

Analysis in Practical Usability Evaluation: A Survey Study

Asbjørn Følstad

SINTEF

P.O. Box 124, Blindern,
N-0134 Oslo, Norway.

asf@sintef.no

Effie Lai-Chong Law

Dept. of Computer Science
University of Leicester,
Leicester, LE1 7RH, UK.

elaw@mcs.le.ac.uk

Kasper Hornbæk

Department of Computing,
University of Copenhagen,
Njalsgade 128, DK-2300
Copenhagen S, Denmark.

kash@diku.dk

ABSTRACT

Analysis is a key part of conducting usability evaluations, yet rarely systematically studied. Thus, we lack direction on how to do research on supporting practitioners' analysis and lose an opportunity for practitioners to learn from each other. We have surveyed 155 usability practitioners on the analysis in their latest usability evaluation. Analysis is typically flexible and light-weight. At the same time, practitioners see a need to strengthen reliability in evaluation. Redesign is closely integrated with analysis; more than half of the respondents provide visual redesign suggestions in their evaluation deliverables. Analysis support from academic research, including tools, forms and structured formats, does not seem to have direct impact on analysis practice. We provide six recommendations for future research to better support analysis.

Author Keywords

Usability evaluation method; analysis; questionnaire survey; state-of-the-practice.

ACM Classification Keywords

H.5.m. [Information interfaces and presentation (e.g., HCI)]: Miscellaneous;

General Terms

Experimentation, Human Factors.

INTRODUCTION

Analysis in usability evaluation is challenging. Analysis is the process by which observations of users or inspections of interfaces are turned into prioritized, coherent descriptions of usability problems, including descriptions of causes, implications, and potential solutions [13]. Doing analysis is described as “the ultimate detective work” [32], involving multiple data sources [10], and highly dependent on practitioners' knowledge and expertise [9].

Despite these well-known challenges in analysis, there is a surprising lack of research on analysis practices. We are unaware of comprehensive studies of this topic, though many studies treat broader aspects of usability evaluation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1015-4/12/05...\$10.00.

practices, such as method preferences [3,35,36], or specific aspects, such as the practice of doing think aloud [4].

At the same time, introductory texts provide only high-level descriptions of analysis in usability evaluation. Texts on usability testing [10,32], as well as on heuristic evaluation [27,28], lack detail on how to do analysis. For instance, Rubin and Chisnell use only two pages of their well known handbook [32] to discuss user error identification and source of error analysis (pp. 260-261).

The lack of knowledge on analysis implies that researchers who develop methods and tools to support analysis in usability evaluation will find it hard to meet practitioners' needs. In turn, this may impede the transfer of methods and tools from research to practice [14] (and vice versa) and lead to a gap between research and practice [5].

This paper presents a survey of usability practitioners' analysis in their latest usability evaluation. We present data on how analysis is supported, how usability problems are identified, and how evaluation and redesign are integrated. Usability researchers may benefit by improved knowledge on what practitioners need from future methods and tools for analysis. Also, we hope that practitioners can learn from insights into the state of the practice, and use these in their professional development.

RELATED WORK

Analysis in Research

Analysis methods and tools

In spite of the lack of research-based knowledge of analysis practices, in the last two decades research has been conducted to develop methods and tools for analysis in usability evaluation. Below we discuss these.

Analysis process improvements have been suggested both as light-weight approaches, such as instant data analysis upon conclusion of usability testing sessions [23], and as formal procedures and structures, such as the user action framework (UAF) [1] and the framework for structured usability problems extraction (SUPEX) [7]. Concern has been voiced regarding the usability and learnability of the latter [22].

Problem description formats have been developed to facilitate usability problem merging, either as templates [8,24], guidelines [6], or integrated into software tools [21].

Research-based analysis procedures and problem description formats include guidelines or strategies for problem identification. These strategies are mainly task-oriented, using task performance as the key indicator for usability problems. However, user behavior and verbalization are acknowledged as separate data sources to complement task performance data.

Tools for problem identification and consolidation have been developed in the research communities as well as in industries. Research-based tools include software tools such as the Usability Problem Inspector [2], and simple conceptual tools such as Skov and Stage's tool to support problem identification and prioritization [33]. Commercial video analysis software such as Morae [26] includes facilities for usability problem description and grouping. Howarth, Smith-Jackson and Hartson provided evidence that novice analysts may benefit from having structured formats for problem description included in such software [21].

Collaboration in analysis

The research literature clearly shows the importance of collaboration between usability practitioners in evaluation. The use of multiple evaluators in problem identification and analysis has been found useful for improving the thoroughness [28] and reliability [17] of problem identification; that is increasing the likelihood that all real problems are found and improving consistency in analysis.

To serve the intended purpose, collaboration typically happens among two or more usability practitioners who perform independent analysis of the same data set [17,28].

The relation between evaluation and design

There is an ongoing debate in usability research on the relation between evaluation and design [15,20]. Some suggest an integration of evaluation and redesign, as in Rapid Iterative Testing and Evaluation (RITE), where fixes to usability problems with "an obvious cause and an obvious solution" are encouraged as part of the usability evaluation procedure [25], or by including ideas or redesign suggestions in the findings from a usability evaluation [12,18,19,31]. Others argue against such integration of evaluation and redesign. It is held that the findings from a usability evaluation should only drive the next iteration in the development process, requiring a "grounding of [the findings] in business goals, usage contexts, design principles, and design constraints" prior to considering redesign solutions [9].

