

## RESEARCH

## Open Access



# Users' design feedback in usability evaluation: a literature review

Asbjørn Følstad\* 

\*Correspondence:  
asf@sintef.no  
SINTEF, Forskningsveien 1,  
0373 Oslo, Norway

## Abstract

As part of usability evaluation, users may be invited to offer their reflections on the system being evaluated. Such reflections may concern the system's suitability for its context of use, usability problem predictions, and design suggestions. We term the data resulting from such reflections *users' design feedback*. Gathering users' design feedback as part of usability evaluation may be seen as controversial, and the current knowledge on users' design feedback is fragmented. To mitigate this, we have conducted a literature review. The review provides an overview of the benefits and limitations of users' design feedback in usability evaluations. Following an extensive search process, 31 research papers were identified as relevant and analysed. Users' design feedback is gathered for a number of distinct purposes: to support budget approaches to usability testing, to expand on interaction data from usability testing, to provide insight into usability problems in users' everyday context, and to benefit from users' knowledge and creativity. Evaluation findings based on users' design feedback can be qualitatively different from, and hence complement, findings based on other types of evaluation data. Furthermore, findings based on users' design feedback can hold acceptable validity, though the thoroughness of such findings may be questioned. Finally, findings from users' design feedback may have substantial impact in the downstream development process. Four practical implications are highlighted, and three directions for future research are suggested.

**Keywords:** Usability evaluation, Users' design feedback, User reports, Literature review

## Introduction

Involving users in usability evaluation is valuable when designing information and communication technology (ICT), and a range of usability evaluation methods (UEM) support user involvement. Relevant methods include adaptations of usability testing [1], usability inspection methods such as pluralistic walkthrough [2], and inquiry methods such as interviews [3], and focus groups [4].

Users involved in usability evaluation may generate two types of data. We term these *interaction data* and *design feedback*. Interaction data are recordings of the actual use of an interactive system, such as observational data, system logs, and data from think-aloud protocols. Design feedback are data on users' reflections concerning an interactive system, such as comments on experiential issues, considerations of the system's suitability for its context of use, usability problem predictions, and design suggestions.

The value of interaction data in evaluation is unchallenged. Interaction data is held to be a key source of insight in the usability of interactive systems and has been the object of thorough scientific research. Numerous empirical studies concern the identification of usability problems on the basis of observable user behaviour [5]. Indeed, empirical UEM assessments are typically done by comparing the set of usability problems identified through the assessed UEM with a set of usability problems identified during usability testing (e.g. [6, 7]).

The value of users' design feedback is, however, disputed. Nielsen [8] stated, as a first rule of usability, "don't listen to users" and argued that users' design feedback should be limited to preference data after having used the interactive system in question. Users' design feedback may be biased due to a desire to report what the evaluator wants to hear, imperfect memory, and rationalization of own behaviour [8, 9]. As discussed by Gould and Lewis [10], it can be challenging to elicit useful design information from users as they may not have considered alternative approaches or may be ignorant of relevant alternatives; users may simply be unaware of what they need. Furthermore, as discussed by Wilson and Sasse [11], users do not always know what is good for them and may easily be swayed by contextual factors when making assessments.

Nevertheless, numerous UEMs that involve the gathering and analysis of users' design feedback have been suggested (e.g. [12–14]), and textbooks on usability evaluations typically recommend gathering data on users' experiences or considerations in qualitative post-task or post-test interviews [1, 15]. It is also common among usability practitioners to ask for the opinion of the participants in usability testing pertaining to usability problems or design suggestions [16].

Our current knowledge of users' design feedback is fragmented. Despite the number of UEMs suggested to support the gathering of users' design feedback, no coherent body of knowledge on users' design feedback as a distinct data source has been established. Existing empirical studies of users' design feedback typically involve the assessment of one or a small number of UEMs, and only to a limited degree build on each other. Consequently, a comprehensive overview of existing studies on users' design feedback is needed to better understand the benefits and limitation of this data source in usability evaluation.

To strengthen our understanding of users' design feedback in usability evaluation we present a review of the research literature on such design feedback.<sup>1</sup> Through the review, we have sought to provide an overview the benefits and limitations of users' design feedback. In particular, we have investigated users' design feedback in terms of the purposes for which it is gathered, its qualitative characteristics, its validity and thoroughness, as well as its downstream utility.

Our study is not an attempt to challenge the benefit of interaction data in usability evaluation. Rather, we assume that users' design feedback may complement other types

---

<sup>1</sup> The review is based on the author's Ph.D. thesis on users' design feedback, where it served to position three studies conducted by the authors relative to other work done within this field. The review presented in this paper includes these three studies as they satisfy the inclusion criteria for the review. It may also be noted that, to include a broader set of perspectives on the benefits and limitations of users' design feedback, the inclusion criteria applied in the review presented here is more relaxed compared to that of the Ph.D. thesis. The thesis was accepted at the University of Oslo in 2014.

of evaluation data, such as interaction data or data from inspections with usability experts, thereby strengthening the value of involving users in usability evaluation.

The scope of the study is delimited to qualitative or open-ended design feedback; such data may provide richer insight into the potential benefits and limitations of users' design feedback than do quantitative or set-response design feedback. Hence, design feedback in the form of data from set-response data gathering methods, such as standard usability questionnaires, are not considered in this review.

## Background

### Users' design feedback

In usability evaluation, users may engage in interaction and reflection. During interaction the user engages in behaviour that involves the user interface of an interactive system or its abstraction, such as a mock-up or prototype. The behaviour may include think-aloud verbalization of the immediate perceptions and thoughts that accompany the user's interaction. The interaction may be recorded through video, system log data, and observation forms or notes. We term such records interaction data. Interaction data is a key data source in usability testing and typically leads to findings formulated as usability problems, or to quantitative summaries such as success rate, time on task, and number of errors [1].

During reflection, the user engages in analysis and interpretation of the interactive system or the experiences made during system interaction. Unlike the free-flowing thought processes represented in think-aloud data, user reflection typically is conducted after having used the interactive system or in response to a demonstration or presentation of the interactive system. User reflection can be made on the basis of system representations such as prototypes or mock-ups, but also on the basis of pre-prototype documentation such as concept descriptions, and may be recorded as verbal or written reports. We refer to records of user reflection as design feedback, as their purpose in usability evaluation typically is to support the understanding or improvement of the evaluated design. Users' design feedback often lead to findings formulated as usability problems, (e.g. [3, 17]), but also to other types of findings such as insight into users' experiences of a particular design [18], input to user requirements [19], and suggestions for changes to the design [20].

What we refer to as *users' design feedback* eclipses what has been termed *user reports* [9], as its scope includes data on user' reflections not only from inquiry methods but also from usability inspection and usability testing.

