

RESEARCH

Open Access

# Social design feedback: evaluations with users in online ad-hoc groups

Asbjørn Følstad<sup>1\*</sup>, Kasper Hornbæk<sup>2</sup> and Pål Ulleberg<sup>3</sup>

\* Correspondence: asf@sintef.no  
<sup>1</sup>SINTEF, Pb124 Blindern, 0314 Oslo, Norway  
Full list of author information is available at the end of the article

## Abstract

Social design feedback is a novel approach to usability evaluation where user participants are asked to comment on designs asynchronously in online ad-hoc groups. Two key features of this approach are that (1) it supports interaction between user participants and development team representatives and (2) user participants can see and respond to other participants' comments. Two design cases, involving 250 user participants, were studied to explore the output of social design feedback and investigate the effect of the two key features of this approach. Of all the design feedback, 17% was rated highly useful, and 21% contained change suggestions. The presence of an active moderator, representing the development team and interacting with the user participants, increased the usefulness of the design feedback. The opportunity to see and respond to others' design feedback had a minor effect on the kind of design feedback provided, but no effect on usefulness. Based on the findings, we offer advice on how to implement social design feedback and suggest future research.

**Keywords:** Human-computer interaction; Social design feedback; Usability evaluation; Online user involvement

## Introduction

Involving users in usability evaluation is valuable when designing information and communication technology (ICT). Traditionally, evaluation with users has been conducted face-to-face with methods such as usability testing, participatory evaluation and post-experience interviews [1]. Increasingly, however, online evaluation methods are more used.

Online evaluation methods include (a) methods that require user participants and development team representatives to be present synchronously and (b) methods that require user participants to contribute asynchronously. Synchronous evaluation methods comprise, among others, moderated remote usability testing [2] and online focus groups [3]. Asynchronous evaluation methods include online questionnaires [4], unmoderated remote usability testing [5], and involvement of online user communities in design and development [6].

The established asynchronous online evaluation methods have inherent benefits and limitations. Questionnaires and unmoderated remote usability testing allow for relatively easy access to participants as participation does not require long-term engagement. However, these methods restrict interaction between the development team and the user participants, barring the development team from asking participants to clarify or elaborate

their contributions. The involvement of online user communities enables interaction between the development team and user participants, for example in user forums or as part of beta testing [7]. As the establishment of active online user communities requires dedicated and long-term efforts, these typically are not available to a development project.

The popular uptake of social internet solutions, often referred to as social media [8], supports new opportunities for gathering design feedback online. In particular, designs at any level of maturity may be presented online for groups of colleagues, peers, clients or users to contribute design feedback, by way of comments and discussion threads. Available services for such design feedback include Notable [9] and ConceptShare [10].

We term this emerging approach *social design* feedback, as it exploits social internet solutions. By this term we mean asynchronous feedback in ad-hoc online groups. Compared to other methods for asynchronous evaluation with users, social design feedback has two key features: (1) There may be interaction between user participants and development team representatives and (2) the user participants can see and respond to other participants' contributions. No pre-existing online user community is required, since participants for social design feedback can be recruited in the same manner as for online surveys or unmoderated usability testing and the feedback is gathered in ad-hoc groups.

Our knowledge of social design feedback as an approach to usability evaluation is limited. Little is known concerning what level of usefulness to expect from the output of social design feedback, and the qualitative characteristics of this output is not sufficiently explored. Furthermore, we have no knowledge on the effect of the key characteristics of social design feedback. This lack in knowledge is critical if the Human-Computer Interaction (HCI) community is to judge the relevance of this approach and develop it further as an HCI evaluation method.

We present a study on social design feedback that investigates the output of social design feedback and the effect of the two key features of this method on output quality.

The contribution of the study is to increase our understanding of social design feedback as an approach to usability evaluation. The study contributes an exploration of the usefulness of the output from social design feedback in terms of its relevance and ability to inspire subsequent design work, as well as the qualitative characteristics of this output in terms of users' concerns and change suggestions. Furthermore, the effects of the key features of this novel approach to usability evaluation are examined. In particular, we examine how the interaction between study participants and an active moderator, as well as the participants' direct view of each other's contributions, affect the usefulness and qualitative characteristics of the evaluation output. The study is of particular interest for HCI researchers and practitioners concerned with new ways of conducting usability evaluations.

## **Background**

### **Design feedback in evaluations with users**

Involving users in evaluations may generate two types of output: *Interaction data* and *design feedback*. Interaction data are data from the users' interaction with a 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 [11], considerations on the system's suitability for its context of use [12], predictions of usability issues [13], and suggestions on design improvements [14]. Design feedback may address any aspect of a design, such as visual layout, interaction design, content categories, and technical or performance issues [13].

Methods that generate interaction data are well exemplified by usability testing, the most commonly used method for involving users in evaluation when designing ICT [15,16]. It should, however, be noted that usability testing also may involve design feedback, in particular through the application of post-test questionnaires or interviews [17].

Evaluation methods that may be used to gather design feedback include enquiry methods (for example workshop evaluations [18], focus groups [19], and questionnaire methods [20]), but also usability testing methods that allow users to contribute their reflections on the evaluated designs (for example cooperative usability testing [21]) and inspection methods supporting users as inspectors (for example pluralistic walkthrough [22] and group-based expert walkthrough [23]). It has been suggested that evaluation methods used to gather design feedback should provide substantial guidance and support so as to enable users inexperienced with usability evaluation to generate useful feedback [23].

In current usability evaluation practice, users are often asked for design feedback. In a recent survey, involving 112 participants reporting on their latest usability test [24], 80% reported having asked test participants for their opinion on usability problems (64%), redesign suggestions (48%), and/or other issues (28%).

Social design feedback, which is our approach to gathering design feedback, has only sporadically been studied in the field of HCI. Hagen and Robertson [25] discussed user participation through social technologies. Apart from this, we are aware of three empirical studies of evaluation methods resembling social design feedback, all concerning the use of online forums for usability evaluation.

Smilowitz, Darnell and Benson [26] and Bruun, Gull, Hofmeister, and Stage [27] compared online forums for usability evaluation to usability testing and individual self-report of usability problems. Self-reports were gathered as part of beta testing [26], remote asynchronous usability testing [27], and diary reports [27]. In both studies, the online forum approach was found to identify fewer usability problems than did the usability test. However, none of the studies were set up to exploit the opportunity for increased numbers of user participants in online forum evaluations, as the number of participants in the forum conditions were the same as in the usability test conditions.

Cowley and Radford-Davenport [14] compared evaluations in online forums to evaluations in focus groups with respect to design suggestions and participant conversations. They found that the online forum evaluations generated more design suggestions, but that the focus group evaluations to a greater degree induced conversations that could lead to unexpected findings.

None of the three studies considered design feedback as encompassing both usability problems and change suggestions. Smilowitz et al. and Bruun et al. studied evaluation output in terms of usability problems only, whereas Cowley and Radford-Davenport studied the output in terms of design suggestions. Only one of the studies [14] concerned the usefulness of the users' design feedback, by including an analysis of the feasibility of the resulting design suggestions. None of the studies included more than one case,

something that may represent a threat to the external validity [28] of the studies' conclusions.

The three studies provided only limited insight in the effect of the two key features of social design feedback. The effect of interaction between user participants and development team representatives was not investigated in any of the studies. The effect of allowing user participants to access each other's comments could have been investigated in two of the studies [26,27] as these compared methods for individual self-reporting of problems to forum methods. However, as the compared methods differed on several features, the findings from these comparisons cannot be directly attributed to whether or not user participants were allowed to access each other's comments.

### **Current solutions for social design feedback**

Emerging solutions for social design feedback follow one of two approaches. One approach is to enable feedback on a visual presentation of a user interface by adding comments as annotations in a separate visual layer, as for example Notable [9], Notebox [29], Cage [30], and ConceptShare [10]. All contributed annotations are visualised on-screen. This allows all participants to see the feedback that has already been given, but also serves to limit the number of contributions that may be handled due to the on-screen clutter resulting from large numbers of annotations. Some solutions, including ConceptShare, Cage and Notable, support discussion threads associated with the annotations.

Another approach is to enable feedback as comments in an adjacent discussion thread without on-screen annotations, such as the Open Web Lab, OWELA [31] and the RECORD online Living Lab [32]. All contributions are available to all participants for reading and commenting. This approach allows larger amounts of feedback as the discussion thread avoids the problem of on-screen clutter that may result from large numbers of comments added as annotations directly to the visual presentation. However, the lack of on-screen annotations may make it more difficult for participants to get an overview of all comments addressing a given element in the visual presentation.