Analysis in Practice

Our knowledge of general usability evaluation practices is extensive. In particular, survey studies provide information on evaluation method use and preferences among practitioners [3,35,36]. Both usability testing and inspection are widely applied, and informal usability inspection methods (expert evaluations) are used more frequently than inspection methods with higher degree of formality, such as cognitive walkthrough. A recent PhD-thesis provided in-

depth knowledge on usability evaluation practices, identifying contextual factors affecting the uptake of research-based evaluation methods [14].

To the best of our knowledge, only one prior study concern analysis practice in usability evaluation [13]. The study was presented as a pilot, and was based on interviews with 11 usability practitioners. It concluded that (a) analysis is informal and pragmatic with little use of structured formats, (b) professional experience is the most important analysis resource while resources like standards and guidelines are less important, (c) redesign suggestions seem to be an integrated part of analysis, and (d) collaboration on analysis is frequent, but mainly used to help identify more usability problems and redesign suggestions, rather than to reduce evaluator bias.

Other studies related to analysis practices are Boren and Ramey's [4] and Nørgaard and Hornbæk's [29] observational studies of usability practitioners during think-aloud testing sessions. These studies, however, offer only indirect data on analysis [29] or no data at all [4].

RESEARCH QUESTIONS

Previous research has provided methods and tools, knowledge on collaboration benefits, and perspectives on the relationship between evaluation and design. However, we lack knowledge about analysis practices and, consequently, knowledge about how the research-based methods and tools support such practices. Our research questions aim to generate such knowledge. The main question is: *What is the state of practice of analysis in usability evaluation?*

Based on this question and the related work, four sub-questions are asked:

- Q1. How is analysis supported?*
- Q2. How are usability problems identified?*
- Q3. How do usability practitioners collaborate in analysis?*
- Q4. How is redesign integrated into the evaluation process?*

METHOD

As analysis practices are expected to vary substantially among practitioners, it is important to get a broad sample of respondents. Consequently, we employed a web-based questionnaire survey. We regarded this methodological approach as adequate given that this study was to complement the existing knowledge generated through observation [29] and qualitative interviews [13]. Our choice of method is in line with previous survey studies of usability practices [3,36].

Recruiting Respondents

The survey respondents were required to be usability practitioners who had conducted a usability evaluation within the last six months. The time limitation was set to increase the likelihood of rich and accurate memory of the

analysis. The respondents were recruited through several channels; invitations were distributed via local SIGCHI chapters (151 respondents), local UPA chapters (31 respondents), the UTEST mailing list (19 respondents), the European COST project TwinTide (11 respondents), and fliers at CHI 2011 (2 respondents). As incentives to participate in the survey, all respondents leaving their e-mail address could get early access to the project results and were included in a lottery for a \$250 Amazon gift card.

Developing the Questionnaire

There were two versions of the questionnaire; one for respondents reporting on usability testing and one for usability inspection. Both versions were piloted with non-author respondents.

The first question of the questionnaire served to direct the respondents towards the *Usability testing* or the *Usability inspection* version of the questionnaire. The respondents were explicitly asked to report on their latest usability evaluation, and they were repeatedly reminded to focus on this latest evaluation throughout the questionnaire.

The questionnaire on usability testing consisted of 32 questions and that on usability inspection consisted of 29. Of these, 26 questions overlapped. The questions concerned background data (the practitioner and the evaluation context), perceived challenges, perceived impact, analysis support (notes, tools, resources, measures, and structured formats), usability problem identification strategy, collaboration, redesign, and prioritization of findings. Questions on note taking forms, tools, problem identification strategies, usability inspection methods, as well as the respondents’ educational background, were free text items. The other questions were fixed response items.

To avoid bias related to the order of response categories, fixed response alternatives were randomized for each question where the order bias was judged to be a threat.

Questions and response categories are quoted throughout the result section in italic font. The questionnaires are accessible at <http://tinyurl.com/7uegbwa>. Due to the space limitation, this paper only presents findings on analysis support, problem identification strategies, collaboration and redesign.

Respondents’ work context	UT (n=112)	UI (n=43)
<i>Consultancy or development projects for external clients</i>	40 %	63 %
<i>Inhouse development projects</i>	46 %	23 %
<i>Scientific research projects</i>	7 %	12 %
<i>Adaptation of commercial software</i>	1 %	0 %
<i>Other</i>	5 %	2 %

Table 1. Distribution of usability testing (UT) and usability inspection (UI) respondents across work contexts.

Respondents were informed that data collection was fully anonymous so as to avoid bias. Respondents who chose to leave their email address to receive the report on the study findings and participate in the lottery did so using an independent form; the e-mail addresses could not be connected to the questionnaire data.

Analysis

Quantitative analyses were descriptive, due to the exploratory aim of the study. Qualitative analyses of free text answers were conducted as thematic analysis [11]. Itemized free text answers related to Q2 (usability problem identification) were coded by two independent analysts to check agreement; free marginal *kappa* coefficients were in the range of .70 - .78 indicating adequate agreement [30].

