Yes, both problems and suggestions
- Yes, only redesign suggestions
- Yes, only usability problems

■ Usability inspection
■ Usability testing

0% 20% 40% 60% 80% 100%
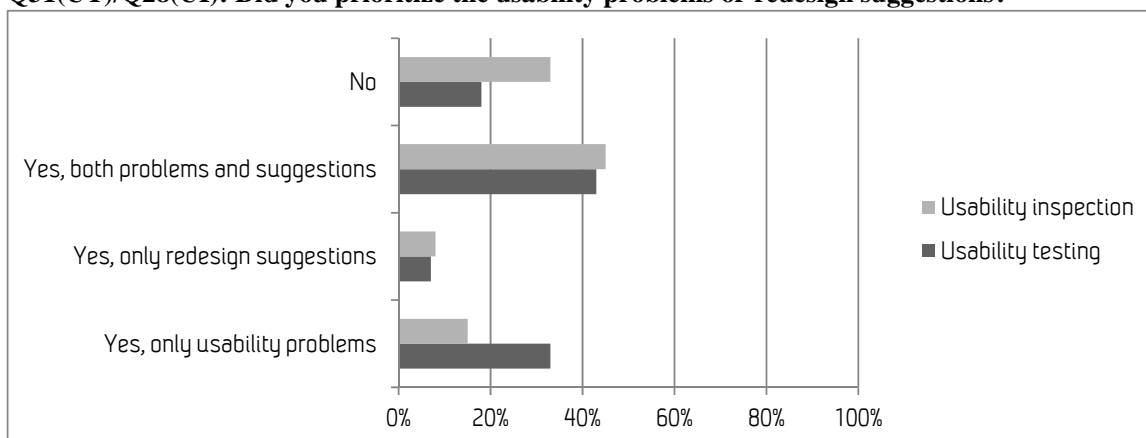
Q32(UT)/Q29(UI): How did you prioritize the usability problems or redesign suggestions in your latest usability [test / inspection]? (Choose the alternative closest to describing the classification you used)

- Urgency ratings
- Severity ratings
- Other

■ Usability inspection
■ Usability testing

0% 20% 40% 60% 80% 100%

App

Intro

Welcome to our study on how usability professionals conduct the later phase of usability evaluation.

For usability testing, the later phase is the process of going from user observations to the results hand-over to the client. For usability inspection, the process from walking through the user interface to the results hand-over.

We invite usability practitioners who have conducted a usability evaluation within the last six months.

This study is important. Currently we do not have a sufficient understanding of how the later phases of usability evaluation is conducted. The results will provide valuable guidance for future research on usability evaluation methods.

Your participation will take about 15 minutes. As an appreciation of your effort, **we draw a prize between the survey participants – the winner can choose between an iPod Touch 32 GB or a 250$ (US) Amazon gift card.**. You will also be given free access to the study report prior to its publication.

The questionnaire consists of **32 questions**, mainly targeting your latest usability evaluation.

It is important for the quality of the study that you answer honestly. All responses are collected anonymously and cannot be tracked back to you. To participate in the drawing of the prize you have to leave your e-mail address, but the address will be fully disconnected from your other responses. You are free to resign from the study at any time.

Thanks for participating!

(Study organiser names removed)

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

31 of 59

Your latest usability evaluation

✱1. Yes, I have conducted a usability evaluation within the last six months and would like
My latest usability evaluation was:

- ◯ Usability test (evaluation that involves observation of users)
- ◯ Usability inspection (analytical evaluation)
- ◯ Have not conducted a usability evaluation within the last six months

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

32 of 59

About you

2. How many years have you been working as a usability practitioner?

[ ]

3. How many usability evaluations have you conducted the last 12 months? (approximately)

[ ]

4. What kind of usability projects are you usually engaged in?

◯ Inhouse development projects

◯ Consultancy / development projects for external clients

◯ Adaptation of commercial software

◯ Scientific research projects

◯ Other (please specify)

[ ]

5. Your educational background (subject/field, level)

[ ]

6. Your country of work.

[ ]

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

33 of 59

Your latest usability test

The remainder of the questionnaire will target your latest usability test. Please make sure that you have one particular test in mind when answering the remaining questions.