### **Assessing the output of social design feedback**

Within the field of HCI, usability evaluation output is typically assessed on *thoroughness* and *validity* [33]; that is, the proportion of real usability problems that are identified in the evaluation and the proportion of problem predictions that actually correspond to real usability problems (as opposed to false positives). This approach to assessment, however, is hardly viable when assessing output from social design feedback; partly because such output may include more than just usability problem predictions, in particular positive feedback and change suggestions, and partly because the thoroughness and validity assessments require a *comparison criterion* [33], a comprehensive set of real usability problems against which the evaluation output can be assessed. Such comparison criteria are typically established through usability testing, something that is not possible for ideas and early concepts which are important objects of evaluation in social design feedback.

Another approach to the assessment usability evaluation output is to assess its impact on the subsequent design process [12,13,34]. This approach, however, requires access to the development team at a later stage in the design process. Furthermore, this

approach may be vulnerable to spurious effects caused by conditions in the design process not related to the evaluation, for example, management decisions to prioritise a particular area of functionality in subsequent development.

For the purpose of our study of social design feedback, a viable approach to the assessment of usability evaluation output may be that of Følstad and Knutsen [35]. They presented a study where an online survey tool was used to collect design feedback from more than 200 user participants across four student design cases. They assessed the design feedback using two approaches: (a) The feedback was rated on usefulness by the involved student designers, and (b) the feedback was categorised according to its qualitative characteristics – *positive*, *negative*, *constructive*; constructive feedback included suggestions on needed design changes. About one-third of the design feedback was rated as useful. The feedback categorised as *constructive* was judged by the student designers as more useful than the other feedback. Negative feedback was judged as slightly more useful than positive feedback, which was not judged as useful at all. In follow-up interviews, the student designers were concerned with the lack of detail in the design feedback and suggested online dialogue between the participants as a means to improve the level of detail in the feedback.

A strength of the analysis scheme suggested by Følstad and Knutsen [35] is that it supports analysis of design feedback in the context of usability evaluations. In particular, *usefulness* may be seen as an early measure of the possible impact of design feedback; design feedback that is not seen as useful will not have any impact on the subsequent design process and, conversely, design feedback seen as useful is likely to have an impact if this is feasible within the practical constraints of the design process. Likewise, the qualitative characteristics *negative* and *constructive* correspond to key outputs of a usability evaluation: usability problems and change suggestions. Though other schemes for analysing data from online social interaction exist, such as the one presented by Agichtein, Castillo, Donato, Gionis, and Mishne [36], the scheme of Følstad and Knutsen is used in this study as it has been developed particularly for the context of usability evaluations.

### **The effect of the key features of social design feedback**

Though the effect of the two key features of social design feedback has not been sufficiently studied within the field of HCI, research from other fields may provide some indications of the effects to be expected.

#### ***The effect of interaction between user participants and development team representatives***

Social design feedback supports interaction between user participants and development team representatives. For the user participants, such interaction will serve as feedback on their contributions from the development team. Research on online social networks indicates that visible feedback, in particular others' comments, increases the motivation to make future contributions [37-40]. In the field of online learning communities, moderators' comments and summaries have been found to strengthen collaboration [41]. Likewise, in the field of online political debate, it has been found that the presence of an active moderator may increase the quality of the debate [42].

For the development team representatives, interaction with user participants will serve as an opportunity to acknowledge good feedback, ask follow-up questions, and provide direction for future comments. User participants are likely to have an imperfect

understanding of the kind of design feedback that is expected from them. Consequently, the feedback from the development team representatives may have an *uncertainty reducing function* [43], clarifying what kinds of contributions are relevant to and appreciated by the development team. Such clarification may be valuable for improving the usefulness of the user contributions gathered in social design feedback.

#### ***The effect of access to other participants' contributions***

In social design feedback, user participants are given immediate access to other participants' contributions to enable participants to build on each other's contributions. The literature on electronic brainstorming provides relevant insight concerning this feature of social design feedback. Studies on electronic brainstorming indicate that access to others' contributions may have a *synergy* effect; that is, ideas from one participant may trigger new ideas in others [44]. Synergy seems to be dependent on there being a sufficiently high number of participants in the group – DeRosa, Smith and Hantula [45] suggested more than eight – as well as sufficient time for each participant to take advantage of the potential synergy from others' ideas [44]. Consequently, access to others' contributions may be expected to have a beneficial effect on the output of social design feedback if the number of participants is sufficiently high and the participants spend enough time to be able to use each other's contributions as a basis for their own feedback.

Access to others' contributions may also invoke the detrimental effect of *social loafing* [46]; that is, the tendency of individuals to perform worse when part of a group than on their own. Research on electronic brainstorming suggests that social loafing is reduced if participants are clearly identifiable as individuals by the use of nicknames or pseudonyms [47], as they were in the setup for social design feedback in this study.

#### **Summary: needed knowledge on social design feedback**

The presented background shows three aspects of social design feedback for which we need new knowledge.

Firstly, we need knowledge concerning the output to be expected from social design feedback, in particular for studies exploiting the potential of an online medium to involve large numbers of users. Knowledge is needed both on the usefulness of such output and on its qualitative characteristics.

Secondly, we need knowledge concerning the interaction between development team representatives and user participants, a key feature of social design feedback. No studies have previously addressed the effect of such interaction. However, studies from other fields indicate that such interaction may be beneficial.

Thirdly, we need knowledge concerning the effect of user participants' access to other's comments, the second key feature of social design feedback. No studies have previously studied this effect systematically. However, studies on online brainstorming indicate that access to other's contributions may lead to beneficial synergy.

#### **Research questions**

Our research questions are formulated so as to address each of the three knowledge needs summarized above.

**RQ1: What is the usefulness and qualitative characteristics of social design feedback?**

Following Følstad and Knutsen [35], we wanted to explore the output of social design feedback in terms of its usefulness and qualitative characteristics. The exploration should provide insight in the potential downstream value of the output from social design feedback. Furthermore, it should provide a more comprehensive exploration of the qualitative characteristics of such output than what has been provided in previous studies [14,26,27]. It was seen as important to conduct this exploration with a sufficiently high number of participants so as to take advantage of the capacity for large scale user involvement in social design feedback.

**RQ2: How does the active participation of a moderator affect the output of social design feedback?**

Social design feedback allows two-way interaction between user participants and development team representatives. Our study focused, for reasons discussed in the Method section below, on the possible interaction between user participants and a moderator serving as the recipient of feedback to the development team. On the basis of earlier work on uncertainty reduction [43], we hypothesised that moderator feedback on user participants' comments would increase the usefulness in the design feedback. Furthermore, we hypothesised that moderator feedback would increase the proportion of suggestions for change or redesign in the design feedback, as constructive user feedback previously has been found to be closely associated with high-usefulness comments [35].

**RQ3: How does access to others' contributions affect the output of social design feedback?**

Access to other participants' contributions can lead to synergy, as is seen in research on electronic brainstorming [44]. Consequently, we hypothesised that giving user participants' access to others' contributions would increase the usefulness of their design feedback. We also hypothesised that this increase in usefulness would be associated with an increase in the proportion of suggestions for change or redesign in the design feedback. However, such synergy may depend on multiple factors, for example on the participants spending enough time on others' feedback [48].

## **Method**

To study the usefulness and qualitative characteristics of the output of social design feedback we explored such output in two ICT design cases (RQ1). In both cases, the participants contributed design feedback as free text comments in an online environment. All comments were displayed in discussion threads adjacent to the visual representation of the design. The environment did not support annotations in the visual representation thereby avoiding problems with visual clutter.

To allow conclusions on the effects of the key features of social design feedback (RQ2 and RQ3), each case was designed as a  $2 \times 2$  factorial experiment where the participants were randomly assigned to one of four conditions (see Table 1). This design also allowed us to check for interaction effects between these two features, though no such effects were hypothesised.

The purpose of conducting our study in two cases was to generalize and challenge our findings. Our research design, however, does not support conclusions about differences

**Table 1 Overview of the experimental conditions**

		Direct view of other participants comments' prior to own comment	
		Yes	No
Moderator providing feedback on participants' comments	Yes	Condition 1	Condition 2
	No	Condition 3	Condition 4

between the two cases. Rather, we assumed that the two cases would yield the same experimental findings.