Results are presented separately for usability testing and inspection throughout the paper except for those results where the difference between these two groups is less than 5 percentage points. In those cases the results for all the respondents are combined to compress the description. It could also have been relevant to break up results according to respondent background variables, such as work context or educational background. This, however, is not done in order to reduce complexity.

RESULTS

The respondents

In total, 224 people attempted the questionnaire. The responses from 155 of these were included in the analysis. Sixty-nine were excluded for (a) providing no response in any of free text fields in the questionnaire (40), (b) nonsense responses in the free text fields (3), (c) no response beyond the second page (the sixth item) of the questionnaire (6), and (d) not having conducted a usability evaluation in the last 6 months (20); the latter group left the questionnaire after the first question.

The majority of the respondents (62%) had started their latest usability evaluation within the last two months. They had a median experience of 5 years as usability practitioner (25th percentile = 3 years, 75th percentile = 11 years) and had conducted a median of 5 usability evaluations in the

Educational background	UT (n=112)	UI (n=43)
Computer science	19 %	19 %
HCI or Human factors	29 %	14 %
Psychology, behavioral science or cognitive science	13 %	19 %
Science and engineering	11 %	7 %
Media or communication	10 %	9 %
Other	18 %	32 %

Table 2. Distribution of respondents across education categories. Reported in free text and thematically coded.

last 12 months (25th percentile = 4, 75th percentile = 10). The respondents' typical work context and educational background are presented in Table 1 and Table 2.

The respondents worked in 21 countries: 55% in Europe, 36% in the US, 4.5% in Latin America and 4.5% in other parts of the world. In Europe, respondents worked in 14 countries. Three countries had remarkably large numbers of respondents: Switzerland and Czech Republic each had 15% of the respondents, and Poland had 10% of the respondents. The skew towards these three countries is likely the result of particularly effective dissemination of the survey locally. A check of responses from these countries showed no noteworthy difference to the others and so all responses were retained.

The Usability Evaluations

Usability testing

112 respondents reported a usability test as their latest usability evaluation. Usability testing was most typically conducted with 8 users (25th percentile = 5; 75th percentile = 12). The most frequently used measure was *Task completion* (84%), followed by *Satisfaction measures* (80%), *Error rate* (45%) and *Task time* (33%). Eighteen respondents reported *Other measures*, but the free-text explanations did not show any pattern.

Video recordings were made by 73% of the respondents. Half of the respondents also reviewed the video recordings after the evaluation, either in part (34 %) or in full (19 %).

Besides, 43% of the respondents reported that key findings of the test were summarized immediately following each test participant, whereas 24% reported that such summarizing was done at the end of the day. About half of these summarizing exercises were done with observers. 14% reported no summarizing; neither after each test participant nor at the end of the day.

Usability inspection

43 respondents reported a usability inspection as their latest usability evaluation. The respondents classified their inspection method as either *Heuristic evaluation* (30%), *Informal expert evaluation* (26%), *No particular method / mix of methods* (21%), *Cognitive walkthroughs* (14%) or *Other methods* (9%; including “compliance check with internal standards”, “interviews in the field”, “participatory

Development phase	UT (n=112)	UI (n=43)
Concepts	5 %	19 %
Early prototypes	23 %	21 %
Advanced prototypes	43 %	12 %
Running systems	30 %	49 %

Table 3. Distribution of respondents across the development phase of the system they evaluated.

workshops”, and “usability-workshop”). The third and the last group reflect the trend described by Rosenbaum [31] towards the adoption of a broader spectrum of methods in evaluation.

Thirty-four respondents provided free-text descriptions of their inspection method. The descriptions of the methods from the literature (*Heuristic evaluation* and *Cognitive walkthrough*) indicate flexible use. For heuristic evaluation, three respondents described combinations of heuristic evaluation and other methods, and one described an inspection process including ad-hoc construction of heuristics. For cognitive walkthrough, only two of five respondents clearly presented tasks as the starting point of the analysis, and one of these respondents also described a combined walkthrough and heuristic evaluation approach.

Similarly, the descriptions of the other inspection methods (*Informal expert evaluations* and *No particular method / mix of methods*) indicated variation and flexible combination of elements from different methods. Respondents detailing their *Informal expert review* mentioned “walkthrough”, “team discussion”, “comparison study”, “checklist”, and “business field study”. The respondents with *No particular method / mix of methods* described combined approaches such as heuristics plus walkthrough, heuristics plus best practice, and heuristics plus some form of expert review.

The evaluation context

For usability testing the median time spent on preparing, conducting and reporting the test was 48 working hours (25th percentile = 24; 75th percentile = 80), for usability inspection the median time spent was 24 working hours (25th percentile = 10; 75th percentile = 40). Most of the respondents claimed that these were typical time constraints (UT: 81%; UI: 72%).

The majority of the evaluations were conducted on mature systems or prototypes. However, early phase evaluations were also fairly frequent. Details are given in Table 3.