7. When did you start working on your latest usability test?

○ More than 6 months ago

○ 3-6 months ago

○ 1-2 months ago

○ Less than 1 month ago

8. In which development phase was the object of evaluation?

○ Concept *(overall presentation of solution but without presentation of interaction mechanisms)*

○ Early prototype

○ Advanced prototype

○ System in use

9. How many user participants were involved?

[ ]

10. How many hours were spent on preparing, conducting and reporting the usability test

(Approximately, in total for all usability professionals involved. Just leave blank if you do not know.)

[ ]

11. Would you say this usability test was typical with respect to time and resource constraints?

○ Yes

○ No, we had less time and/or resources than we typically have

○ No, we had more time and/or resources than we typically have

SINTEF

Perceptions of your latest usability test

12. How challenging did you consider the following when conducting this usability test?

	Not at all challenging	Little challenging	Somewhat challenging	Challenging	Very cha
To identify all relevant usability problems	○	○	○	○	○
To avoid "false" usability problems (that is, issues that seem to be a problem during evaluation but that turns out not to be a problem in the real world context).	○	○	○	○	○
To make good redesign suggestions	○	○	○	○	○
To prioritize the problems and/or redesign suggestions	○	○	○	○	○
To convince the development team and/or customer	○	○	○	○	○

13. Please indicate your agreement with the following statements

	Strongly disagree	Disagree	Neither	Agree	Strongly
My latest usability test was highly successful	○	○	○	○	○
My latest usability inspection has caused important changes to the solution, or will so in the immediate future	○	○	○	○	○
My latest usability test has lead to extensive dialogue on how to improve the design	○	○	○	○	○
The results of my latest usability test have been actively used in the development project	○	○	○	○	○
It was easy to communicate the results of my latest usability test to the people in the development project	○	○	○	○	○

During the user sessions

14. What kind of notes was taken during the user sessions of your latest usability test? (Choose one or more)

☐ Informal notes

☐ Notes on simple forms *(for example forms structured as a few high-level sections)*

☐ Notes on detailed forms *(for example forms including tables or similar for filling in user data)*

If you used simple or detailed forms for note taking, please describe the form briefly.

```

```

15. Were the user participants or the observers asked for their opinion on …
(choose one or more for each column)

	User participants	Obse
Possible usability problems	☐	☐
Possible redesign suggestions	☐	☐
Other issues (please specify)	☐	☐
None of the above	☐	☐

If you answered "Other", please specify here

```

```

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

36 of 59

After the user sessions - I

16. Did you do a summary of key findings immediately following each test person?

○ Yes, the present observers and usability personnel summarized key findings together immediately following each test session

○ Yes, I summarized key findings by myself immediately following each test person

○ No, but the present observers and usability personnel summarized key findings at the end of the day

○ No, but I summarized key findings by myself at the end of the day

○ No such immediate summary of key findings was conducted.

17. Did you review video recordings from the user sessions of your latest usability test?

○ No recordings were made

○ Recordings were made, but not reviewed

○ Recordings were made, but reviewed in part only

○ Recordings were made and fully reviewed

18. Which of the following resources did you make explicit use of during analysis in your la test?

(Choose one or more)

☐ My professional experience

☐ Previously established personas

☐ Heuristics or guidelines

☐ Design patterns

☐ Standards (*norm or technical requirement established within a given field or application area*)

☐ Other resources

[   ]

19. Which measures, if any, did you use during analysis in your latest usability test?

☐ Task completion

☐ Task time

☐ Error rate

☐ Satisfaction

☐ Other (please specify)

[   ]

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

37 of 59

After the user sessions - II

20. Which tools, if any, did you use for usability problem description and analysis in your latest usability test?

21. Please provide a brief description of how you decided on something being a usability problem.

22. If an incident was observed with *only one* of the users participating in your latest usability test,

how did you decide whether this was a usability problem or not?