### Cases

The cases were from different sectors: football (Case 1) and telecommunications (Case 2). Furthermore, the cases were in different design phases with designs of different levels of maturity: a running prototype website and a non-functional visualisation of a competing user interface design (Case 1), and pre-prototype concepts presented through simple storyboards (Case 2). For both cases, the purpose of the design feedback was to guide subsequent design and development. The purpose of selecting cases that differed on multiple characteristics was to study social design feedback in different contexts, thereby checking for the potential threat to external validity due to only using one particular study setting [28]; conducting the same experiment in two different contexts allowed us to challenge the findings of one case with reference to the findings of the other.

*Case 1* was an early running prototype of a blog feed aggregator for a Norwegian premier league football club (see Figure 1). Its purpose was to provide team supporters one place to be updated on blogs concerning the team. The study commenced on the day of the launch of the prototype. The participants were asked to provide feedback on (a) the running prototype and (b) a visual presentation of an alternative user interface for the blog aggregator.

*Case 2* was about novel concepts for social text-based communication on mobile devices, designed at the Oslo School of Architecture and Design. The concepts were presented as story-boards outlined as cartoon strips. The design feedback was meant to support prioritising concepts and subsequent design work. See Figure 2 for an example concept from Case 2.

### Setup for social design feedback

The online environment for social design feedback consisted of a set of webpages, each structured as a frameset with four frames containing (a) instructions, (b) a free text comment field and a discussion thread, (c) the object of feedback, and (d) buttons to navigate between feedback topics. Figure 3 shows an example webpage.

The instruction frame was placed horizontally at the top of the webpage. The instructions were intended to be short and precise, while allowing room for discussion and reflection. In the feedback topic in Figure 3, the instructions read: "The look and functions of the blog portal. Currently there is a lot of text on the blog portal. The content is presented under headings indicating its origin. Do you have suggestions on how the blog portal should look in the future? Ideas on functionality? Thoughts on how the blog portal should be tied to other webpages, for example, football club webpage and Facebook?"





Figure 1 Home-page of the prototype web site evaluated in Case 1.

The comment field and the discussion thread were placed vertically at one side of the screen; the comment field above the discussion thread, and the thread sorted chronologically with the newest comment on top. Each comment in the threads included the contributor's nickname and a timestamp. Commenting on others' comments was available as a "reply" function associated with each comment in the thread. When commented on, a participant received an e-mail notification with a description of the reply and a link to access the relevant feedback topic.

The object of feedback was presented in the frame next to the comment field and discussion thread. When the object of feedback was a website (parts of Case 1), the participants could navigate the website while retaining the frames containing the instructions, comment field, and discussion thread. In parts of both cases, the objects of feedback were presented as images.

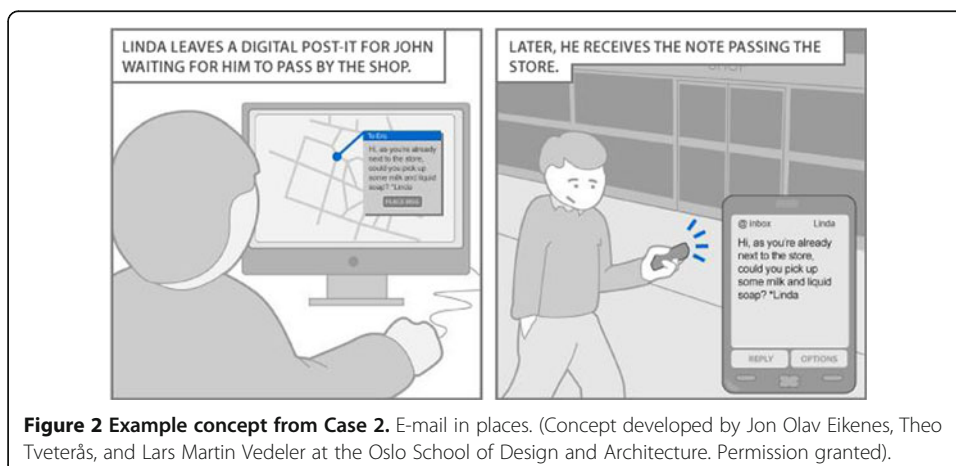


Figure 2 Example concept from Case 2. E-mail in places. (Concept developed by Jon Olav Eikenes, Theo Tvetervås, and Lars Martin Vedeler at the Oslo School of Design and Architecture. Permission granted).



The navigation buttons, *next* and *previous*, were located immediately below the discussion thread. The participants could move between feedback topics at will.

Case 1 included five feedback topics; Case 2 included six. The feedback topics concerned different functionalities and design suggestions and were selected in cooperation with the case owners. All participants were shown the feedback topics in the same order and asked to contribute feedback for at least three of the topics. The participants were allowed to move on to the next topic even if they had not contributed feedback to the current one.

We were aware that the setup for social design feedback would be unfamiliar to the study participants, and consequently included explanatory texts for guidance and support in the invitation and recruitment process as well as for each feedback topic. The participants were explained that their feedback was meant to advise future design. They were also, for each discussion thread, asked to provide feedback in a manner consistent with this purpose; for example, to provide their “impression of the design,” what they perceive to be “good / bad” in the design, or to “suggest changes.” In particular, change suggestions and problems (“bad” in the design) were meant to trigger useful feedback.

### Participant recruitment

In Case 1, invitations were included in an electronic newsletter to the football club supporters. In Case 2, participants were invited from a national market research panel provided they reported that they used e-mail on their mobile phones several times a week or more. The recruitment strategy allowed us to get user participants experienced with similar solutions.

Upon accepting the invitation, the participants clicked a link taking them to the social design feedback solution where they entered background data, including a nickname

and an e-mail address for notifications, and were given instructions. No one participated in more than one case.

As compensation for their time, all participants entered a lottery with a prize worth about \$300. The participants' chances in the lottery were not dependent on the content of the participants' comments. Participant fallout was calculated as the proportion of participants entering a nickname and e-mail address but not providing any design feedback. The fallout rate was 22% in Case 2 (unavailable for Case 1). We assume that the fallout rate was mainly due to the novelty of this kind of data collection and that some participants upon registration found that they did not want to participate because of the study setup. Indeed, the setup was duly described in the study invitation, but some participants may have overlooked this information.

#### **Data collection and analysis – to explore the output of social design feedback (RQ1)**

The study data consisted of the comments made in the environment for social design feedback, as well as the participant background data. To explore the output of social design feedback (RQ1), the comments were analysed in terms of their *usefulness* and *qualitative characteristics*, following Følstad and Knutsen [35]. This choice was made as the analysis scheme used by these authors has been developed particularly to analyse online design feedback.

**The usefulness of the comments** as input to a design process was rated by two independent analysts. Both analysts rated all participant comments to check inter-rater agreement. This rating was assumed to require special training in user-centred design, as judgments on the usefulness of design feedback require experience and understanding of the design process. One of the analysts had been working as a concept designer in an IT development company for three years. The other (the first author of this paper) had been working as a researcher on user-centred design in IT for ten years.

None of the analysts was responsible for the designs in any of the two cases, but one (the first author of this paper) had served as moderator in the two cases. Their distance to the design process allowed the design feedback to be rated without being affected by spurious idiosyncrasies in the two design processes, as, for example, could happen if a development team representative were to rate design feedback corresponding to design ideas previously suggested by this representative but for some reason not being pursued in the current design. Avoidance of such idiosyncrasies is arguably beneficial to the reliability of the rating. However, insufficient understanding of the designs could compromise the validity of the rating. Consequently, prior to the ratings, both analysts familiarised themselves thoroughly with the designs at hand.

Usefulness scores were calculated as the average of the analysts' ratings on two scales: *Relevance* and *Inspiration*. *Relevance* was defined as "the comment directly concerns a key part of the solution or its context of use"; *Inspiration* was defined as whether "the comment is suited to contribute to a change in the design." The two scales were motivated by Amabile's [49] work on creativity assessment, where the main components of creativity are held to be relevance and novelty.

Both aspects of usefulness were rated on scales from 0 to 10, the latter being the best. For a comment to receive a *Relevance* score above 5, it should be judged as suited to provide new insight. For it to receive an *Inspiration* score above 5, it should be judged to build the idea further, not just motivate the removal of something that does not work. Inter-rater agreement  $r$  ranged from 0.65–0.83 for the two scales across the two cases. The correlation  $r$  between *Relevance* and *Inspiration* ranged between 0.84–0.90.