### UEMs for users' design feedback

The gathering and analysis of users' design feedback is found in all the main UEM groups, that is, usability inspection methods, usability testing methods, and inquiry methods [21].

Usability inspection, though typically conducted by trained usability experts [22], is acknowledged to be useful also with other inspector types such as "end users with content or task knowledge" [23]. Specific inspection methods have been developed to involve users as inspectors. In the pluralistic walkthrough [2] and the participatory heuristic evaluation [13] users are involved in inspection groups together with usability

experts and developers. In the structured expert evaluation method [24] and the group-based expert walkthrough [25] users can be involved as the only inspector type.

Several usability testing methods have been developed where interaction data is complemented with users' design feedback, such as cooperative evaluation, cooperative usability testing, and asynchronous remote usability testing. In the cooperative evaluation [14] the user is told to think of himself as a co-evaluator and encouraged to ask questions and to be critical. In the cooperative usability testing [26] the user is invited to review the task solving process upon its completion and to reflect on incidents and potential usability problems. In asynchronous remote usability testing the user may be required to self-report incidents or problems, as a substitute of having these identified on the basis of interaction data [27].

Inquiry methods typically are general purpose data collection methods that have been adapted to the purpose of usability evaluation. Prominent inquiry methods in usability evaluation are interviews [3], workshops [28], contextual inquiries [29], and focus groups [30]. Also, online discussion forums have been applied for evaluation purposes [17]. Inquiry methods used for usability evaluation are generally less researched than methods for usability inspection methods and usability testing [21].

#### **Motivations for gathering users' design feedback**

There are two key motivations for gathering design feedback from users: users as a source of knowledge and users as a source of creativity.

Knowledge of a system' context of use is critical in design and evaluation. Such knowledge, which we in the following call domain knowledge, can be a missing evaluation resource [22]. Users have often been pointed out as a possible source of domain knowledge during evaluation [12, 13]. Users' domain knowledge may be most relevant for usability evaluations in domains requiring high levels of specialization or training, such as health care or gaming. In particular, users' domain knowledge may be critical in domains where the usability expert cannot be expected to have overlapping knowledge [25]. Hence, it may be expected that the user reflections that are captured in users' design feedback are more beneficial for applications specialized to a particular context of use than for applications with a broader target user group.

A second motivation to gather design feedback from users is to tap into their creative potential. This perspective has, in particular, been argued within participatory design. Here, users, developers, and designers are encouraged to exchange knowledge, ideas, and design suggestions in cooperative design and evaluation activities [31]. In a survey of usability evaluation state-of-the-practice, Følstad, Law, and Hornbæk [16] found that it is common among usability practitioners to ask participants in usability testing questions concerning redesign suggestions.

#### **How to review studies of users' design feedback?**

Through a wide range of UEMs that involve users' design feedback have been suggested, current knowledge on users' design feedback is fragmented; in part, because the literature on relevant UEMs often do not present detailed empirical data on the quality of users' design feedback (e.g. [2, 13, 31]).

We do not have a sufficient overview of the purposes for which users' design feedback is gathered. Furthermore, we do not know the degree to which users' design feedback serves its purpose as usability evaluation data. Does users' design feedback really complement other evaluation data sources, such as interaction data and usability experts' findings? To what degree can users' design feedback be seen as a credible source of usability evaluation findings; that is, what levels of validity and thoroughness can be expected? And to what degree does users' design feedback have an impact in the downstream development process?

To get an answer to these questions concerning users' design feedback, we needed to single out that part of the literature which presents empirical data this topic. We assumed that this literature typically would have the form of UEM assessments, where data on users' design feedback is compared to some external criterion to investigate its qualitative characteristics, validity and thoroughness, or downstream impact. UEM assessment as form of scientific enquiry has deep roots in the field of human-computer interaction (HCI); flourishing since the early nineties, typically pitting UEMs against each other to investigate their relative strengths and limitations (e.g. [32, 33]). Following Gray and Salzman's [34] criticism of early UEM assessments, studies have mainly targeted validity and thoroughness [35]. However, also aspects such as downstream utility [36, 37] and the qualitative characteristics of the output of different UEMs (e.g. [38, 39]) have been investigated in UEM assessments.

In our literature review, we have identified and analysed UEM assessments where the evaluation data included in the assessment at least in part are users' design feedback.

### Research question

Due to the exploratory character of the study, the following main research question was defined:

*Which are the potential benefits and limitations of users' design feedback in usability evaluations?*

The main research question was then broken down into four sub-questions, following from the questions raised in the section "[How to review studies of users' design feedback?](#)":

- RQ1: For which purposes are users' design feedback gathered in usability evaluation?
- RQ2: How do the qualitative characteristics of users' design feedback compare to that of other evaluation data (that is, interaction data and design feedback from usability experts)?
- RQ3: Which levels of validity and thoroughness are to be expected for users' design feedback?
- RQ4: Which levels of downstream impact are to be expected for users' design feedback?

## Methods

The literature review was set up following the guidelines of Kitchenham [40], with some adaptations to fit the nature of the problem area. In this "Methods" section we describe the search, selection, and analysis process.

### Search tool and search terms

Before conducting the review, we were aware of only a small number of studies concerning users' design feedback in usability evaluation; this in spite of our familiarity with the literature on UEMs. Hence, we decided to conduct the literature search through the Google Scholar search engine to allow for a broader scoping of publication channels than what is supported in other broad academic search engines such as Scopus or Web of Knowledge [41]. Google Scholar has been criticized for including a too broad range of content in its search results [42]. However, for the purpose of this review, where we aimed to conduct a broad search across multiple scientific communities, a Google Scholar search was judged to be an adequate approach.

To establish good search terms we went through a phase of trial and error. The key terms of the research question, *user* and "*design feedback*", were not useful even if combined with "*usability evaluation*"; the former due to its lack of discriminatory ability within the HCI literature, the latter because it is not an established term within the HCI field. Our solution to the challenge of establishing good search terms was to use the names of UEMs that involve users' design feedback. An initial list of relevant UEMs was established on the basis of our knowledge of the HCI field. Then, whenever we were made aware of other relevant UEMs throughout the review process, these were included as search terms along with the other UEMs. We also included the search term "*user reports*" (combined with "*usability evaluation*") as this term partly overlaps the term design feedback. The search was conducted in December 2012 and January 2013.

Table 1 lists the UEM names forming the basis of the search. For methods or approaches that are also used outside the field of HCI (cooperative evaluation, focus group, interview, contextual inquiry, the ADA approach, and online forums for evaluation) the UEM name was combined with the term *usability* or "*usability evaluation*".

