Facilitating Redesign with Design Cards: Experiences with Novice Designers

Nis Bornoe
Aalborg University
Aalborg, Denmark
nis@cs.aau.dk

Anders Bruun
Aalborg University
Aalborg, Denmark
bruun@cs.aau.dk

Jan Stage
Aalborg University
Aalborg, Denmark
jans@cs.aau.dk

ABSTRACT
While effort has been put into developing and evaluating usability evaluation methods less attention has been paid to shifting usability feedback into improved designs. We report from a study with 44 novice designers creating redesign suggestions. Some were provided with domain specific design cards to facilitate the redesign process. Design cards are physical cards used to structure a collaborative process, and providing design cues such as keywords and questions. Afterward, three developers assessed the quality of the suggestions. We found that the cards diversified the range of system aspects that novices considered, supported ideation, and kept the discussion going. However, the cards did not compensate the limited design experience, and the participants had challenges understanding the value of the cards, and implement them in the process. Having developers assessing the subjective quality of the suggestions turned out to be challenging due to low inter-rater reliability.

Author Keywords
Usability, evaluation, redesign, design cards

ACM Classification Keywords
D.2.2 Design Tools and Techniques

INTRODUCTION
The output from empirical and analytical usability evaluations reveal insights about how interactive systems are used and perceived with the goal of improving the quality of the interaction design. A lot of effort has been put into developing and evaluating usability evaluation methods (Nørgaard & Hornbæk 2009). In contrast, it has been argued that too little attention has been paid to shifting the insights from evaluations into improved designs (Wixon 2003; Hornbæk 2010). The essence of the output is to support the downstream utility (Law 2006; Hornbæk & Frøkjær 2008) defined as *"…the extent to which the improved or deteriorated usability of a system can directly be attributed to fixes that are induced by the results of usability evaluations performed on the system"* (Law 2006). Feedback also serve as a foundation for making informed design choices and influencing future designs (Hornbæk 2010; Lewis 2014; Smith & Dunckley 2002). Although it is a resource demanding process, studies suggest accompanying usability problem lists with redesign proposals (Nørgaard & Hornbæk 2009; Hornbæk & Frøkjær 2005; Sawyer et al. 1996). This has received positive feedback as proposals are more concrete, information rich (Nørgaard & Hornbæk 2009), and can seed ideas for problem fixing (Hornbæk & Frøkjær 2005). While studies have investigated improvements of the evaluation and feedback process, the redesign process itself is still relatively unexplored (Lottridge & Mackay 2009; Smith & Dunckley 2002). Usability evaluations are suitable for locating problems, but not for gathering redesign proposals (Tohidi et al. 2006), so methods are needed to utilize usability feedback (Smith & Dunckley 2002). One example is structured redesign workshops (Bornoe et al. 2014; Bruun et al. 2014; Garnik et al. 2014), and using theories and methods used to facilitate creativity is also suggested (Hornbæk 2008). One approach receiving attention for facilitating creativity is different types of design cards. They come in many forms and shapes and are developed for a variety of different purposes (Wölfel & Merritt 2013). For example, cards are developed for creating new design concepts (Halskov & Dalsgård 2006) including exertion game designs (Mueller et al. 2014), tangible designs (Hornecker 2010), and "playful experiences" (Lucero & Arrasvuori 2010). A common denominator is that they can facilitate creativity by acting as a source of inspiration (Kwiatkowska et al. 2014) as they can be used to present keywords, pictures and questions (Hornecker 2010), and the tangibility can support integration with objects such as notes and sketches (Buur & Soendergaard 2000). When conducting and analyzing usability evaluations it is, not surprisingly, reported that novice evaluators do not have the same skills as expert evaluators. Especially regarding problem identification (Skov & Stage 2012). To support novice evaluators usability engineering tools have been proposed. This has been found to improve reliability and quality of problem identification (Howarth et al. 2009). Based on this logic we hypothesize that facilitating creativity with design cards can support novice designers creating redesign suggestions. During three workshops 44 novice designers participated in creating redesign suggestions of a webshop. Afterward, the developers of the webshop assessed the quality. We address the questions: 1) Are design cards practicable for facilitating the creation of redesign suggestions based on usability feedback? 2) What is novice designers perceived usefulness of the cards? 3) How can we assess the quality of the redesign suggestions?

RELATED WORK

In the following sections, we review literature related to redesign, and design cards.

Redesign

Iterative design is encouraged in Human-Computer Interaction (HCI), but the act of redesigning and the resulting impact is not well explored (Uldall-Espersen et al. 2008) since redesigning not is differentiated from designing, and has received less attention in HCI literature (Lottridge & Mackay 2009). It is argued that redesigning a system, in development or production, can be differentiated from the initial system design process because more balance and compromises are required (Smith & Dunckley 2002). Improving interaction design is a continuous refinement process based on multiple sources of inspiration and, an ongoing work process (Uldall-Espersen et al. 2008) including relevant knowledge from stakeholders (Garnik et al. 2014).