Note taking structure	UT	UI
General topics and questions	8	2
Tasks from scenarios of use	8	-
Test script or protocol	6	-
Checkpoints or checklists	4	2
UI location	-	2
Heuristics	-	2
Sources of observation, including participant behavior, participant comment, and observer comment	2	-

Table 4. Structuring principles for note taking forms with number of respondents making free text report of each.

How is Analysis Supported? (Q1)

Notes, tools, and problem description formats

Notes were used during analysis by all the respondents. About one-third took *Notes on simple forms* (UT: 22%; UI: 42%), and less than one-fifth took *Notes on detailed forms* (UT: 18%; UI: 15%). Fewer than half reported using any kind of form to structure note taking. For this question, multiple answers were allowed.

The respondents who used forms for note taking were asked to describe these in free text. We subsequently coded the text through thematic analysis. The reported types of structuring principles are presented in Table 4.

In addition, seven usability-testing respondents reported to include scales for task performance; four to include fields for notes on task time. Six participants reported use of the note functionality in their video analysis software.

Tools used for *problem description and analysis* were reported as free text answers. Spread sheets, text editors and presentation tools were frequently mentioned. More specialized tools are presented in Table 5.

Structured formats for usability problem description were used by the majority of the respondents (n=141). The structured formats used were, however, mainly home grown: 55% of the respondents reported that *The problems were described according to my/our own format*. Only 4% reported that *The problems were described according to a structured format described in standards or the literature*. This latter finding is highly interesting, given the attention and persistent effort ascribed to structured formats in usability research [8,24].

The remaining 41% reported that *The problems were described in plain prose*. No noteworthy difference was found between usability testing and inspection respondents.

Tools	UT	UI
Screen recording and analysis software including Morae, Silverback, and BeGaze	11	-
Screen recording tools including Camtasia, SnagIt, and Snap'nDrag	5	-
Eye tracking tools including Tobii	2	-
Drawing and prototyping tools including Balsamiq, Axure, Visio, Photoshop and Inkscape	2	6
Web analytics facilities including Google analytics, Seevolution and Site Catalyst	-	5

Table 5. Tools for problem description and analysis, with number of respondents making free text report of each.

Resources for knowledge and insight

The respondents were asked which resources they had used during analysis. Details are presented in Table 6.

Free text elaborations of the *Other* category (provided by 6 user testing and 6 inspection respondents) indicated two additional resource categories: Knowledge and skills from particular domains (mobile computing, visual design, psychology) (5) and input from clients (3).

Resources specific to usability testing

Participating users were a frequently consulted resource for usability testing respondents (n=106); 80% of the respondents reported that they had asked test participants for opinions on either *Usability problems* (64%), *Redesign suggestions* (48%), or *Other issues* (28%). Other issues, specified in free text, concerned:

- Acceptance and preferences (reported by 9; including subjective feedback or ratings)
- Utility and applicability (reported by 3; related to e.g. “fit in context” and “use cases”).
- Workflow (reported by 3; targeting the “process flow” or “workflow” in the object of evaluation)
- Contextual information (reported by 2; on “culture and uptake” and “contextual info [on] problems or redesign”).

Observers, which could include clients or colleagues, were used as a resource by 59% of the usability testing respondents (n=106), by asking their opinion on *Usability problems* (48%), *Redesign suggestions* (53%), or *Other issues* (27%). None of the *Other issues* was reported more than once. As we did not ask the respondents whether observers were present during the tests, it is likely that some did not report to seek observer opinion simply because no observer was present.

Analysis resources	UT (n=107)	UI (n=41)
<i>My professional experience</i>	90 %	95 %
<i>Heuristics and guidelines</i>	60 %	76 %
<i>Design patterns</i>	41 %	54 %
<i>Standards (norm or technical requirement established within a given field or application area)</i>	40 %	42 %
<i>Previously established personas</i>	18 %	27 %
<i>Previously established scenarios of use</i>	NA	39 %
<i>Other</i>	5 %	2 %

Table 6. Distribution of respondents across analysis resources. Multiple answers allowed. (NA = not available)

How are Usability Problems Identified? (Q2)

Rubin and Chisnell described the identification of the sources of user error, that is, the usability problems, as “the ultimate detective work” [32]. Knowing how practitioners perform this detective work is critical to understand analysis. The respondents were asked for *A brief description of how you decided something being a usability problem.*

Problem identification in usability testing

Ninety-three usability testing respondents provided descriptions of their problem identification strategy; from these, 214 items were extracted for thematic analysis. Two main strategies were related to (a) observed consequences for task performance and (b) the users’ responses.

Observed consequences for task performance were reported in 32% of the items. Task (in)completion was, not surprisingly, most frequently reported (32 items), followed by Other consequences for task performance (24 items), and Task completion time (12 items). Other consequences for task performance included “user error”, “deviations from the expected path”, and “users misjudging their status as completed / not completed”.

The users’ responses, irrespective of task performance, were reported as indicative of usability problems in 19% of the items. Observations included Emotional responses such as ‘boredom’, ‘confusion’, ‘frustration’, ‘impatience’ and ‘reduced engagement’ (18 items), and Behavioral responses such as ‘hesitation’, ‘seeking help’, ‘struggling’, ‘giving up’, and ‘discomfort’ (9 items). Fourteen items concerned Verbal responses collected through think aloud.