To balance the aim for a broad search with the resources available, we set a cut-off at the 100 first hits for each search. For searches that returned fewer hits, we included all. The first 100 hits is, of course, an arbitrary cut-off and it is possible that more relevant papers had been found if this limit was extended. Hence, while the search indeed is broad it cannot claim complete coverage. We do not, however, see this as a problematic limitation. In practice, the cut-off was found to work satisfactorily as the last part of the included hits for a given search term combination typically returned little of interest for the purposes of the review. Increasing the number of included hits for each search combination would arguably have given diminishing returns.

### Selection and analysis

Each of the search result hits was examined according to publication channel and language. Only scientific journal and conference papers were included, as the quality of these is verified through peer review. Also, for practical reasons, only English language publications were included.

**Table 1 UEMs and associated search terms used in the literature review**

UEM group	UEM	Search terms	Hits (analysed and total)
Usability inspection methods	Pluralistic walkthrough (e.g. [2])	"Pluralistic walkthrough"	First 100 of 315
	Participatory heuristic evaluation (e.g. [13])	"Participatory heuristic evaluation"	First 100 of 110
	Structured expert evaluation method—SEEM (e.g. [24])	"Structured expert evaluation method"	All of 31
	Group-based expert walkthrough (e.g. [25])	"Group-based expert walkthrough"	All of 17
Usability testing methods	Cooperative evaluation (e.g. [14])	Usability AND "Cooperative evaluation"	First 100 of 343
	Cooperative usability testing—CUT (e.g. [26])	"Cooperative usability testing"	All of 39
	Developer–user contextual evaluation sessions—DUCE (e.g. [61])	"Developer–user contextual evaluation"/"designer user contextual enquiry"	All of 9
	Remote asynchronous usability testing (e.g. [44])	"Remote usability testing" AND asynchronous	First 100 of 112
Inquiry methods	Focus group (e.g. [4])	"Usability evaluation" AND "focus group"	First 100 of 1630
	Evaluation workshop [60] or Workshop test [28]	Usability AND "evaluation workshop" OR "workshop test"	First 100 of 485
	Interview [3]	"Usability evaluation" AND "Interview"	First 100 of 4890
	Contextual inquiry (e.g. [29])	"Usability evaluation" AND "Contextual inquiry" OR "Contextual enquiry"	First 100 of 733
	The ADA approach (e.g. [12])	Usability AND "ADA approach"	All of 45
	Online forums used for evaluation (e.g. [17])	Usability AND "Forum test" OR "forum evaluation"	All of 60
Other	Participatory evaluation through redesign and analysis—PETRA (e.g. [62])	"Participatory evaluation through redesign and analysis"	All of 35
	NA	"Usability evaluation" AND "user reports"	All of 94

All papers were scrutinized with regard to the following inclusion criterion: *Include papers with conclusions on the potential benefits and limitations of users' design feedback*. Papers excluded were typically conceptual papers presenting evaluation methods without presenting conclusions, studies on design feedback from participants (often students) that were not also within the target user group of the system, and studies that did not include qualitative design feedback but only quantitative data collection (e.g. set-response questionnaires). In total 41 papers were retained following this filtering. Included in this set were three papers co-authored by the author of this review [19, 25, 43].

The retained papers were then scrutinized according to possible overlapping studies and errors in classification. Nine papers were excluded as these presented the same data on users' design feedback as had already been presented in other of the identified papers,

but in less detail. One paper was excluded as it had been erroneously classified as a study of evaluation methods.

In the analysis process, all papers were coded on four aspects directly reflecting the research question: the purpose of the gathered users' design feedback (RQ1), the qualitative characteristics of the evaluation output (RQ2), assessments of validity and thoroughness (RQ3), and assessments of downstream impact (RQ4). Furthermore, all papers were coded according to UEM type, evaluation output types, comparison criterion (the criteria used, if any, to assess the design feedback), the involved users or participants, and research design.

## Results

The papers included for analysis concerned users' design feedback gathered through a wide range of methods from all the main UEM groups. The papers presented studies where users' design feedback was gathered through usability inspections, usability testing, and inquiry methods. Among the usability testing studies, users' design feedback was gathered both as extended debriefs and for users' self-reporting of problems or incidents. The inquiry methods were used both for stand-alone usability evaluations and as part of field tests (see Table 2). This width in studies should provide a good basis for making general claims on the benefits and limitations of users' design feedback.

Of the analysed studies, 19 provided detailed empirical data supporting their conclusions. The remaining studies presented the findings only summarily. The studies which

**Table 2 Overview of the UEM group and approach to users' design feedback in the retained papers**

UEM group	Users' design feedback	Number of papers	References
Usability inspection	Gathered by having users take the role of inspectors in usability inspections	4	Barcelos et al. [53], Edwards et al. [54], Følstad [25], Følstad et al. [43]
Usability testing (complemented with inquiry method)	Gathered during the test session or the debrief	8	Cunliffe et al. [49], Donker and Markopoulos [51], Følstad and Hornbæk [19], Henderson et al. [3], Marsh et al. [57] <sup>a</sup> , O'Donnel et al. [4], Wright and Monk [14], Yeo [59]
Usability testing (self-reports of problems or incidents)	Gathered in the form of users' self-reported problems	6	Andreasen et al. [27], Bruun and Stage [46], Bruun et al. [44], Hartson and Castillo [47], Marsh et al. [57] <sup>a</sup> , Petrie et al. [48]
Inquiry methods (as stand-alone evaluations)	Gathered through inquiry methods implemented as a self-sufficient UEMs	11	Åborg et al. [12], Choe et al. [30], Cowley and Radford-Davenport [20], Ebenezer [58], Kontio et al. [55], Obrist et al. [56], Hertzum [28] <sup>a</sup> , Lamanauskas et al. [63], Rector et al. [60], Sullivan [50], Sylaiou et al. [64]
Inquiry methods (part of field testing)	Gather by having the inquiry method implemented as part of a field test	4	Christensen and Frøkjær [45], Hertzum [28] <sup>a</sup> , Horsky et al. [52], Smilowitz et al. [17]

<sup>a</sup> The papers of Hertzum [28] and Marsh et al. [57] both concern UEMs belonging to more than one UEM group. These papers are therefore listed for more than one UEM group



provided detailed empirical data ranged from problem-counting head-to-head UEM comparisons, (e.g. [3, 17, 27, 44]) to in-depth reports on lessons learnt concerning a particular UEM (e.g. [30, 45]). All but two of the studies with detailed presentations of empirical data [20, 30] compared evaluation output from users' design feedback to output from interaction data and/or data from inspections with usability experts.

In the presented studies, users' design feedback was typically treated as a source to usability problems or incidents; this in spite that users' design feedback may serve as a gateway also to other types of evaluation output such as experiential issues, reflections on the system's context of use, and design suggestions. The findings from this review therefore mainly concern usability problems or incidents.