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

38 of 59

Collaboration in your latest usability test

23. What kind of collaboration did you have with other usability professionals during the analysis of your latest usability test? (Choose one or more)

☐ Short discussions immediately following user test sessions

☐ Review of draft report

☐ Guidance when conducting analysis

☐ Individual analysis of data from **different** user sessions by myself and at least one other colleague

☐ Individual analysis of data from **the same** user sessions by myself and at least one other colleague

☐ The data analysis was conducted as a group activity by two or more usability professionals.

☐ **No collaboration**

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

39 of 59

Motivation for collaboration

24. If you were to choose one, what would you say was the main purpose of collaboration the period between user test sessions and the final report? (Choose one)

◯ Identify more problems

◯ Generate better redesign suggestions

◯ Quality assurance

◯ Improve reliability / avoid that the findings were biased by personal perspective

◯ Other (please specify)

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

40 of 59

The usability test deliverable

25. How would you characterize the deliverable from your latest usability test?

○ A set of redesign suggestions in part motivated from usability problems

○ A set of redesign suggestions in response to a set of usability problems

○ A set of usability problems with some redesign suggestions

○ A set of usability problems with no redesign suggestions

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

41 of 59

Redesign suggestions

26. When did you make the redesign suggestions?

○ First all usability problems were identified, then the redesign suggestions were made

○ Whenever a usability problem was identified, a redesign suggestion was immediately made

○ Some redesign suggestions were immediately when a usability problem was identified, others were made only when all usability problems had been identified

○ Other (please specify)

27. How did you reach the redesign suggestions that were included in the deliverable? (Chose one or more)

☐ In response to usability problems

☐ On basis of recognition that the chosen solution was not optimal, even though no usability problem had been observed.

☐ Test participant suggestions

☐ Observer suggestions

☐ Client suggestions

☐ Other (please specify)

28. How did you present redesign suggestions in the deliverable? (Chose one or more)

☐ Textual descriptions

☐ Annotated screen shots

☐ Sketching

☐ UI digital mock-up

☐ Graphical elements or code that could be immediately implemented

☐ Oral presentation

☐ Other (please specify)

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

42 of 59

Structured formats and severity classifications

29. Did you use screen-shots to illustrate usability problem descriptions?

○ No

○ Yes, for some of the problems

○ Yes, for most or all of the problems

30. Did you use a structured format for usability problem description?

○ No, the problems were described in plain prose.

○ Yes, the problems were described according to my/our own format

○ Yes, the problems were described according to a structured format described in standards or literature

If you used a structured format described in standards or literature, please specify

[   ]

31. Did you prioritize the usability problems or redesign suggestions?

○ Yes, but only usability problems

○ Yes, but only redesign suggestions

○ Yes, both usability problems and redesign suggestions

○ No, neither usability problems nor redesign suggestions were prioritized

Type of ratings

32. How did you prioritize the usability problems or redesign suggestions in your latest usability test? (Choose the alternative closest to describing the classification you used)

○ Severity ratings *(classifying the predicted impact of a usability problem on user behavior)*

○ Urgency ratings *(classifying how important it is to make a design change in response to a usability problem or redesign suggestion)*

○ Other (please specify)

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

44 of 59

Thank you

Thank you for participating! This is a valuable contribution to the research on analysis in usability evaluation.

For distribution of report results, and for notification if you win the participant prize, **please provide your e-mail address on the next page**.

The answers you have given on the previous pages are anonymous even if you leave your e-mail address. The e-mail address will at no point in ti be connected to your answers.

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

45 of 59

Appendix 3: Usability inspection questionnaire

Intro

Welcome to our study on how usability professionals conduct the later phase of usability evaluation.

For usability testing, the later phase is the process of going from user observations to the results hand-over to the client. For usability inspection, the process from walking through the user interface to the results hand-over.

We invite usability practitioners who have conducted a usability evaluation within the last six months.

This study is important. Currently we do not have a sufficient understanding of how the later phases of usability evaluation is conducted. The results will provide valuable guidance for future research on usability evaluation methods.

Your participation will take about 15 minutes. As an appreciation of your effort, **we draw a prize between the survey participants – the winner can choose between an iPod Touch 32 GB or a 250$ (US) Amazon gift card.**. You will also be given free access to the study report prior to its publication.

The questionnaire consists of **32 questions**, mainly targeting your latest usability evaluation.