The remaining items concerned strategies related to Frequency and severity (37 items), Causal explanations (38 items), Usability expertise (18 items), and Other approaches (11 items), such as the use of rating scales, and web analytics. One item was coded as incomprehensible.

Problem identification in usability inspection

Thirty-two usability inspection respondents answered this question; 61 items were extracted for thematic analysis.

When describing strategies for problem identification in usability inspection, the respondents typically referred to the usability expertise involved in the inspection (50% of the items). Expertise included professional experience / expert knowledge and mindset (18 items), and the use of resources such as heuristics, standards, specifications, checklists, best practices, design patterns, and research literature (12 items).

Consequences for task performance (8 items) were mentioned more often than consequences for the users’ responses (2 items), but much more rarely than for the usability testing respondents reported above.

The remaining items targeted Causal explanations (4 items) and Other approaches (11 items), such as benchmarking comparison and web analytics. Four items discussed general issues or were coded Incomprehensible.

Types of collaboration	UT (n=102)	UI (n=40)
<i>Short discussions immediately following user test sessions / startup of the inspection</i>	58 %	40 %
<i>Review of draft report</i>	54 %	43 %
<i>Data analysis / inspection conducted as a group activity by two or more usability professionals</i>	29 %	30 %
<i>Individual analysis of data from the same user sessions / the same parts of the system by myself and at least one colleague</i>	23 %	43 %
<i>Individual analysis of data from different user sessions / different parts of the system by myself and at least one other colleague</i>	16 %	18 %
<i>Guidance when conducting analysis</i>	20 %	15 %

Table 7. Distribution of respondents across types of peer-collaboration. Multiple answers allowed.

How is Collaboration Practiced? (Q3)

Collaboration with fellow usability practitioners was reported by most respondents (n=142); only 26% reported *No collaboration*. The reported types of collaboration are presented in Table 7.

Collaboration purposes were typically reported as *Improve reliability / avoid that the findings were biased by personal perspective* (UT: 46%; UI: 30%), *Generate better redesign suggestions* (UT: 17%; UI: 30%), and *Quality assurance* (UT: 13%; UI: 30%). Less frequent purposes were to *Identify more problems* (UT: 12%; UI: 3%) and *Other* (UT: 11%; UI: 7%). This practitioner awareness of reliability is in marked contrast to previous research indicating a lack of practitioner attention to reliability as a rationale for deploying multiple evaluators [13].

Integrating Redesign in the Evaluation Process (Q4)

The respondents characterized their evaluation deliverable with respect to the inclusion of redesign suggestions and usability problems (n=141). Redesign suggestions were included in the deliverables of nearly all the respondents. Half of the respondents described their deliverable as *A set of redesign suggestions either in response to a set of usability problems or in part motivated from usability problems*. Details are presented in Table 8.

The respondents were divided on whether or not all usability problems were identified prior to generating redesign suggestions (n=105). 49% reported that *First all usability problems were identified, then the redesign suggestions were made*. 46% reported that some (37%) or all (9%) redesign suggestions were generated immediately when a usability problem was identified. 5% chose the

Deliverable characterized as ...	UT (n=101)	UI (n=40)
<i>A set of redesign suggestions in part motivated from usability problems</i>	12 %	23 %
<i>A set of redesign suggestions in response to a set of usability problems</i>	39 %	30 %
<i>A set of usability problems with some redesign suggestions</i>	46 %	43 %
<i>A set of usability problems with no redesign suggestions</i>	4 %	5 %

Table 8. Distribution of respondents across evaluation deliverable characteristics.

Other response alternative. No noteworthy differences between usability testing and inspection respondents were detected.

The respondents also answered how they *Reached the redesign suggestions that were included in the deliverable*. Redesign suggestion were typically provided in response to usability problems. However, more than one-third reported to reach redesign suggestions even though no usability problem had been observed. Details are presented in Table 9.

The respondents were also asked how they *present redesign suggestions in the deliverable*. More than half of the respondents provided visual redesign suggestions such as UI-digital mock-ups, sketching, and graphical elements / code (UT: 57%; UI: 68%). The others provided redesign suggestions as oral presentation only, plain text only, or text supported by screen shots. Oral presentations were frequent, but only six respondents reported this as their only means of redesign presentation. Table 10 presents details.

This integration of evaluation and design was perhaps the most surprising finding in the study, and an important

Sources of redesign suggestions	UT (n=67)	UI (n=38)
<i>In response to usability problems</i>	94 %	74 %
<i>On basis of recognition that the solution was not optimal even though no usability problem had been observed</i>	38 %	47 %
<i>Test participant suggestions</i>	41 %	NA
<i>Observer suggestions</i>	33 %	NA
<i>Colleague or peer suggestions</i>	NA	40 %
<i>Client suggestions</i>	15 %	18 %

Table 9. Distribution of respondents across sources of redesign suggestions. Multiple answers allowed.

contribution for moving forward the discussion on evaluation and design in usability research [9,20].

DISCUSSION

General findings

The evaluation context: Practitioners appropriate and adapt
One key finding of the survey concerns the flexibility of method use. A variety of methods is employed across process phases and project types, and a range of adaptations and method combinations is reported. Web analytics and tools for drawing and prototyping are flexibly integrated in usability evaluation, in particular in usability inspection.