### **The purpose of gathering users' design feedback (RQ1)**

In the reviewed studies, different data collection methods for users' design feedback were often pitted against each other. For example, Bruun et al. [44] compared online report forms, online discussion forum, and diary as methods to gather users' self-reports of problems or incidents. Henderson et al. [3] compared interviews and questionnaires as means of gathering details on usability problems as part of usability testing debriefs. Cowley and Radford-Davenport [20] compared online discussion forum and focus groups for purposes of stand-alone usability evaluations.

These comparative studies surely provide relevant insight into the differences between specific data collection methods for users' design feedback. However, though comparative, most of these studies mainly addressed one specific purpose for gathering users' design feedback. Bruun et al. only considered users' design feedback in the context of users' self-reporting of problems in usability tests. Henderson et al. [3] only considered users' self-reporting during usability testing debriefs. Cowley and Radford-Davenport [20] only considered methods for users' design feedback as stand-alone evaluation methods. We therefore see it as beneficial to contrast the different purposes for gathering users' design feedback in the context of usability evaluations.

Four specific purposes for gathering users' design feedback were identified: (a) a budget approach to problem identification in usability testing, (b) to expand on interaction data from usability testing, (c) to identify problems in the users' everyday context, and (d) to benefit from users' knowledge or creativity.

### ***The budget approach***

In some of the studies, users' design feedback was used as a budget approach to reach findings that one could also have reached through classical usability testing. This is, in particular, seen in the five studies of usability testing with self-reports where the users' design feedback consisted mainly of reports of problems or incidents [27, 44, 46–48]. Here, the users were to run the usability test and report on the usability problems independently of the test administrator, potentially saving evaluation costs. For example, in their study of usability testing with disabled users, Petrie et al. [48] compared the self-reported usability problems from users that self-administer the usability test at home to those that participate in a similar usability test in the usability laboratory. Likewise, Andreasen et al. [27], Bruun et al. [44] compared different approaches to remote asynchronous usability testing. In these studies of self-reported usability problems, users'

design feedback hardly generated findings that complemented other data sources. Rather, the users' design feedback mainly generated a subset of the usability problems already identified through interaction data.

#### ***Expanding on interaction data***

Other reviewed studies concerned how users' design feedback may expand on usability test interaction data. This was seen in some of the studies where users' design feedback is gathered as part of the usability testing procedure or debrief session [4, 14, 19, 49, 59]. Here, users' design feedback generated additional findings rather than merely reproducing the findings of the usability test interaction data. For example, O'Donnel et al. [4] showed how the participants of a usability test converged on new suggestions for redesign in focus group sessions following the usability test. Similarly, Følstad and Hornbæk [19] found the participants of a cooperative usability test to identify other types of usability issues when walking through completed tasks of a usability test than the issues already evident through the interaction data. In both these studies, the debrief was set up so as to aid the memory of the users by the use of video recordings from the test session [4] or by walkthroughs of the test tasks [19]. Other studies were less successful in generating additional findings through such debrief sessions. For example, Henderson et al. [3] found that users during debrief interviews, though readily reporting problems, were prone to issues concerning recall, recognition, overload, and prominence. Likewise, Donker and Markopoulos [51], in their debrief interviews with children, found them susceptible of forgetfulness. Neither of these studies included specific memory aids during the debrief session.

#### ***Problem reports from the everyday context***

Users' design feedback may also serve to provide insight that is impractical to gather by other data sources. This is exemplified in the four studies concerning users' design feedback gathered through inquiry methods as part of field tests [17, 28, 45, 52]. Here, users reported on usability problems as they appear in everyday use of the interactive system, rather than usability problems encountered during the limited tasks of a usability test. As such, this form of users' design feedback provides insight into usability problems presumably holding high face validity, and that may be difficult to identify during usability testing. For example, Christensen and Frøkjær [45], gathered user reports on problems with a fleet management systems through an integrated reporting software. Likewise, Horsky et al. gathered user reports on problems with a medical application through emails from medical personnel. The user reports in these studies, hence, provided insight into problems as they appeared in the work-day of the fleet managers and medical personnel respectively.

#### ***Benefitting from users' knowledge and creativity***

Finally, in some of the studies, users' design feedback was gathered with the aim of benefiting from the particular knowledge or creativity of users. This is, in particular, seen in studies where users were involved as usability inspectors [25, 43, 53, 54] and in studies where inquiry methods were applied for stand-alone usability evaluations [20, 28, 30, 55, 56]. Also, some of the studies where users' design feedback was gathered through

extended debriefing sections had such a purpose [3, 4, 19, 57]. For example, in their studies of users as usability inspectors, Barcelos et al. [53], Edwards et al. [54], and Følstad [25] found the user inspectors to be particularly attentive to other aspects of the interactive systems than did the usability expert inspectors. Cowley and Radford-Davenport [20], as well as Ebenezer [58], in their studies of focus groups and discussion forums for usability evaluation, found participants to eagerly provide design suggestions, as did Sylaiou et al. [64] in their study of evaluations based on interviews and questionnaires with open-ended questions. Similarly, O'Donnell et al. [4] found users in focus groups arranged as follow-ups to classical usability testing sessions to identify and develop design suggestions; in particular in response to tasks that were perceived by the users as difficult.

#### **How do the qualitative characteristics of users' design feedback compare to that of other evaluation data? (RQ2)**

Given that users design feedback is gathered with the purpose of expanding on the interaction data from usability testing, or with the aim of benefitting from users knowledge and creativity, it is relevant to know whether users' design feedback actually generate findings that are different to what one could have reached through other data sources. Such knowledge may be found in the studies that addressed the qualitative characteristics of the usability issues identified on the basis on users' design feedback.

The qualitative characteristics of the identified usability issues were detailed in nine of the reviewed papers [17, 19, 20, 25, 28, 52–54, 59]. These studies indeed suggest that evaluations based on users' design feedback may generate output that is qualitatively different from that of evaluations based on other types of data. A striking finding across these papers is the degree to which users' design feedback may facilitate the identification of usability issues specific to the particular domain of the interactive system. In six of the papers addressing the qualitative characteristics of the evaluation output [19, 25, 28, 52–54], the findings based on users' design feedback concerned domain-specific issues not captured by the alternative UEMs. For example, in a heuristic evaluation of virtual world applications, studied by Barcelos et al. [53], online gamers that were representative of the typical users of the applications identified relatively more issues related to the concept of playability than did usability experts. Emergency response personnel and mobile salesforce representatives involved in cooperative usability testing, studied by Følstad and Hornbæk [19], identified more issues concerning needed functionality and organisational requirements when providing design feedback in the interpretation phases of the testing procedure than when providing interaction data in the interaction phases. The users of a public sector work support system, studied by Hertzum [28], identified more utility-problems when in a workshop test, where the users were free to provide design feedback, than they did in a classical usability test. Hertzum suggested that the rigidly set tasks, observational setup, and formal setting of the usability test made this evaluation “biased toward usability at the expense of utility”, whereas the workshop allowed more free exploration on the basis of the participants' work knowledge which was beneficial for the identification of utility problems and bugs.