It is important for the quality of the study that you answer honestly. All responses are collected anonymously and cannot be tracked back to you. To participate in the drawing of the prize you have to leave your e-mail address, but the address will be fully disconnected from your other responses. You are free to resign from the study at any time.

Thanks for participating!

(Study organizer names removed)

Your latest usability evaluation

✱1. Yes, I have conducted a usability evaluation within the last six months and would like
My latest usability evaluation was:

○ Usability test (evaluation that involves observation of users)

○ Usability inspection (analytical evaluation)

○ Have not conducted a usability evaluation within the last six months

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

47 of 59

ABOUT YOU

2. How many years have you been working as a usability practitioner?

3. How many usability evaluations have you conducted the last 12 months? (approximately)

4. What kind of usability projects are you usually engaged in?

○ Inhouse development projects

○ Consultancy / development projects for external clients

○ Adaptation of commercial software

○ Scientific research projects

○ Other (please specify)

5. Your educational background (subject/field, level)

6. Your country of work.

YOUR LATEST USABILITY INSPECION

The remainder of the questionnaire will target your latest usability inspection (analytical evaluation). Please make sure that you have one particular inspection in mind when answering the remaining questions.

7. When did you start working on your latest usability inspection?

- ○ More than 6 months ago
- ○ 3-6 months ago
- ○ 1-2 months ago
- ○ Less than 1 month ago

8. In which development phase was the object of evaluation?

- ○ Concept (overall presentation of solution but without presentation of interaction mechanisms)
- ○ Early prototype
- ○ Advanced prototype
- ○ System in use

9. Which usability inspection method was used?

- ○ Heuristic evaluation
- ○ Cognitive walkthrough
- ○ Informal expert evaluation
- ○ No particular method / mix of methods
- ○ Other method (please specify)

10. Please provide a brief description of the inspection method.

11. How many hours were spent on preparing, conducting and reporting the usability inspection
(Approximately, in total for all usability professionals involved. Just leave blank if you do not know.)

12. Would you say this usability inspection was typical with respect to time and resource constraints?

- ○ Yes
- ○ No, we had less time and/or resources than we typically have
- ○ No, we had more time and/or resources than we typically have

PERCEPTIONS OF YOUR LATEST USABILITY INSPECTION

13. How challenging did you consider the following when conducting this usability inspec

	Not at all challenging	Little challenging	Somewhat challenging	Challenging	Very cha
To identify all relevant usability problems	○	○	○	○	○
To avoid "false" usability problems (that is, issues that seem to be a problem during evaluation but that turns out not to be a problem in the real world context).	○	○	○	○	○
To make good redesign suggestions	○	○	○	○	○
To prioritize the problems and/or redesign suggestions	○	○	○	○	○
To convince the development team and/or customer	○	○	○	○	○

14. Please indicate your agreement with the following statements

	Strongly disagree	Disagree	Neither	Agree	Strongly
My latest usability inspection was highly successful	○	○	○	○	○
My latest usability inspection has caused important changes to the solution, or will so in the immediate future	○	○	○	○	○
My latest usability inspection has lead to extensive dialogue on how to improve the design	○	○	○	○	○
The results of my latest usability inspection have been actively used in the development project	○	○	○	○	○
It was easy to communicate the results of my latest usability inspection to the people in the development project	○	○	○	○	○

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

50 of 59

DURING THE INSPECTION (1)

15. What kind of notes was taken when going through the object of evaluation during your latest usability inspection?

(Choose one or more)

☐ Informal notes

☐ Notes on simple forms *(for example forms structured as a few high-level sections)*

☐ Notes on detailed forms *(for example forms including tables or similar for filling in user data)*

If you used simple or detailed forms for note taking, please describe the form briefly.