Appropriation and adaptation is also evident in the use of analysis resources. *Heuristics or guidelines, Design patterns, and Standards* are almost as frequently used in usability testing as in inspection. This somewhat surprising finding indicates the value of general knowledge resources, not only to compensate for lack of user-based empirical data, but also to understand and organize such data.

The prevalence of seeking test participants’ opinions on usability problems and redesign suggestions also seem to represent a pragmatic exploitation of the available resources, in this case the knowledge of the test participant. As data from usability testing are challenging to interpret, test participants’ knowledge is valuable to explain observations and support analysis.

The observed appropriation and adaptation indicates that practitioners use existing methods and resources as components to be combined in response to the requirements of a given evaluation. These findings support practitioners [31] and researchers [38] who argue that we need to support combining and adapting methods better.

Analysis support: Home-grown and commercially available
The reported uses of notes, tools, and problem description formats indicate that practitioners tend to use commercial or home-grown support, rather than support from usability research. Usability research on forms, formats, and tools

Means of redesign presentation	UT (n=66)	UI (n=38)
<i>Textual descriptions</i>	68 %	71 %
<i>Annotated screen shots</i>	50 %	55 %
<i>UI digital mock-ups</i>	32 %	47 %
<i>Sketching</i>	29 %	21 %
<i>Graphical elements or code that could be immediately implemented</i>	5 %	2 %
<i>Oral presentation</i>	53 %	34 %
<i>Other</i>	13 %	8 %

Table 10. Distribution of respondents across means of presenting redesign suggestions. Multiple answers allowed.

does not seem to have much direct impact on analysis practice; if such impact exists at all, it is likely to be restricted to the development of home-grown and commercial analysis support.

Possibly, this preference for commercially available and home-grown support may be attributed to the fact that tools developed by the research community tend to be complex (as is arguably the case for SUPEX and UAF) and difficult to learn. Complexity is not compatible with the time demands of the practical evaluation context; when a usability test is to be completed in 48 hours, and an inspection in 24, supporting tools clearly need to be light weight. For new analysis support to be successful, it will have to fit the fast-paced analysis context.

Problem identification: Task performance and users' responses

For usability testing, two ways of problem identification in analysis were identified: (a) Issues that compromise task performance and (b) issues identified from users' responses. In the research literature, data on users' responses are typically used to complement task performance data. However, the importance given to the users' responses may also reflect the emphasis on user experience in HCI in the last decade [16] where user experience may be compromised even though task performance in a usability test is found satisfactory.

For usability inspection, problem identification typically is grounded in general usability knowledge, rather than in speculations on users' task performance or responses. Interestingly, this expertise-based approach to problem identification is also clearly present in usability testing: 38% of the usability testing respondents reported to make redesign suggestions even though no usability problem had been observed. It seems as if problem identification practices from usability inspection are being used also in usability testing. This is in line with the argument that practitioners view methods as components to be combined and appropriated as needed.

Peer-collaboration: Early discussions, group analysis, and awareness of reliability

The drive for efficient analysis is evident also in peer-collaboration. The most frequent form of collaboration is short discussions at the outset of analysis. At the same time, there is substantial awareness on collaboration as a means to improve reliability. Nearly half of the usability-testing respondents reported that improved reliability was their main purpose for peer-collaboration. An unanswered question, however, is whether the collaboration done in analysis actually serves to improve reliability. Research suggests that reliability is improved by independent analysis of the same data set [17], something that was reported by only one-fourth of the usability testing respondents.

The findings on purposes of peer-collaboration partially go against the findings of Følstad et al. [13], who suggested

that the identification of more problems, as well as the generation of better redesign suggestions, were key drivers for such collaboration, rather than improved reliability. Their suggestions on (a) low awareness of reliability and (b) collaboration mainly motivated by an intent to increase the pool of problems are not corroborated.

A surprising finding was the prevalence of group discussions in analysis for both usability testing and inspection. In usability inspection, group discussion following individual assessments is a recommended practice [28,37]. However, group analysis of usability testing data is rarely emphasized in the literature.

Integrating redesign: Ever more relevant

The results clearly support the current trend within usability research to strengthen the integration of evaluation and redesign. An evaluation deliverable without redesign suggestions is reported by a minority of respondents only, and almost half of the respondents provide redesign suggestions immediately upon problem identification for at least some of the usability problems. The research community seems to be overdue in providing practitioners with facilities to support redesign as part of the evaluation process.

Redesign suggestions can take many forms. In the literature, change recommendations are typically perceived as textual, though exceptions exist [34]; the recommended approach to solving an identified usability problem is to present the solution in prose [9,10], possibly with the support of screenshots [32]. Quite surprisingly, then, we find that more than half of the respondents report visual presentations of redesign suggestions, such as sketches, mock-ups, and graphical elements. Usability practitioners seem to have advanced further than both the current literature and the research field in their integration of redesign and evaluation, indicating a potential for transfer of knowledge from practice to research.

Implications

On basis of the above findings, we suggest six implications for future usability research.