In two of the studies, however, the UEMs involving users' design feedback were not reported to generate more domain-specific issues than did the other UEMs [17, 59].

These two studies differed from the others on one important point: the evaluated systems were general purpose work support systems (one spreadsheet system and one system for electronic Post-It notes), not systems for specialized work support. A key motivation for gathering users' design feedback is that users possess knowledge not held by other parties of the development process. Consequently, as the contexts of use for these two systems most likely were well known to the involved development teams, the value of tapping into user's domain knowledge may have been lower than for the evaluations of more specialized work support systems.

The studies concerning the qualitative characteristics of users' design feedback also suggested the importance of not relying solely on such feedback. In all the seven studies, findings from UEMs based on users' design feedback were compared with findings from UEMs based on other data sources (interaction data or usability experts' findings). In all of these, the other data sources generated usability issues that were not identified from the users' design feedback. For example, the usability experts in usability inspections studied by Barcelos et al. [53] and Følstad [25] identified a number of usability issues not identified by the users; issues that also had different qualitative characteristics. In the study by Barcelos et al. [53], the usability expert inspectors identified more issues pertaining to system configuration than did the user inspectors. In the study by Følstad [25], the usability expert inspectors identified more domain-independent issues. Hence, depending only on users' design feedback would have limited the findings with respect to issues related to what Barcelos et al. [53] referred to as "the classical usability concept" (p. 303).

These findings are in line with our assumption that users' design feedback may complement other types of evaluation data by supporting qualitatively different evaluation output, but not replace other evaluation data. Users' design feedback may constitute an important addition to other evaluation data sources, by supporting the identification of domain specific usability issues and, also, user-based suggestions for redesign.

### **Which levels of validity and thoroughness are to be expected for users' design feedback?**

#### **(RQ3)**

To rely on users' design feedback as data in usability evaluations, we need to trust the data. To be used for any evaluation purpose, the findings based on users' design feedback need to hold adequate levels of *validity*; that is, the usability problems identified during the evaluation should reflect problems that the user can be expected to encounter when using the interactive system outside the evaluation context. Furthermore, if users' design feedback is to be used as the only data in usability evaluations, it is necessary to know the levels of thoroughness that can be expected; that is, the degree to which the evaluation serves to identify all relevant usability problems that the user can be expected to encounter.

Following Hartson et al. [35], validity and thoroughness scores can be calculated on the basis of (a) the set of usability problems predicted with a particular UEM and (b) the set of *real* usability problems, that is, usability problems actually encountered by users outside the evaluation context. The challenge of such calculations, however, is that we need to establish a reasonably complete set of real usability problems. This challenge has typically been resolved by using the findings from classical usability testing as an

approximation to such a set [65], though this approach introduces the risk of erroneously classifying usability problems as false alarms [6].

A substantial proportion of the reviewed papers present general views on the validity of the users' design feedback. However, only five of the papers included in the review provide sufficient detail to calculate validity scores. This, provided that we assume that classical laboratory testing can serve as an approximation to the complete set of real usability problems. In three of these [44, 46, 47], the users' design feedback was gathered as self-reports during remote usability testing, in one [3] users' design feedback was gathered during usability testing debrief, and in one [43] users' design feedback was gathered through usability inspection. The validity scores ranged between 60% [43] and 89% [47], meaning that in all of the studies 60% or more of the usability problems or incidents predicted by the users were also confirmed by classical usability testing.

The reported validity values for users' design feedback were arguably acceptable. For comparison, in newer empirical studies of heuristic evaluation with usability experts the validity of the evaluation output has typically been found to be well below 50% (e.g. [6, 7]). Furthermore, following from the challenge of establishing a complete set of real usability problems, it may be assumed that several of the usability problems not identified in classical usability testing may nevertheless represent real usability problems [43, 47].

Thoroughness concerns the proportion of predicted real problems relative to the full set of real problems [35]. Some of the above studies also provided empirical data that can be used to assess the thoroughness of users' design feedback. In the Hartson and Castillo [47] study, 68% of the critical incidents observed during video analysis were also self-reported by the users. The similar proportion for the study by Henderson et al. [3] on problem identification from interviews was 53%. For the study on users as usability inspectors by Følstad et al. [43] the median thoroughness score for individual inspectors was 25%; however, for inspectors in nominal groups of seven thoroughness scores were raised to 70%. Larger numbers of evaluators or users is beneficial to thoroughness [35]. This is, in particular, seen in the study of Bruun et al. [44] where 43 users self-reporting usability problems in remote usability evaluations were able to identify 78% of the problems identified in classical usability testing. For comparison, in newer empirical studies of heuristic evaluation with usability experts thoroughness is typically well above 50% (e.g. [6, 7]).

The empirical data on thoroughness seem to support the conclusion that users typically underreport problems in their design feedback, though the extent of such underreporting varies widely between evaluations. In particular, involving larger numbers of users may mitigate this deficit in users' design feedback as an evaluation data source.

#### **Which levels of downstream impact are to be expected for users' design feedback? (RQ4)**

Seven of the papers presented conclusions concerning the impact of users' design feedback on the subsequent design process; that is, whether the issues identified during evaluations lead to change in later versions of the system. Rector et al. [60], Obrist et al. [56], and Wright and Monk [14] concluded that the direct access to users' reports served to strengthen the understanding in the design team of the users' needs. The remaining four studies concerning downstream impact, provided more detailed evidence on this.

In a study by Hertzum [28], the impact ratio for a workshop test was found to be more than 70%, which was similar to that of a preceding usability test in the same development process. Hertzum argued that a key factor determining the impact of an evaluation is its location in time: evaluations early in the development process are argued to have more impact than late evaluations. Følstad and Hornbæk [19], in their study of cooperative usability testing, found the usability issues identified on the basis of users' design feedback during interpretation phases to have equal impact to those identified on the basis of interaction data. Følstad [25] in his study of users and usability experts as inspectors for applications for three specialized domains, found usability issues identified users on average to have higher impact than those of usability experts. Horsky et al. [52] studied usability evaluations of a medical work support system by way of users' design feedback through email and free-text questionnaires during field trial, and compared the findings from these methods to findings from classical usability testing and inspections conducted by usability experts. Here, 64% of the subsequent changes to the system were motivated from issues reported in users' self-reports by email. E-mail reports were also the most prominent source of users' design feedback; 85 of a total of 155 user comments were gathered through such reports. Horsky et al. suggested the problem types identified from the e-mail reports to be an important reason for the high impact of the findings from this method.