[text box]

16. Which of the following resources did you make explicit use of during analysis in your la[...] inspection?

(Choose one or more)

☐ My professional experience

☐ Previously established personas

☐ Heuristics or guidelines

☐ Design patterns

☐ Standards *(norm or technical requirement established within a given field or application area)*

☐ Previously established scenarios of use

☐ Other resources

[text box]

DURING THE INSPECTION (2)

17. Which tools, if any, did you use for usability problem description and analysis in your latest usability inspection?

18. Please provide a brief description of how you decided on something being a usability problem.

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

52 of 59

COLLABORATION IN YOUR LATEST USABILITY INSPECTION

19. What kind of collaboration did you have with other usability professionals during your latest usability inspection? (Choose one or more)

- [ ] Short discussions immediately following startup of the inspection
- [ ] Review of draft report
- [ ] Guidance when conducting the inspection
- [ ] Individual inspections of **different** parts of the system by myself and at least one other colleague
- [ ] Individual inspections of **the same** part of the system by myself and at least one other colleague
- [ ] The inspection was conducted as a group activity by two or more usability professionals.
- [ ] **No collaboration**

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

53 of 59

MOTIVATION FOR COLLABORATION

20. If you were to choose one, what would you say was the main purpose of collaboration the later phases of your latest usability inspection? (Choose one)

○ Identify more problems

○ Generate better redesign suggestions

○ Quality assurance

○ Improve reliability / avoid that the findings were biased by personal perspective

○ Other (please specify)

[   ]

21. If *only one* of the usability personnel participating in your latest usability inspection identified something as a usability problem, how did you decide whether this was to be reported as a usability problem or not?

[   ]

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

54 of 59

THE USABILITY INSPECTION DELIVERABLE

22. How would you characterize the deliverable from your latest usability inspection?

○ A set of redesign suggestions in part motivated from usability problems

○ A set of redesign suggestions in response to a set of usability problems

○ A set of usability problems with some redesign suggestions

○ A set of usability problems with no redesign suggestions

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

55 of 59

REDESIGN SUGGESTIONS

23. When did you make the redesign suggestions?

○ First all usability problems were identified, then the redesign suggestions were made

○ Whenever a usability problem was identified, a redesign suggestion was immediately made

○ Some redesign suggestions were immediately when a usability problem was identified, others were made only when all usability problems had been identified

○ Other (please specify)

24. How did you reach the redesign suggestions that were included in the deliverable? (Chose one or more)

☐ In response to usability problems

☐ On basis of recognition that the chosen solution was not optimal, even though no usability problem had been observed.

☐ Client suggestions

☐ Colleague or peer suggestions

☐ Other (please specify)

25. How did you present redesign suggestions in the deliverable? (Chose one or more)

☐ Textual descriptions

☐ Annotated screen shots

☐ Sketching

☐ UI digital mock-up

☐ Graphical elements or code that could be immediately implemented

☐ Oral presentation

☐ Other (please specify)

STRUCTURED FORMATS AND SEVERITY CLASSIFICATIONS

26. Did you use screen-shots to illustrate usability problem descriptions?

○ No

○ Yes, for some of the problems

○ Yes, for most or all of the problems

27. Did you use a structured format for usability problem description?

○ No, the problems were described in plain prose.

○ Yes, the problems were described according to my/our own format

○ Yes, the problems were described according to a structured format described in standards or literature

If you used a structured format described in standards or literature, please specify

[ ]

28. Did you prioritize the usability problems or redesign suggestions?

○ Yes, but only usability problems

○ Yes, but only redesign suggestions

○ Yes, both usability problems and redesign suggestions

○ No, neither usability problems nor redesign suggestions were prioritized

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

57 of 59

TYPE OF RATINGS

29. How did you prioritize the usability problems or redesign suggestions in your latest usability inspection? (Choose the alternative closest to describing the classification you used)

○ Severity ratings *(classifying the predicted impact of a usability problem on user behavior)*

○ Urgency ratings *(classifying how important it is to make a design change in response to a usability problem or redesign suggestion)*

○ Other (please specify)

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

58 of 59

Thank you

Thank you for participating! This is a valuable contribution to the research on analysis in usability evaluation.

For distribution of report results, and for notification if you win the participant prize, **please provide your e-mail address on the next page**.

The answers you have given on the previous pages are anonymous even if you leave your e-mail address. The e-mail address will at no point in ti be connected to your answers.

PROJECT NO.
90B273

REPORT NO.
SINTEF A22665

VERSION
1

59 of 59