1. Develop and study method components

The flexible use of evaluation methods and analysis support indicates a need to move from usability research on methods to research on method components, as was recently suggested within usability research [38]. Method components may, for example, concern task-scenario development, participant recruitment, problem identification, problem consolidation, and redesign. Research is needed to establish a basic taxonomy of method components, identify the components in most need of strengthening, as well as provide empirical knowledge on their strengths and weaknesses for typical evaluation contexts.

2. Support light-weight analysis

As opposed to analysis in usability research, where it is hard to get a paper accepted without having two independent analysts review the video material from the

evaluations, analysis in usability practice typically need to be light-weight. Research is needed on support for fast and effective analysis, following the lead of Kjeldskov, Skov and Stages' instant data analysis [23]. New approaches to analysis should involve analysis support already used by the practitioner, ideas include structuring forms for note taking to better support subsequent analysis, and improving post-task data collection of test participants' opinions.

3. Align research with commercial software

Practitioners in our study reported only using commercially available software tools for analysis. As a consequence, researchers should consider aligning their research with commercial software, for example, by developing plug-ins to existing software, as done by Howarth et al. [21] who developed and studied a plug-in to the Morae video analysis software.

4. Support home-growing

Forms and structured formats are widely used, but they are typically home-grown, rather than retrieved from the literature. In consequence, researchers should consider supporting home-growing of simple analysis support, rather than to provide fixed forms and templates. A good example of support for home-growing is Capra's guidelines for usability problem description, which suggest important description elements without presenting a strict template [6].

5. Explore analysis in groups

Group analysis is prevalent in usability testing as well as inspection. Such analysis is not well understood for usability testing and its strengths and weaknesses, as well as guidelines for practice, need to be established. Also, processes for group analysis could be established so as to improve reliability, for example, by presenting brief individual summaries prior to the group interaction – learning from the oscillation between individual analysis and group discussion recommended by Nielsen [28].

6. Facilitate the integration of evaluation and design

The integration of evaluation and design has been an important topic in the research literature of the last decade. Even so, it seems as if usability practitioners are ahead of researchers in bridging the gap between evaluation and design. Research is needed to refine emerging approaches among leading practitioners into sustainable practices. In a recent doctoral thesis, Furniss argues for the need to study usability evaluation in context in order to facilitate transfer of research results to usability practice [14]. In the case of the integration of evaluation and design, research on evaluation in context is also needed in order to identify method innovations as they happen in practice.

Limitations

The main limitation of the study is that it was conducted as a questionnaire survey: what respondents say may differ from what they do. In particular, forgetfulness and a bias towards providing socially desirable responses pose threats to questionnaire studies. We addressed these threats by (a) targeting one single usability evaluation rather than general

practice, (b) accepting only respondents that had conducted a usability evaluation during the last six months, and (c) providing the respondents full anonymity.

However, the questionnaire survey design is also a strength of the study as it enables data collection from a broad spectrum of usability practitioners. This would not be practically possible with methods supporting data collection on the respondents' behavior. Given the broad variation in current practices identified in this study, our choice of a survey approach to data collection appears supported.

Conclusion and future work

Our findings paint a picture of usability evaluation where analysis practices evolves as practitioners adapt and appropriate tools and methods to the needs of their evaluation context. The picture, however, is incomplete. In particular, the findings of the current study should be complemented with observational studies and case studies of actual analysis. We hold that usability research may provide valuable contributions to the future evolution of analysis practices. For this to happen, however, we need to a much larger extent to inform our research efforts by the needs of usability practitioners.

ACKNOWLEDGMENTS

This study was done as part of the R2D2 Networks project, supported by the Norwegian research council's VERDIKT programme, and by the COST Action TwinTide, IOC0904, supported by the European Commission. Thanks to all who participated as respondents and to all who helped us distribute the invitations to participate.

REFERENCES

1. Andre, T.S., Hartson, H.R., and Williges, R.C. The user action framework: a reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies* 54, 1 (2001), 107-136.
2. Andre, T.S., Hartson, H.R., and Williges, R.C. Determining the effectiveness of the usability problem inspector: a theory-based model and tool for finding usability problems. *Human Factors* 45, 3 (2003), 455-482.
3. Bark, I., Følstad, A., Gulliksen, J. Use and Usefulness of HCI methods. Results From an Exploratory Survey among Nordic HCI Practitioners. In *Proc. HCI 2005*, Springer (2006), 200-217.
4. Boren, M., and Ramey, J. Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication* 43, 3 (2000), 1-23.
5. Buie, E., Dray, S., Instone, K., Jain, J., Lindgaard, G., and Lund A. How to bring HCI research and practice closer together? In *Proc. CHI 2010 extended abstracts*, ACM Press (2010), 3181-3184.
6. Capra, M. *Usability Problem Description and the Evaluator Effect in Usability Testing*. PhD thesis. Virginia Tech, Blacksburg, VA, 2006.