## **Discussion and conclusion**

### **The benefits and limitations of users' design feedback**

The literature review has provided an overview concerning the potential benefits and limitations of users' design feedback. We found that users' design feedback can be gathered for four purposes. When users' design feedback is gathered to expand on interaction data from usability testing, as in usability testing debriefs (e.g. [4]), or benefitting from the users' knowledge or creativity, as in usability inspections with user inspectors (e.g. [53]), it is critical that the evaluation output include findings that complement what could be achieved through other evaluation data sources; if not, the rationale for gathering users' design feedback in such studies is severely weakened. When users' design feedback is gathered as a budget approach to classical usability testing, as in asynchronous remote usability testing (e.g. [44]), or a way to identify problems in the users' everyday context, as in inquiry methods as part of field tests (e.g. [45]), it is critical that the evaluation output holds adequate validity and thoroughness.

The studies included in the review indicate that users' design feedback may indeed complement other types of evaluation data. This is seen in the different qualitative characteristics for findings made on the basis of users' design feedback compared to those made from other evaluation data types. This finding is important, as it may motivate usability professionals to make better use of UEMs particularly designed to gather of users' design feedback to complement other evaluation data. Such UEMs may include the pluralistic walkthrough, where users participate as inspectors in groups with usability experts and development team representatives, and the cooperative usability testing, where users' design feedback is gathered through dedicated interpretation phases added to the classical usability testing procedure. Using UEMs that support users' design feedback seems to be particularly important when evaluating systems for specialized

domains, such as that of medical personnel or public sector employees. Possibly, the added value of users' design feedback as a complementary data source may be reduced in evaluations of interactive systems for the general public; here, the users' design feedback may not add much to what is already identified through interaction data or usability experts' findings.

Furthermore, the reviewed studies indicated that users' can self-report incidents or problems validly. For usability testing with self-reporting of problems, validity values for self-reports were consistently 60% or above; most identified incidents or problems made during self-report were also observed during interaction. In the studies providing validity findings, the objects of evaluation were general purpose work support systems or general public websites, potentially explaining why the users did not make findings more complementary to that of the classical usability test.

Users were, however, found to be less able with regard to thoroughness. In the reviewed studies, thoroughness scores varied from 25 to 78%. A relatively larger number of users' seems to be required to reach adequate thoroughness through users' design feedback than through interaction data. Evaluation depending solely on users' design feedback may need to increase the number of users relative to what would be done e.g. for classical usability testing.

Finally, issues identified from users' design feedback may have substantial impact in the subsequent development process. The relative impact of users' design feedback compared to that of other data sources may of course differ between studies and development process, e.g. due to contextual variation. Nevertheless, the reviewed studies indicate users' design feedback to be at least as impactful as evaluation output from other data sources. This finding is highly relevant for usability professionals, whom typically aim to get the highest possible impact on development. One reason why findings from users' design feedback were found to have relatively high levels of impact may be that such findings, as opposed to, for example, the findings of usability experts in usability inspections, allow the development team to access the scarce resource of users' domain knowledge. Hence, the persuasive character of users' design feedback may be understood as a consequence of it being qualitatively distinct from evaluation output from other data sources, rather than merely being a consequence of this feedback coming straight from the users.

#### **Implications for usability evaluation practice**

The findings from the review may be used to advice usability evaluation practice. In the following, we summarize what we find to be the most important take-away for practitioners:

- Users' design feedback may be particularly beneficial when conducting evaluation of interactive systems for specialized contexts of use. Here, users' design feedback may generate findings that complement those based on other types of evaluation data. However, for this benefit to be realized, the users' design feedback should be gathered with a clear purpose of benefitting from the knowledge and creativity of users.
- When users' design feedback is gathered through extended debriefs, users are prone to forgetting encountered issues or incidents. Consider supporting the users recall

by the use of, for example, video recordings from system interaction or by walking through the task.

- Users' design feedback may support problem identification, in evaluations where the purpose is a budget approach to usability testing or problem reporting from the field. However, due to challenges in thoroughness, it may be necessary to scale up such evaluations to involve more users than would be needed e.g. for classical usability testing.
- Evaluation output based on users' design feedback seems to be impactful in the downstream development process. Hence, gathering users' design feedback may be an effective way to boost the impact of usability evaluation.

### **Limitations and future work**

Being a literature review, this study is limited by the research papers available. Though evaluation findings from interaction data and inspections with usability experts have been thoroughly studied in the research literature, the literature on users' design feedback is limited. Furthermore, as users' design feedback is not used as a term in the current literature, the identification of relevant studies was challenging to the point that we cannot be certain that not some relevant study has passed unnoticed.

Nonetheless, the identified papers, though concerning a wide variety of UEMs, were found to provide reasonably consistent findings. Furthermore, the findings suggest that users' design feedback is a promising area for further research on usability evaluation.

The review also serves to highlight possible future research directions, to optimize UEMs for users' design feedback and to further investigate which types of development processes that in particular benefit from users' design feedback. In particular, the following topics may be highly relevant for future work:

- More systematic studies of the qualitative characteristics of UEM output in general, and users' design feedback in particular. In the review, a number of studies addressing various qualitative characteristics were identified. However, to optimize UEMs for users' design feedback it may be beneficial to study the qualitative characteristics of evaluation output according to more comprehensive frameworks where feedback is characterized e.g. in terms of being general or domain-specific as well as being problem oriented, providing suggestions, or concerning the broader context of use.
- Investigating users' design feedback across types of application areas. The review findings suggest that the usefulness of users' design feedback in part may be decided by application area. In particular, application domains characterized by high levels of specialization may benefit more from evaluations including users' design feedback, as the knowledge represented by the users are not as easily available through other means as for more general domains. Future research is needed for more in-depth study of this implication of the findings.
- Systematic studies of users' design feedback across the development process. It is likely, as seen from the review, that the usefulness of users' design feedback may be dependent on which stage of the development process in which the evaluation is conducted. Furthermore, different stages of the development process may require different UEMs for gathering users' design feedback. In the review, we identified four



typical motivations for gathering users' design feedback. These may serve as a starting point for further studies of users' design feedback across the development process.

While the review provides an overview of our current and fragmented knowledge of users' design feedback, important areas of research still remain. We conclude that users' design feedback is a worthy topic of future UEM research, and hope that this review can serve as a starting point for this endeavour.

#### Acknowledgements

The presented work was supported the Research Council of Norway Grant Numbers 176828 and 203432. Thanks to Professor Kasper Hornbæk for providing helpful and constructive input on the manuscript and for supervising the Ph.D. work on which it is based.