The rating was conducted blind; that is, no information was provided during analysis on the conditions which the comments belonged to. This was done to avoid possible biases associated with expectations related to the different conditions. However, as one of the raters was also the moderator of the cases, this rater might remember which comments belonged to which condition. To check this possible source of bias, an additional set of usefulness analyses was run with the usefulness scores obtained only from the other analyst. These additional analyses showed the same pattern as the analyses using the average usefulness scores. The *Usefulness* scores obtained by the other analyst were only found to be significantly affected ( $p < .05$ ) by the *Moderator* conditions in both cases, but not affected by the *Direct view* conditions in any of the cases. Thus, we can rule out the possibility that the analysis of usefulness was biased by analyst expectations.

**The qualitative characteristics of the comments** were coded by two independent analysts. As this was not expected to require special training in HCI, the analysis was done by two student assistants who received initial training and piloting. Both analysts rated all participant comments to check inter-rater agreement.

The comments were coded on:

- *Negative/problem* (yes/no). Comments expressing a general negative attitude to the function or solutions manifested in the design and/or identifying a particular problem with the same function or solution. (Inter-rater agreement: Cohen's kappa = 0.87 indicating almost perfect agreement [50]).
- *Suggestions* (yes/no). Comments explicitly suggesting a change or redesign to the function or solution manifested in the design (Inter-rater agreement: Cohen's kappa = 0.76 indicating substantial agreement [50]).

*Negative/Problem* was initially treated as two distinct characteristics during coding, but merged because the analysts expressing difficulties differentiating between *negative* and *problem*, something that also was reflected in lower inter-rater agreement for these initial characteristics (Cohen's kappa 0.68 and 0.55, respectively).

Comments containing content corresponding to a coding category, for example a *suggestion*, were coded *yes* for this category. All other comments were coded *no*. The exception to this coding system was the initial coding category *negative*, which was coded as positive, neutral, or negative, upon which *negative* was recoded as *yes*, and *positive* and *neutral* were recoded as *no*. In the case of disagreement between the raters, only those comments coded *yes* by both raters were counted within a given category.

A comment could potentially be coded as both *Negative/Problem* and *Suggestion*. Such overlap did not have implications for the subsequent analyses, as the different codes were never included in the same analysis.

The comments were also coded on other characteristics. These were *references to similar solutions*, *references to other participants*, *information on the intended context of use*, and *comments on good details in the design*. The first three of these were not included in the following analysis, as each of them covered less than 10% in any of the cases. The last of these was not included as the inter-rater agreement for this characteristic was too low.

All coding of the comment characteristics was conducted blind. None of the analysts were aware of which conditions the different comments were made in. This was done to control for possible biases associated with analyst expectations.

### **Experimental conditions - to investigate the effect of the two key features of social design feedback (RQ2 and RQ3)**

For each case, RQ2 and RQ3 were investigated by implementing the four experimental conditions of Table 1 as four instances of the online environment. After being presented to the study instructions, the participants were randomly directed to one of the four conditions by a JavaScript. The participants were not aware of there being different conditions.

#### ***Moderator feedback***

In Conditions 1 and 2, a moderator provided feedback on the participants' comments. Using a moderator as the main point of interaction between the user participants and the development team made it practically possible to conduct and analyse the interaction in a systematic manner, this being a necessary condition for the experimental setup. The moderator gave feedback as comments in the discussion thread, clearly specifying the nickname of the user participant being addressed. The moderator feedback was phrased as praise/thank you, enquiries for more detail, or requests for others to offer their viewpoint. More than two-thirds of the participants in the moderated conditions received moderator feedback on one or more of their comments. In Conditions 3 and 4, no such moderator comments were given. Examples of moderator feedback are given in Table 2.

#### ***Direct view of others' comments***

In Conditions 1 and 3, the participants could see the discussion thread with the other participants' comments before making a comment. In Conditions 2 and 4, the participants could see a given discussion thread only after having made a root comment, a *root comment* being understood as a participant's first comment for a particular feedback topic. As all participants were allowed to contribute more than one comment in any thread, the participants in Conditions 2 and 4 could reply to other participants' comments if they made a root comment to get to see the thread. The direct view of other participants' comments was meant to allow participants in Conditions 1 and 3 to benefit from synergy with the other participants' comments when making their root comment in the thread. Participants in Conditions 2 and 4 were not allowed such a potential benefit of synergy when making their root comment for any feedback topic. To check that participants' in Conditions 2 and 4 did not make bogus root comments just to get access to other participants' comments, we particularly reviewed the participant comments to detect such bogus statements.

**Table 2 Examples of moderator feedback**

Moderator feedback types	Example feedback case 1	Example feedback case 2
Praise/thank you	Hi <b>@nemezizz</b> . Thanks for your good feedback on the blog portal. (AsbjornF – partly responsible for the study :-)	Really like your rich reflections on the function for delayed sending of e-mail, <b>@idieh</b> . Free choice of time for sending, and that it should distract from “send” (as in “send now”). Thanks! (AsbjornF – partly responsible for the study :-)
Enquiry for more detail	Good feedback on the blog portal, <b>@csandoy</b> . Would be great to get to know more on what you think of the new blog portal suggestion vs. the existing portal, and why? (AsbjornF – partly responsible for the study :-)	Thanks for your enthusiastic feedback on the function for delayed sending of e-mail, <b>@Lisa</b> . Would be great to get some examples on how you would use this. Is it possible to ask you for a couple of these? (AsbjornF – partly responsible for the study :-)
Request others to offer their viewpoint	Hi <b>@Anon1</b> . Thanks for telling us that we need more pictures. Anybody else having an opinion on this? (AsbjornF – partly responsible for the study :-)	Some of you, e.g. <b>@bestorp</b> , suggest automatic zoom. I definitely see the point, but I wonder if one easily gets a kind of key-hole effect where you see too little of a long message. [...] What do all of you think about this? (AsbjornF – partly responsible for the study :-)

## Results

### Initial analyses

In total, 250 participants took part in the study and provided one or more comments; 86 participants in Case 1 (35% females; mean age = 29 years,  $SD = 13$  years) and 164 participants in Case 2 (32% females; mean age = 28 years,  $SD = 5$  years). The mean number of comments per participant was 3.0 in Case 1 ( $SD = 1.6$ ) and 4.7 in Case 2 ( $SD = 2.0$ ). No participant provided more than 13 comments.

The participants provided 1036 comments across the two cases. Of these comments, 19 were discarded for being unintelligible or the same comment submitted twice by the same participant. The remaining 1017 comments were analysed; 980 of these were root comments, whereas 37 were follow-ups. As a *root comment* is understood as a participant's first comment for a particular feedback topic, a participant could have as many as five or six root comments depending on the number of feedback topics in the case.

The moderator contributed a total of 103 comments across the two cases. Details of the distribution of participant and moderator comments are given in Table 3.

Only a small proportion of the participant comments were follow-ups. This was surprising to us and will be treated further in the Discussion section. Because of this lack of follow-up comments, the subsequent analyses include only the participants' root comments. We have done this to simplify interpretation, as the low number of follow-up comments (37) makes it difficult to draw general conclusions on the nature of such comments.

### The usefulness and qualitative characteristics of the comments (RQ1)

The usefulness of the participant comments was skewed towards the lower end of the scale. In Case 1, only 17.1% of the comments received usefulness scores above 5, and in Case 2, only 17.0% were above 5; this means that 167 of the 980 comments were judged as potentially giving new insight and/or being suited to build the design further. Details are presented in Table 4.

Example comments for different levels of usefulness, as well as different characteristics, are presented in Table 5. All comments are from Case 2, from the feedback topic on the concept presented in Figure 2. The instruction for the topic was as follows: "Suggestion 2: E-mail in places. If the recipient has GPS on the phone you can send e-mails that are only received when the recipient is where you want the message to be read; as in the example in the cartoon below. How would you use such a function? What should we think of when developing this function further?"

### The effect of Moderator (RQ2) and Direct view of others' contributions (RQ3)

The effects of an active moderator and direct view of other participants' contributions were analysed by using data on comment *Usefulness*, as well as the comment characteristics

**Table 3 Distribution of participants and comments across the cases**

	Case 1 (n = 86)	Case 2 (n = 164)	Total (n = 250)
Participant root comments	246	734	980
Participant follow-up comments	8	29	37
Moderator comments	42	61	103

**Table 4 Distribution of usefulness scores**

	Usefulness 0–2.50		Usefulness 2.51–5.00		Usefulness 5.01–7.50		Usefulness 7.51–10		All comments	
	Count	%	Count	%	Count	%	Count	%	Count	%
Case 1	112	46%	92	37%	41	17%	1	0%	246	100%
Case 2	330	45%	279	38%	103	14%	22	3%	734	100%

*Negative/Problem* and *Suggestion* as dependent variables in two-way ANOVAs, based on the 2 × 2 factorial experimental design.