7. Cockton, G., and Lavery, D. A Framework for Usability Problem Extraction, In *Proc. of INTERACT '99*, IOS Press (1999), 344-352.
8. Cockton, G., Woolrych, A., and Hindmarch, M. Reconditioned Merchandise: Extended Structured Report Formats in Usability Inspection. In *Proc. CHI 2004 Extended Abstracts*, ACM Press (2004), 1433-1436.
9. Cockton, G., Lavery, D. and Woolrych, A. Inspection-based Evaluations. In A. Sears and J. Jacko (eds.), *The Human-Computer Interaction Handbook*, 2nd edition. Lawrence Erlbaum Associates (2008), 1171-1190.
10. Dumas, J.S., and Redish, J. *A practical guide to usability testing*. Ablex Publishing Corporation, 1993.
11. Ezzy, D. *Qualitative Analysis: Practice and Innovation*. Routledge, 2002.
12. Følstad, A., and Hornbæk, K. Work-domain knowledge in usability evaluation: Experiences with Cooperative Usability Testing. *Journal of Systems and Software* 83, 11 (2010), 2019-2030
13. Følstad, A., Law, E.L-C., Hornbæk, K. Analysis in Usability Evaluations: an Exploratory Study. In *Proc. NordiCHI 2010*, ACM Press (2010), 647-650.
14. Furniss, D. *Beyond Problem Identification: Valuing methods in a 'system of usability practice'*. PhD thesis. University College London, London, UK, 2008.
15. Gulliksen, J, Boivie, I., and Göransson, B. Usability professionals - current practices and future development. *Interacting with Computers* 18, 4 (2006), 568-600.
16. Hassenzahl, M., and Tractinsky, N. User Experience – a research agenda. *Behavior & Information technology* 25, 2 (2006), 91-97
17. Hertzum, M., and Jacobsen, N. E. The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction* 15, 1, Lawrence Erlbaum Associates (2003), 183-204.
18. Hornbæk, K. Usability Evaluation as Idea Generation. In E. L-C. Law, E.T. Hvannberg, and G. Cockton (eds.) *Maturing Usability: Quality in Software, Interaction and Value*, Springer (2008), 267-286.
19. Hornbæk, K., and Frøkjær, E., Comparing usability problems and redesign proposals as input to practical systems development. In *Proc. CHI 2005*, ACM Press (2005), 391-400.
20. Hornbæk, K., and Stage, J. The Interplay between Usability Evaluation and User Interaction Design - Introduction to Special Issue. *International Journal of Human-Computer Interaction* 22, 2 (2006), 117-123.
21. Howarth, J., Smith-Jackson, T., and Hartson, R. Supporting novice usability practitioners with usability engineering tools. *International Journal of Human-Computer Studies* 67, 6 (2009), 533–549.
22. Hvannberg, E. T., Law, E. L-C., and Lárusdóttir, M. Heuristic Evaluation: Comparing Ways of Finding and Reporting Usability Problems. *Interacting with Computers* 19 (2007), 225-240.
23. Kjeldskov J., Skov M. B., and Stage J. Instant Data Analysis: Evaluating Usability in a Day. In *Proc. NordiCHI 2004*, ACM Press (2004), 233-240.
24. Lavery, D., Cockton, G., and Atkinson, M.P. Comparison of Evaluation Methods Using Structured Usability Problem Reports. *Behaviour and Information Technology* 16, 4 (1997), 246-266.
25. Medlock, MC, Wixon, D., Terrano, M., Romero, R.L. Fulton, B. Using the RITE method to improve products; a definition and a case study. Presented at UPA 2002.
26. Morae. <http://www.techsmith.com/morae.asp>
27. Nielsen J., and Molich, R. Heuristic evaluation of user interfaces. In *Proc. CHI 1990*, ACM Press (1990), 249-256.
28. Nielsen, J. Heuristic Evaluation. In J. Nielsen and R.L. Mack (eds.) *Usability Inspection Methods*, John Wiley and Sons (1994), 25-62.
29. Nørgaard, M., and Hornbæk, K. What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing. *Proc. of DIS 2006*, ACM Press (2006), 209-218
30. Randolph, J.J. Online Kappa Calculator. <http://justus.randolph.name/kappa>
31. Rosenbaum, S. The Future of Usability Evaluation: Increasing Impact on Value. In E. L-C. Law, E.T. Hvannberg, and G. Cockton (eds.) *Maturing Usability: Quality in Software, Interaction and Value*, Springer (2008), 344-380.
32. Rubin, J. and Chisnell, D. *Handbook of usability testing* (2nd.edition). Wiley Publishing, 2008.
33. Skov, J., and Stage, J. *Supporting Problem Identification in Usability Evaluations*. Paper presented at OzCHI 2005.
34. Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. User sketches: a quick, inexpensive, and effective way to elicit more reflective user feedback. In *Proc. NordiCHI 2006*, ACM Press (2006), 105-114
35. UPA 2009 Salary Survey. http://www.usabilityprofessionals.org/usability_resources/surveys/2009salarysurvey_PUBLIC.pdf
36. Vredenburg, K., Mao, J., Smith, P. W., and Carey, T. A Survey of User-centered Design Practice. In *Proc. CHI'02*, ACM Press (2002), 471-478.
37. Wharton, C., Rieman, J., Lewis, C., and Polson, P. The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen and R.L. Mack (eds.) *Usability Inspection Methods*, John Wiley and Sons (1994), 105-140.
38. Woolrych, A. Hornbæk K. Frøkjær, E. and Cockton, G. 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 (2011), 940-970.