#### Competing interests

The author declares no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 2 July 2016 Accepted: 18 May 2017

Published online: 03 July 2017

#### References

- Rubin J, Chisnell D (2008) Handbook of usability testing: how to plan, design, and conduct effective tests, 2nd edn. Wiley, Indianapolis
- Bias RG (1994) The pluralistic usability walkthrough: coordinated empathies. In: Nielsen J, Mack RL (eds) Usability inspection methods. Wiley, New York, pp 63–76
- Henderson R, Podd J, Smith MC, Varela-Alvarez H (1995) An examination of four user-based software evaluation methods. *Interact Comput* 7(4):412–432
- O'Donnel PJ, Scobie G, Baxter I (1991) The use of focus groups as an evaluation technique in HCI. In: Diaper D, Hammond H (eds) People and computers VI, proceedings of HCI 1991. Cambridge University Press, Cambridge, pp 212–224
- Lewis JR (2006) Sample sizes for usability tests: mostly math, not magic. *Interactions* 13(6):29–33
- Chattrichart J, Brodie J (2004) Applying user testing data to UEM performance metrics. In: Dykstra-Erickson E, Tscheligi M (eds) CHI'04 extended abstracts on human factors in computing systems. ACM, New York, pp 1119–1122
- Hvannberg ET, Law EL-C, Lárusdóttir MK (2007) Heuristic evaluation: comparing ways of finding and reporting usability problems. *Interact Comput* 19(2):225–240
- Nielsen J (2001) First rule of usability? don't listen to users. Jakob Nielsen's Alertbox: August 5, 2001. <http://www.nngroup.com/articles/first-rule-of-usability-dont-listen-to-users/>
- Whitefield A, Wilson F, Dowell J (1991) A framework for human factors evaluation. *Behav Inf Technol* 10(1):65–79
- Gould JD, Lewis C (1985) Designing for usability: key principles and what designers think. *Commun ACM* 28(3):300–311
- Wilson GM, Sasse MA (2000) Do users always know what's good for them? Utilising physiological responses to assess media quality. People and computers XIV—usability or else!. Springer, London, pp 327–339.
- Åborg C, Sandblad B, Gulliksen J, Lif M (2003) Integrating work environment considerations into usability evaluation methods—the ADA approach. *Interact Comput* 15(3):453–471
- Muller MJ, Matheson L, Page C, Gallup R (1998) Methods & tools: participatory heuristic evaluation. *Interactions* 5(5):13–18
- Wright PC, Monk AF (1991) A cost-effective evaluation method for use by designers. *Int J Man Mach Stud* 35(6):891–912
- Dumas JS, Redish JC (1999) A practical guide to usability testing. Intellect Books, Exeter
- Følstad A, Law E, Hornbæk K (2012) Analysis in practical usability evaluation: a survey study. In: Chi E, Höök K (eds) Proceedings of the SIGCHI conference on human factors in computing systems, CHI '12. ACM, New York, pp 2127–2136
- Smilowitz ED, Darnell MJ, Benson AE (1994) Are we overlooking some usability testing methods? A comparison of lab, beta, and forum tests. *Behav Inf Technol* 13(1–2):183–190
- Vermeeren AP, Law ELC, Roto V, Obrist M, Hoonhout J, Väänänen-Vainio-Mattila K (2010) User experience evaluation methods: current state and development needs. In: Proceedings of the 6th Nordic conference on human-computer interaction: extending boundaries, ACM, New York, p 521–530

19. Følstad A, Hornbæk K (2010) Work-domain knowledge in usability evaluation: experiences with cooperative usability testing. *J Syst Softw* 83(11):2019–2030
20. Cowley JA, Radford-Davenport J (2011) Qualitative data differences between a focus group and online forum hosting a usability design review: a case study. *Proceedings of the human factors and ergonomics society annual meeting* 55(1): 1356–1360
21. Jacobsen NE (1999) Usability evaluation methods: the reliability and usage of cognitive walkthrough and usability test. (Doctoral thesis. University of Copenhagen, Denmark)
22. Cockton G, Lavery D, Woolrych A (2008) Inspection-based evaluations. In: Sears A, Jacko J (eds) *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, 2nd edn. Lawrence Erlbaum Associates, New York, pp 1171–1190
23. Mack RL, Nielsen J (1994) Executive summary. In: Nielsen J, Mack RL (eds) *Usability inspection methods*. Wiley, New York, pp 1–23
24. Baauw E, Bekker MM, Barendregt W (2005) A structured expert evaluation method for the evaluation of children's computer games. In: Costabile MF, Paternò F (Eds.) *Proceedings of human-computer interaction—INTERACT 2005*, lecture notes in computer science 3585, Springer, Berlin, p 457–469
25. Følstad A (2007) Work-domain experts as evaluators: usability inspection of domain-specific work support systems. *Int J Human Comp Interact* 22(3):217–245
26. Frøkjær E, Hornbæk K (2005) Cooperative usability testing: complementing usability tests with user-supported interpretation sessions. In: van der Veer G, Gale C (eds) *CHI'05 extended abstracts on human factors in computing systems*. ACM Press, New York, pp 1383–1386
27. Andreasen MS, Nielsen HV, Schrøder SO, Stage J (2007) What happened to remote usability testing? An empirical study of three methods. In: Rosson MB, Gilmore D (Eds.) *CHI'97: Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, New York, p 1405–1414
28. Hertzum M (1999) User testing in industry: a case study of laboratory, workshop, and field tests. In: Kobsa A, Stephanidis C (Eds.) *Proceedings of the 5th ERCIM Workshop on User Interfaces for All*, Dagstuhl, Germany, November 28–December 1, 1999. [http://www.interaction-design.org/references/conferences/proceedings\\_of\\_the\\_5th\\_ercim\\_workshop\\_on\\_user\\_interfaces\\_for\\_all.html](http://www.interaction-design.org/references/conferences/proceedings_of_the_5th_ercim_workshop_on_user_interfaces_for_all.html)
29. Rosenbaum S, Kantner L (2007) Field usability testing: method, not compromise. *Proceedings of the IEEE international professional communication conference, IPCC 2007*. doi: 10.1109/IPCC.2007.4464060
30. Choe P, Kim C, Lehto MR, Lehto X, Allebach J (2006) Evaluating and improving a self-help technical support web site: use of focus group interviews. *Int J Human Comput Interact* 21(3):333–354
31. Greenbaum J, Kyng M (eds) (1991) *Design at work*. Lawrence Erlbaum Associates, Hillsdale
32. Desurvire HW, Kondziela JM, Atwood ME (1992) What is gained and lost when using evaluation methods other than empirical testing. In: Monk A, Diaper D, Harrison MD (eds) *People and computers VII: proceedings of HCI 92*. Cambridge University Press, Cambridge, pp 89–102
33. Karat CM, Campbell R, Fiegel T (1992) Comparison of empirical testing and walkthrough methods in user interface evaluation. In: Bauersfeld P, Bennett J, Lynch G (Eds.) *CHI'92: Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, New York, p 397–404
34. Gray WD, Salzman MC (1998) Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human Comput Interact* 13(3):203–261
35. Hartson HR, Andre TS, Williges RC (2003) Criteria for evaluating usability evaluation methods. *Int J Human Comput Interact* 15(1):145–181
36. Law EL-C (2006) Evaluating the downstream utility of user tests and examining the developer effect: a case study. *Int J Human Comput Interact* 21(2):147–172
37. Uldall-Espersen T, Frøkjær E, Hornbæk K (2008) Tracing impact in a usability improvement process. *Interact Comput* 20(1):48–63
38. Frøkjær E, Hornbæk K (2008) Metaphors of human thinking for usability inspection and design. *ACM Trans Comput Human Interact (TOCHI)* 14(4):20:1–20:33
39. Fu L, Salvendy G, Turley L (2002) Effectiveness of user testing and heuristic evaluation as a function of performance classification. *Behav Inf Technol* 21(2):137–143
40. Kitchenham B (2004) *Procedures for performing systematic reviews* (Technical Report TR/SE-0401). Keele, UK: Keele University. <http://www.scm.keele.ac.uk/ease/sreview.doc>
41. Harzing AW (2013) A preliminary test of Google Scholar as a source for citation data: a longitudinal study of Nobel prize winners. *Scientometrics* 94(3):1057–1075
42. Meho LI, Yang K (2007) Impact of data sources on citation counts and rankings of LIS faculty: web of Science versus Scopus and Google Scholar. *J Am Soc Inform Sci Technol* 58(13):2105–2125
43. Følstad A, Anda BC, Sjøberg DIK (2010) The usability inspection performance of work-domain experts: an empirical study. *Interact Comput* 22:75–87
44. Bruun A, Gull P, Hofmeister L, Stage J (2009) Let your users do the testing: a comparison of three remote asynchronous usability testing methods. In: Hickley K, Morris MR, Hudson S, Greenberg S (Eds.) *CHI'09: Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, New York, p 1619–1628
45. Christensen L, Frøkjær E (2010) Distributed usability evaluation: enabling large-scale usability evaluation with user-controlled instrumentation. In: Blandford A, Gulliksen J (Eds.) *NordiCHI'10: Proceedings of the 6th Nordic conference on human-computer interaction: extending boundaries*, ACM, New York, p 118–127
46. Bruun A, Stage J (2012) The effect of task assignments and instruction types on remote asynchronous usability testing. In: Chi EH, Höök K (Eds.) *CHI'12: Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, New York, p 2117–2126
47. Hartson H R, Castillo JC (1998) Remote evaluation for post-deployment usability improvement. In: Catarci T, Costabile MF, Santucci G, Tarafino L, Levialdi S (Eds.) *AVI'98: Proceedings of the working conference on advanced visual interfaces*, ACM Press, New York, p 22–29