The ANOVAs were carried out on the level of individual participants. The dependent variables were calculated as follows:

- Individual scores on *Usefulness* were calculated as the proportion of a participant’s comments with usefulness score above 5. This approach to individual scoring, instead of, for example, using the mean of the usefulness ratings for an individual participant’s comments, was chosen to clearly differentiate between highly useful comments and other comments. In the case that a larger proportion of the comments had received high usefulness scores, we would have chosen the mean of the usefulness ratings. However, to check whether the analysis would have yielded a different result if we had chosen a different calculation, we also replicated our analyses using *mean usefulness rating* (as outlined above) and *number of usefulness ratings above 5* (a measure which disregards the number of low-usefulness scores provided by an individual).
- Individual scores on *Negative/Problem* and *Suggestion* were calculated as the proportion of a participant’s comments being coded as *Negative/Problem* and *Suggestion* respectively.

**Descriptive analyses**

To provide an initial overview of the findings, prior to presenting the results from the ANOVAs, we present mean scores for *Usefulness*, *Negative/Problem*, and *Suggestion* as bar charts for each of the two independent variables (*Moderator* and *Direct view*). This mode of presentation, where the independent variables are seen independently, is justified as no interaction effects were observed in the ANOVAs (to be presented below).

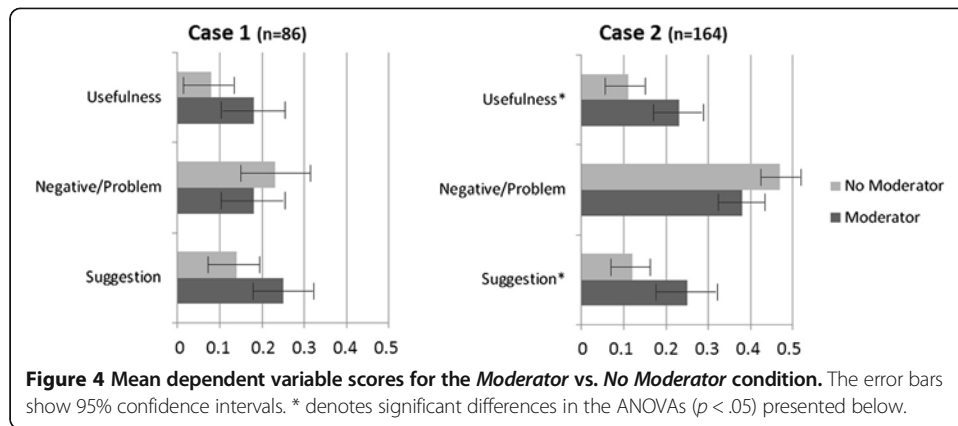
The mean scores for the dependent variables in the *Moderator* and *No Moderator* conditions are presented in Figure 4. We see that in both cases, *Moderator* is associated with higher scores on *Usefulness* and *Suggestion*. However, this difference only approached significance in Case 1. Also, *Moderator* is associated with lower scores for *Negative/Problem*, but these differences are not statistically significant in either of the cases. The results indicate that *Moderator* had a positive effect on the usefulness of the comments, as well as the participants’ tendency to contribute suggestions. However, the presence of a moderator did not have a positive effect on the participants’ tendency to contribute negative- / problem-oriented comments.

In the same manner, the mean scores for the dependent variables for *Direct view* and *No Direct view* are presented in Figure 5. There seems to be no common pattern across the two cases for *Usefulness* and *Suggestion*. The scores for *Negative/Problem* are higher for the *Direct view* condition in both cases, though this difference was statistically



**Table 5 Example participant comments, Case 2**

Usefulness rating	Qualitative characteristic	Example comments
0 - 3	Negative/Problem	John turning off the GPS may have catastrophic consequences.
	Suggestion	NA
	Both	NA
	Neither	This is perfect. It is much easier than having notes or similar lying around to remember things.
4.5 - 5.5	Negative/Problem	I would not use this kind of function. It is a type of function that limits the freedom of people to move about as they wish, and as Anonymous2 writes, this assumes that the recipient is moving in a "given pattern" ...
	Suggestion	If this is to be interesting, the recipient needs to be updated on whether the message is read or not.
	Both	A bit too 1984, except that it is Linda who sees you and not big brother. I would not use this function and would turn it off if it was default. If you absolutely need this, it would be good if it could (if possible) find its position by way of mobile signals, this would save battery capacity dramatically. (Think HTC Hero does this if the GPS is turned off and one uses the map, with a 5–800 meter error margin, but still).
	Neither	This is a function I would use for road descriptions and meeting information.
7 - 10	Negative/Problem	Everything is wrong with this functionality. First the GPS reception is poor in the pocket or in the purse. Second, this will drain the phone battery empty even faster. Third, I doubt that John finds it nice that Linda can do 24/7 surveillance on him. All in all GPS on the phone is a double-edged sword in a world where it seems as if the EU data directorate may be accepted (use all your influence and VOTE AGAINST!). This can be useful for Taxi drivers and couriers etc. as a working tool but involves too much surveillance for my liking.
	Suggestion	Would be a cool function that can be used for a lot of useful things. To limit misuse the user should choose who are allowed to do this, for example persons in the contact list, anybody who wants, just some contacts, block some contacts, etc. And for it to work everywhere it needs to use a radius that is bigger than the exact shop door, maybe 50–100 meters from the chosen point.
	Both	I find this superfluous. It should be designed to fit better for people not familiar with the area, rather than people living nearby and therefore knowing where the closest grocery store is. Maybe a kind of a GPS showing guests or visitors where they should go from the nearest bus stop to the party they are to visit.
	Neither	NA



significant only in Case 2. The results indicate that *Direct view* had a positive effect on the participants' tendency to contribute negative/problem-oriented comments, whereas there was no such effect on either the usefulness of the comments or the participants' tendency to contribute suggestions.

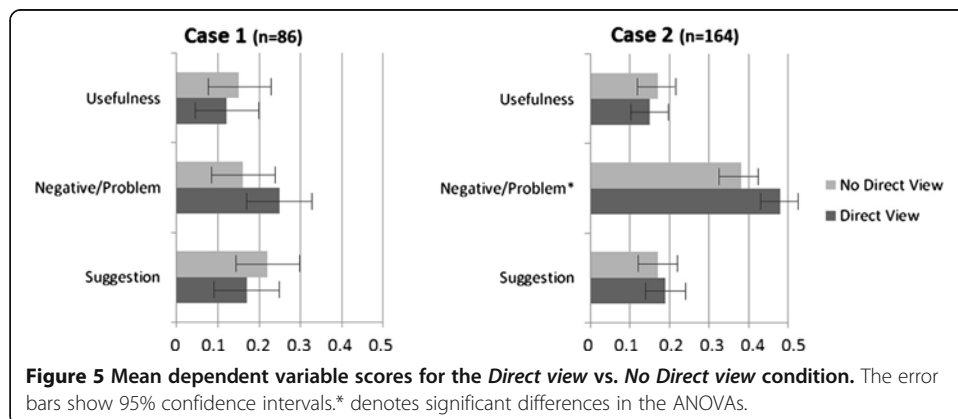
*Moderator* and *Direct view* clearly had different effects on the participants' feedback. Whereas the presence of a moderator increased usefulness and the participants' tendency to contribute suggestions, a direct view of other participants' comments increased the participants' tendency to contribute negative or problem-oriented comments.

#### The effect of Moderator and Direct view on Usefulness

The effects of *Moderator* and *Direct view* on the dependent variable *Usefulness* were analysed in two-way ANOVAs, one ANOVA for each case. In Case 2, *Usefulness* was significantly higher in the *Moderator* conditions than in the *No Moderator* conditions. In Case 1, we also found a difference in *Usefulness* between the *Moderator* and *No Moderator* conditions, but here it only approached significance ( $p = .05$ ).

*Usefulness* was not affected by *Direct view* in either of the cases. No interaction effect between *Moderator* and *Direct view* was found. See Table 6 for details.

The square root of the effect size  $\omega^2$  is comparable to  $r$  [51]. Following Cohen's rules of thumb [52], the effect sizes associated with *Moderator* were small; the effect sizes associated with *Direct view* and the interaction term were negligible.



**Table 6 The effect of Moderator and Direct view on Usefulness; results from two-way ANOVAs**

Independent variables	Usefulness					
	Case 1			Case 2		
	F(1,82)	p	$\omega^2$	F(1,160)	p	$\omega^2$
Moderator	3.87	.05	.02	9.25	<.01	.02
Direct view	.21	.65	.00	.09	.77	.00
Interaction	.00	.97	.00	.05	.82	.00

We replicated the ANOVAs with two alternative measures for usefulness, *mean usefulness rating* and *number of usefulness ratings above 5*, to check that our choice of usefulness score did not unduly impact our analysis. The results of these analyses paralleled those of Table 6. We found no effect of *Direct view* in either of the cases for either of the two alternative scores. We found a significant effect of *Moderator* in Case 2 for both the alternative scores. Furthermore, the effect of *Moderator* approached significance in Case 1 for *number of usefulness ratings above 5* ( $p = .08$ ). For *mean usefulness rating*, however, the effect of *Moderator* was not significant ( $p = 0.17$ ).

**The effect of Moderator and Direct view on Negative/Problem and Suggestion**

The effect of *Moderator* and *Direct view* on the dependent variables *Negative/Problem* and *Suggestion* were also analysed in two-way ANOVAs. Two ANOVAs, one for each dependent variable, were run for each case.

For *Negative/Problem*, the cases were not consistent concerning the findings. In Case 1, *Negative/Problem* was not significantly affected either by *Moderator* or *Direct view*. In Case 2, *Negative/Problem* was reduced in the *Moderator* condition, the reduction bordering significance ( $p = .05$ ), and significantly increased in the *Direct view* condition. That is, in Case 2 when a moderator was present, the participants generated a smaller proportion of comments containing dislikes or concerns, whereas when they had immediate access to the other participants' contributions, the participants generated a larger proportion of such comments. No interaction effect was observed between *Moderator* and *Direct view*. See Table 7 for details.

The same kind of inconsistency between the cases was not found for *Suggestion*. *Suggestion* was higher for the *Moderator* conditions than the non-*Moderator* conditions in both cases, but the difference was significant only in Case 2. In Case 1, this difference only bordered significance ( $p = .07$ ). *Suggestion* was not affected by *Direct view* in either of the cases. No interaction effect was observed between *Moderator* and *Direct view*. See Table 8 for details.

**Table 7 The effect of Moderator and Direct view on Negative/Problem; results from two-way ANOVAs**

Independent variable	Negative/problem					
	Case 1			Case 2		
	F(1,82)	p	$\omega^2$	F(1,160)	p	$\omega^2$
Moderator	.94	.34	.00	3.85	.05	.01
Direct view	1.73	.19	.00	5.25	<.05	.01
Interaction	1.54	.22	.00	.98	.32	.00

**Table 8 The effect of Moderator and Direct view on Suggestion; results from two-way ANOVAs**

Independent variable	Suggestion					
	Case 1			Case 2		
	<i>F</i> (1,82)	<i>p</i>	$\omega^2$	<i>F</i> (1,160)	<i>p</i>	$\omega^2$
Moderator	3.28	.07	.01	11.93	<.01	.03
Direct view	.53	.47	.00	.64	.43	.00
Interaction	.15	.33	.00	.50	.48	.00

**Non-parametric replications of the findings**

Due to the fairly low proportion of comments scoring above 5 on *Usefulness*, the data for the *Usefulness* score used in the analyses did not follow a normal distribution. This is a violation of the assumptions of ANOVA. ANOVA has been found to be robust against such violations as long as the experimental groups are of equal size [51]. Even so, we found it desirable to conduct non-parametric tests of all group differences as an additional verification of our findings.

For all three dependent variables we conducted Mann–Whitney U tests for the effect of *Moderator* and *Direct view* respectively. We conducted two sets of these tests, one for each case. The output of the tests followed almost exactly the pattern of the ANOVAs. All significant differences observed in the ANOVAs were also found in the Mann–Whitney U tests. Furthermore, the Mann–Whitney U tests showed non-significance for all non-significant differences observed in the ANOVAs; the only exception was that the Case 2 analysis for *Negative/Problem*, which only bordered significance in the ANOVA ( $p = .05$ ), was found to be significant in the Mann–Whitney U test ( $p < .05$ ).

**Discussion**

The two cases have provided insights into the kind of output that social design feedback may give (RQ1), as well as the effect of the key features of social design feedback (RQ2 and RQ3). These insights have theoretical implications and help advise about the practical implementation of social design feedback.

**The usefulness and qualitative characteristics of the output of social design feedback (RQ1)**

The output of social design feedback was explored in terms of its usefulness and qualitative characteristics. Across the two cases, 167 comments (17%) received usefulness scores above 5, indicating that they provided new insights and/or concerned how to build the idea further. Furthermore, 201 comments (21%) contained change suggestions which have previously been found to be the most useful type of design feedback [35]. The user participants also provided 395 (41%) comments containing negative issues or perceived problems. Social design feedback clearly can generate useful output. We find the user participants to be able to generate feedback holding characteristics corresponding to what is expected of the output from a usability evaluation, in particular, usability problems and redesign suggestions.

However, showing that the output of social design feedback can be useful is only half the story. The output of social design feedback can also be littered with comments of

low usefulness. Across the two cases, 813 comments received usefulness scores of 5 or below; that is, four-fifths of the comments did not provide new insight or make constructive input suited to drive the design process. Also, nearly half the comments contained neither suggestions nor participant concerns. An important challenge concerning social design feedback will be to increase the proportion of useful comments or to effectively filter out comments that are not useful in the subsequent development process.

The high frequency of comments not useful to subsequent development indicates that the majority of the participants were not able to comply with the intended purpose of their participation; that is, to provide feedback that could serve to drive the design process. This lack of compliance may be due to a lack in the participants' understanding of the intended purpose of social design feedback. The importance of sufficient guidance and support for user participants providing design feedback in usability evaluation have previously been accentuated [23]. We sought to provide such guidance and support by explaining the purpose of the study in the invitation and recruitment process, as well as in the descriptions for each feedback topic. Possibly, however, the social design feedback method may require even more in the way of guidance and support for the user participants than what was provided in our instantiation of the method.

#### **The effect of an active moderator (RQ2)**

An active moderator clearly affected the participants' contributions. Participants in the *Moderator* conditions received higher usefulness scores on their comments and provided more comments including suggestions for change or redesign. The hypothesis for RQ2 is supported, though the effect of an active moderator only approached significance in Case 1. As an active moderator both improves usefulness scores and causes more suggestions to be made, it seems reasonable to speculate that an active moderator improves the usefulness of design feedback in particular by guiding the participants to provide more constructive feedback, that is, more suggestions. This interpretation reverberates findings on the beneficial effect of moderators from online learning communities, where moderator summaries have been found to enhance collaboration [41], and online political debate where the beneficial effect of moderators have been attributed to their ability to focus the discussion [42].

It is fascinating that the beneficial effect of an active moderator was found even though we only analysed the participants' root comments. The analysed participant comments were affected only by the moderator's comments made previously in response to other participants, not by moderator responses to their own comments. Consequently, the effect of an active moderator is clearly not limited to the interaction between each individual participant and the moderator.

The beneficial effect of an active moderator may be explained by its potential *uncertainty reducing function* [43]. That is, the potential of the moderator's comments to guide the user participants towards providing comments in line with the goal of the design feedback as seen from the moderator's perspective. As the moderator's comments contain praise for useful participant comments, enquiry for more detail, and requests for others to offer their viewpoint, these comments may clarify to the participants what is expected from their participation.

However, an active moderator may also serve as a motivational factor for the participants. We know that others' comments may increase participants' motivation to make future contributions [37-40]. In this light, moderator comments may motivate not only those that are commented on, but possibly also other participants seeing that the design feedback is actually being read and acted upon by its recipients.