48. Petrie H, Hamilton F, King N, Pavan P (2006) Remote usability evaluations with disabled people. In: Grinter R, Rodden T, Aoki P, Cutrell E, Jeffries R, Olson G (Eds.) CHI'06: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, p 1133–1141
49. Cunliffe D, Kritou E, Tudhope D (2001) Usability evaluation for museum web sites. *Mus Manag Curatorship* 19(3):229–252
50. Sullivan P (1991) Multiple methods and the usability of interface prototypes: the complementarity of laboratory observation and focus groups. In: Proceedings of the International Conference on Systems Documentation—SIGDOC'91, ACM, New York, p 106–112
51. Donker A, Markopoulos P (2002) A comparison of think-aloud, questionnaires and interviews for testing usability with children. In: Faulkner X, Finlay J, Détienne F (eds) *People and computers XVI—memorable yet invisible*, proceedings of HCI 2002. Springer, London, pp 305–316
52. Horsky J, McColgan K, Pang JE, Melnikas AJ, Linder JA, Schnipper JL, Middleton B (2010) Complementary methods of system usability evaluation: surveys and observations during software design and development cycles. *J Biomed Inform* 43(5):782–790
53. Barcelos TS, Muñoz R, Chalegre V (2012) Gamers as usability evaluators: A study in the domain of virtual worlds. In: Anacleto JC, de Almeida Nedis VP (Eds.) IHC'12: Proceedings of the 11th Brazilian symposium on human factors in computing systems, Brazilian Computer Society, Porto Alegre, p 301–304
54. Edwards PJ, Moloney KP, Jacko JA, Sainfort F (2008) Evaluating usability of a commercial electronic health record: a case study. *Int J Hum Comput Stud* 66:718–728
55. Kontio J, Lehtola L, Bragge J (2004) Using the focus group method in software engineering: obtaining practitioner and user experiences. In: Proceedings of the International Symposium on Empirical Software Engineering – ISESE, IEEE, Washington, p 271–280
56. Obrist M, Moser C, Alliez D, Tscheligi M (2011) In-situ evaluation of users' first impressions on a unified electronic program guide concept. *Entertain Comput* 2:191–202
57. Marsh SL, Dykes J, Attilakou F (2006) Evaluating a geovisualization prototype with two approaches: remoteinstructional vs. face-to-face exploratory. In: Proceedings of information visualization 2006, IEEE, Washington, p 310–315
58. Ebenezer C (2003) Usability evaluation of an NHS library website. *Health Inf Libr J* 20(3):134–142
59. Yeo A (2001) Global-software development lifecycle: an exploratory study. In: Jacko J, Sears A (Eds.) CHI'01: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, p 104–111
60. Rector AL, Horan B, Fitter M, Kay S, Newton PD, Nowlan WA, Robinson D, Wilson A (1992) User centered development of a general practice medical workstation: The PEN&PAD experience. In: Bauersfeld P, Bennett J, Lurch G (Eds.) CHI'92: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, p 447–453
61. Smith A, Dunckley L (2002) Prototype evaluation and redesign: structuring the design space through contextual techniques. *Interact Comput* 14(6):821–843
62. Ross S, Ramage M, Ramage Y (1995) PETRA: participatory evaluation through redesign and analysis. *Interact Comput* 7(4):335–360
63. Lamanauskas L, Pribeanu C, Vilkonis R, Balog A, Iordache DD, Klanguauskas A (2007) Evaluating the educational value and usability of an augmented reality platform for school environments: some preliminary results. In: Proceedings of the 4th WSEAS/IASME international conference on engineering education p 86–91
64. Sylaiou S, Economou M, Karoulis A, White M (2008) The evaluation of ARCO: a lesson in curatorial competence and intuition with new technology. *ACM Comput Entertain* 6(20):23
65. Hornbæk K (2010) Dogmas in the assessment of usability evaluation methods. *Behav Inf Technol* 29(1):97–111

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---