Though our hypothesis for RQ2 was supported, we observed a non-hypothesised difference in the effect of an active moderator on the two qualitative characteristics of the design feedback, *Suggestion* and *Negative/Problem*. The moderator was associated with higher scores on *Suggestion* and lower scores on *Negative/Problem*, though these differences were not significant for three of the four analyses. This unexpected finding may suggest that an active moderator does not trigger participants' tendency to provide negative feedback or unfiltered voicing of concerns. It may, possibly, be speculated that an active moderator can help the participants transform their negative feedback or concerns into suggestions. If so, this could help explain the beneficial effect of an active moderator. However, as this finding was both unexpected and only partially underpinned by statistically significant differences, it should not be regarded as valid knowledge. Nonetheless, it may serve as inspiration for future research on the causes for the effect of an active moderator.

#### **The effect of seeing other participants' comments (RQ3)**

Seeing other participants' comments prior to making one's root comment did not have a significant effect on either the usefulness of the participants' comments or on the participants' tendency to provide suggestions. This was contrary to our hypotheses for RQ3.

The number of participants in each condition was sufficiently high to allow for the synergy predicted on the basis of research on electronic brainstorming. However, we did not control for the time spent by the participants on the study. Most likely the participants spent some time reading others' comments in the study, given the effect of an active moderator, but we do not know whether the time spent was sufficient for synergy. Possibly, seeing others' contributions would have had a positive effect, given that we had introduced mechanisms or constraints to make sure that the participants spent enough time to utilise each other's contributions fully.

Seeing other participants' comments, however, had a positive effect on *Negative/Problem* in one of the cases. That is, the participants in the *Direct view* conditions contributed more negative and/or problem-oriented feedback than did those of the *No Direct view* conditions. It may be speculated that whereas participants in the *Moderator* conditions to a greater degree utilised the guidance provided by the moderator comments to adjust their contributions, and hence provided more constructive feedback, the participants in the *Direct view* conditions to a greater degree were left with the other participants' comments to adjust their contributions, which did not provide the needed guidance. However, no interaction effect was observed between *Moderator* and *Direct view* even though such an interaction may be inferred from the above speculation.

#### **Differences between the cases?**

RQ2 and RQ3 were investigated through an experimental design conducted within two cases. Our inclusion of two cases was done to generalize and challenge our

findings, our expectation being that the two cases would yield the same empirical findings.

How, then, should we understand the differences observed between the cases? In particular, none of the findings in Case 1 were significant at  $p > 0.5$ , though the pattern of the findings in Case 1 was similar to the pattern of the findings in Case 2 (as can be seen, for example, in Figures 4 and 5).

Our interpretation is that these differences between the cases are likely due to the difference in the number of participants in each case and, consequently, differences in statistical power. As the main effects were small size only, the sample size in Case 1 was insufficient to achieve adequate statistical power. A sample size of 86 in a two-way ANOVA is only sufficient to observe medium to large size effects given a statistical power ( $1-\beta$  error probability) of .80, according to the statistical software G\*Power 3.1.5 [53]. We assume, therefore, that given a larger sample size in Case 1, the differences between the cases concerning statistically significant findings would have been substantially reduced, if not eliminated altogether. In hindsight, it would have been beneficial to include a larger number of participants in Case 1. However, this was judged as impractical at the time of the study due to time-constraints in the recruitment process.

#### **Advice on the practical implementation of social design feedback**

As we have seen, social design feedback elicits a substantial amount of useful feedback for early concepts, visual prototypes, and implemented applications. However, social design feedback also allows user participants to make contributions with low usefulness. Therefore, the successful implementation of social design feedback depends on our ability to either reduce the proportion of low usefulness feedback or to filter out low usefulness comments. Our findings motivate advice on how to reduce the proportion of low usefulness comments. In the following, we summarise three key learning points on the practical application of social design feedback:

- Clearly explain the purpose of the social design feedback to your participants. The participants are likely to be inexperienced in providing such feedback and, consequently, need guidance. Make sure to explain that you want feedback useful for subsequent design activities, in particular, suggestions for changes and redesign.
- Guide your participants by being responsive to participant comments. Comments from development team representatives, such as a moderator, will help the participants understand what kind of feedback you want. Since the participants are affected by moderator comments made to previous participants, it will be particularly important to moderate early in the feedback session in order to establish a norm for what constitutes useful feedback.
- Pay attention to participant motivation. Participants should belong to the main user groups of the solution under development, potentially improving participant motivation. Being responsive to participant comments and commenting on how the participant comments are useful to the design process should also improve participant motivation.

### Limitations

The main limitation of this study is that we were not able to generate substantial interaction between participants and development team representatives. Thereby, some of the conclusions may be limited to social design feedback without such interaction.

The setup for the social design feedback supported asynchronous interaction between participants and moderators. Whenever a participant or a moderator was mentioned in a comment, a notification e-mail was sent to the mentioned person. In total, the moderator made 166 comments to the participants. However, only 37 follow-up comments were made by the participants. This volume of follow-ups was smaller than we had hoped.

This limitation may, in particular, be relevant to findings about the effect of seeing other participants' contributions prior to making one's root comment. The lack of support for the hypothesised effect of *Direct view* may be a consequence of a lack of substantial interaction between the participants, assuming an interaction effect between *Direct view* and the level of interaction.

The participants' limited interaction with each other or the study moderator is, however, not an indication that the social context of the study did not matter. On the contrary, the effect of an active moderator indicates that the participants, at least to some extent, paid attention to other's comments in the discussion thread. This point is particularly demonstrated as the analysis included only the participants root comments on any feedback topic; meaning that for a moderator's comments to have an effect, this had to be caused by the participants reading the moderator's comments on other participants' contributions. Thus, we hold that the studies were social in the sense that the participants were aware of, and to some degree related to, each other's comments.

How to increase participant interaction will be an important issue in future research on social design feedback, as we assume that increased interaction also will increase the level of detail in the feedback and thus its value in the subsequent design process.

### Conclusions and future work

The present study provides new knowledge on the benefits and limitations of social design feedback. However, important future research remains to be done before we have sufficient knowledge about this approach to evaluations with users. We find the following three knowledge areas particularly relevant.

First, we need knowledge on how to improve the interaction between participants and development team representatives in social design feedback. Two strands of research may be relevant for this purpose: (a) research on the effect of the design and layout of the environment for social design feedback and (b) research on process improvements, such as improvements in the instructions to participants as well as in the moderator activity.

Second, we need knowledge on how to filter out low-usefulness comments. Given that we are not able to avoid getting low-usefulness comments we need reliable approaches to easily filter out such comments. Two strands of research may be relevant: (a) research on automatic content analysis for automatic filtering and (b) research on



social filtering where the participants themselves are allowed to vote up and down other participants' contributions.

Third, it will be relevant to study individual differences in feedback usefulness. Given the large volumes of low-usefulness comments, filtering participants on individual differences may be a possible way to improve the ratio of high-usefulness feedback. Furthermore, as some participants may be more prone to engage in interaction with other participants in a social design feedback study, filtering participants on their tendency to engage socially online may provide a possible way to improve the interaction between participants.

Social design feedback is a novel approach to getting design feedback during the design of IT and may complement existing approaches to collecting design feedback. For HCI practitioners, social design feedback represents an opportunity to use the internet to gather design feedback. For HCI researchers, social design feedback may be seen as an exciting new field of method development.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

AF conceived of the study and its design, conducted the data collection and analyses, and drafted the manuscript. KH participated in the design of the study and in drafting the manuscript. PU participated in the analyses. All authors read and approved the final manuscript.

#### Acknowledgements

The presented study was conducted as part of the research projects RECORD and SocialL, supported by the Norwegian Research Council VERDIKT program and the NordForsk LILAN program respectively. The study forms part of a doctoral thesis to be submitted to the Department of Psychology, University of Oslo.

#### Author details

<sup>1</sup>SINTEF, Pb124 Blindern, 0314 Oslo, Norway. <sup>2</sup>Department of Computing, University of Copenhagen, Njalsgade 128, 2300 Copenhagen, Denmark. <sup>3</sup>Department of Psychology, University of Oslo, Forskningsveien 3B, 0373 Oslo, Norway.

Received: 13 September 2012 Accepted: 20 September 2013

Published: 23 September 2013

#### References

1. Maguire M (2001) Methods to support human-centred design. *Int J Hum Comput Stud* 55:587–634
2. Hammtree M, Weiler P, Nayak N (1994) Remote usability testing. *Interactions* 1:21–25
3. Thomset-Scott BC (2006) Web site usability with remote users. *J Libr Adm* 45:517–547
4. Ozok AA (2008) Survey design and implementation in HCI. In: Sears A, Jacko JA (eds) *The human-computer interaction handbook*, 2nd edn. Lawrence Erlbaum Associates, New York, pp 1151–1170
5. Bruun A, Stage J (2012) The effect of task assignments and instruction types on remote asynchronous usability testing. In: Konstan JA, Chi EH, Höök K (eds) *Proceedings of the SIGCHI conference on human factors in computing systems, CHI '12*. ACM Press, New York, pp 2117–2126
6. Holmström H, Henfridsson O (2006) Improving package software through online community knowledge. *Scand J Inf Syst* 18:3–36
7. Hess J, Offenberg S, Pipek V (2008) Community driven development as participation? Involving user communities in a software design process. In: Simonsen J, Robertson T (eds) *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*. ACM Press, New York, pp 31–40
8. Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. *Bus Horiz* 53:59–68
9. Notable <http://notableapp.com>
10. ConceptShare <http://conceptshare.com>
11. Law ELC, Roto V, Hassenzahl M, Vermeeren APOS, Kort J (2009) Understanding, scoping and defining user experience: a survey approach. In: Olsen DR Jr, Arthur RB, Hinckley K, Morris MR, Hudson S, Greenberg S (eds) *Proceedings of the SIGCHI conference on human factors in computing systems, CHI '09*. ACM Press, New York, pp 719–728
12. Følstad A (2007) Work-domain experts as evaluators: usability inspection of domain-specific work-support systems. *Int J Hum Comput Interact* 22:217–245
13. Følstad A, Hornbæk K (2010) Work-domain knowledge in usability evaluation: experiences with cooperative usability testing. *J Syst Softw* 83:2019–2030
14. Cowley JA, Radford-Davenport J (2011) Qualitative data differences between a focus group and online forum hosting a usability design review: a case study. In: *Proceedings of the Human Factors and*

- Ergonomics Society 55th annual meeting, HFES 2011. Human Factors and Ergonomics Society, Santa Monica, pp 1356–1360
15. Vredenburg K, Mao J, Smith PW, Carey T (2002) A survey of user-centered design practice. In: Wixon D (ed) Proceedings of CHI '02. ACM Press, New York, pp 471–478
  16. Bark I, Følstad A, Gulliksen J (2005) Use and usefulness of HCI methods: results from an exploratory survey among Nordic HCI practitioners. In: McEwan T, Gulliksen J, Benyon D (eds) People and computers XIX - the bigger picture: Proceedings of HCI 2005. Springer-Verlag, London, pp 200–217
  17. Rubin J, Chisnell D (2008) Handbook of usability testing: how to plan, design and conduct effective tests, 2nd edn. Wiley Publishing, Indianapolis
  18. 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
  19. Hertzum M, Andersen HHK, Andersen V, Hansen CB (2002) Trust in information sources: Seeking information from people, documents, and virtual agents. *Interact Comput* 14:575–599
  20. Tullis T, Albert B (2008) Measuring the user experience. Morgan Kaufmann, Boston
  21. Frøkjær E, Hornbæk K (2005) Cooperative usability testing: complementing usability tests with user-supported interpretation sessions. In: van der Veer G, Gale C (eds) Proceedings of CHI '05. Extended Abstracts on Human Factors in Computing Systems. ACM Press, New York, pp 1383–1386
  22. Bias RG (1994) The pluralistic usability walkthrough: coordinated empathies. In: Nielsen J, Mack R (eds) Usability inspection methods. Wiley, New York, pp 63–76
  23. Følstad A (2007) Group-based expert walkthrough. In: Scapin D, Law E (eds) R3UEMs: Review, Report and Refine Usability Evaluation Methods. Proceedings of the 3. COST294-MAUSE international workshop. R3UEMs, Athens
  24. Følstad A, Law ELC, Hornbæk K (2012) Analysis in practical usability evaluation: a survey study. In: Konstan JA, Chi EH, Höök K (eds) Proceedings of the SIGCHI conference on human factors in computing systems, CHI '12. ACM Press, New York, pp 2127–2136
  25. Hagen P, Robertson T (2010) Social technologies: Challenges and opportunities for participation. In: Robertson T, Bødker K, Bratteteig T, Loi D (eds) Proceedings of the biennial participatory design conference, PDC '10. ACM Press, New York, pp 31–40
  26. 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* 13:183–190
  27. 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) Proceedings of the SIGCHI conference on human factors in computing systems, CHI '09. ACM Press, New York, pp 1619–1628
  28. Shadish WR, Cook TD, Campbell DT (2002) Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin, Boston
  29. Notebox <http://noteboxapp.com>
  30. Cage <http://cageapp.com>
  31. Owela <http://owela.fi>
  32. RECORD online Living Lab <http://livinglab.origo.no>
  33. Hartson HR, Andre TS, Williges RC (2001) Criteria for evaluating usability evaluation methods. *Int J Hum Comput Interact* 13:373–410
  34. John BE, Marks SJ (1997) Tracking the effectiveness of usability evaluation methods. *Behav Inf Technol* 16:188–202
  35. Følstad A, Knutsen J (2010) Online user feedback in early phases of the design process: lessons learnt from four design cases. *Advances in Human-Computer Interaction*. doi:10.1155/2010/956918
  36. Agichtein E, Castillo C, Donato D, Gionis A, Mishne G (2008) Finding high-quality content in social media. In: Najorc M, Broder A, Chacrabarti S (eds) Proceedings of the 2008 international conference on web search and data mining. ACM Press, New York, pp 183–194
  37. Lampe C, Johnston E (2005) Follow the (Slash) dot: effects of feedback on new members in an online community. In: Pandergast M, Schmidt K, Mark G, Ackerman M (eds) Proceedings of GROUP '05. ACM Press, New York, pp 11–20
  38. Brzozowski MJ, Sandholm T, Hogg T (2009) Effects of feedback and peer pressure on contributions to enterprise social media. In: Teasley S, Havn E, Prinz W, Lutters W (eds) Proceedings of GROUP '09. ACM Press, New York, pp 61–70
  39. Lüders M, Følstad A, Waldal E (2013) Expectations and experiences with MyLabourParty: From right to know to right to participate? *J Comput Mediat Commun* (in press)
  40. Burke M, Marlow C, Lento T (2009) Feed me: motivating newcomer contribution in social network sites. In: Olsen DR Jr, Arthur RB, Hinckley K, Morris MR, Hudson S, Greenberg S (eds) Proceedings of the SIGCHI conference on human factors in computing systems, CHI '09. ACM Press, New York, pp 945–954
  41. Kienle A, Ritterskamp C (2007) Facilitating asynchronous discussions in learning communities: the impact of moderation strategies. *Behav Inf Technol* 26:73–80
  42. Stomer-Galley J, Wichowski A (2011) Political discussion online. In: Consalvo M, Ess C (eds) The handbook of internet studies. Wiley-Blackwell, Chichester
  43. Ashford SJ, Cummings LL (1983) Feedback as an individual resource: personal strategies of creating information. *Organ Behav Hum Perform* 32:370–398
  44. Dennis AR, Valacich JS (1993) Computer brainstorming: more heads are better than one. *J Appl Psychol* 78:531–537
  45. DeRosa DM, Smith CL, Hantula DA (2007) The medium matters: mining the long-promised merit of group interaction in creative idea generation tasks in a meta-analysis of the electronic group brainstorming literature. *Comput Human Behav* 23:1549–1581
  46. Karau SJ, Williams KD (1993) Social loafing: a meta-analytic review and theoretical integration. *J Pers Soc Psychol* 65:681–706

47. Jung JH, Schneider C, Valacich JS (2005) The Influence of real-time identifiability and evaluability performance feedback on group electronic brainstorming performance. In: Proceedings of the Hawaii international conference on system sciences, HICSS '05. IEEE, Los Alamitos. doi:10.1109/HICSS.2005.566
48. Dennis AR, Pinsonneault A, Hilmer KM, Barki H, Galupe B, Huber MW, Bellavance F (2005) Patterns in electronic brainstorming. *International Journal of eCollaboration* 1:38–57
49. Amabile TM (1996) *Creativity in context*. Westview Press, Boulder
50. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
51. Field A (2009) *Discovering statistics using SPSS*. Sage, London
52. Cohen J (1992) A power primer. *Psychol Bull* 112:155–159
53. G\*Power 3 <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3>

doi:10.1186/2192-1962-3-18

**Cite this article as:** Følstad et al.: Social design feedback: evaluations with users in online ad-hoc groups. *Human-centric Computing and Information Sciences* 2013 **3**:18.

**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)

---