Using feedback from usability evaluations to analyze an existing design and systematically creating redesigns in not a straightforward process (Lottridge & Mackay 2009; Smith & Dunckley 2002). In a study, only 17% of the redesign proposals were found to be highly useful and usable (Molich et al. 2007). Several challenges have been found when suggesting redesigns to fix usability problems. This includes translating identified usability problems into specific design improvements, identify trade-off in design options, identify the mapping from an observed problem to the underlying causes within a design (Smith & Dunckley 2002), analyzing and understanding the complexity of problems (Uldall-Espersen et al. 2008), having the needed domain knowledge (Chilana et al. 2010), systematically conducting a redesign process (Bornoe et al. 2014), and let usability feedback inspire radical design changes (Bruun et al. 2014; Uldall-Espersen et al. 2008).

To support this process, different systematic approaches about creating redesigns have been studied and proposed (Bornoe et al. 2014; Bruun et al. 2014; Lottridge & Mackay 2009; Garnik et al. 2014). Lottridge and Mackay have proposed *"Generative walkthroughs"* to facilitate structured creativity by exploring an existing design through a scenario or storyboard, analyze the design, and finally generate new design ideas based on a set of socio-technical principles. They found that designers with no socio-technical skills were able to learn basic principles and apply them in a redesign process (Lottridge & Mackay 2009). Several approaches have been suggested to provide developers with inspiration for design changes. One simple approach is to have developers observe the usability evaluation. The understanding gained about the users and work processes was found to inspire future system development (Høegh et al. 2006). Facilitated redesign workshops involving developers have also been proposed (Bornoe et al. 2014; Bruun et al. 2014; Garnik et al. 2014). In *"creative sprints"* a multi-disciplinary team conduct evaluation and redesign activities with a focus on one group of usability problem in each 'sprint' (Garnik et al. 2014). Bruun and colleagues conducted a collaborative redesign workshop including developers

and usability specialists. They suggest that a collaborative workshop can unite design and domain knowledge and as a group, they can produce several alternative redesign suggestions. While the developers gained deeper insight of the problems, they did not propose radical design changes, but were able to make informed decisions and complete designs proposed by the usability specialists (Bruun et al. 2014). Another redesign workshop study report that by providing developers with basic knowledge of design principles they were able to apply the principles when correcting usability problems and proposing redesigns. While the suggestions were not complete, new visions and ideas for design changes were developed (Bornoe et al. 2014). A study report that redesign proposals as feedback is of higher utility than problem lists and found useful by developers as they gained an improved understanding, inspiration and perspectives on how to solve, especially non-trivial, problems. Of disadvantages, creating redesign proposals requires additional resources, can be too complex to implement, and introduce new problems (Hornbæk & Frøkjær 2005).

Design cards

Design cards have been proposed to facilitate idearation in a wide range of design situations and topics (Wölfel & Merritt 2013). It is difficult to make a precise definition of design cards, but a common denominator is that they are intended to facilitate creativity by acting as a source of inspiration (Kwiatkowska et al. 2014) and design material in a collaborative setting (Halskov & Dalsgård 2006; Wölfel & Merritt 2013). For example, design cards are used to rephrase abstract frameworks into something more operational. Through this transformation the theory can be more tangible and applicable by making cards with keywords, pictures, and questions (Hornecker 2010; Bekker & Antle 2011). One study found that a main advantage of design cards is that they through these cues can provoke new contextual perspectives on problems and thereby extend beyond personal experience when generating ideas (Kwiatkowska et al. 2014). Both the content and physical layout of design cards have been pointed out as having an influence (Buur & Soendergaard 2000; Deng et al. 2014).

In a review of 18 design cards Wölfel and Merritt highlight that design cards have been reported to facilitate a design process, support design dialogues, make the design process visible and less abstract, provide structure in the process, and are easy to use and manipulate as cards are physical tokens. Therefore, they can serve as a common reference among participants, and act as something specific and concrete to engage discussions. They found that the cards could be divided into three different design situations: *'general'* (open-ended inspiration), *'participatory design'* (engage designers and users in the process), and *'context specific/agenda driven'* (focused on a particular context or design agenda). Methodology wise they divided the cards into *'no methodology' 'suggestion for use'*, and with *'specific instructions'* (Wölfel & Merritt 2013).

Regarding the effectiveness and usefulness of design cards, it has been pointed out that it is an ongoing

challenge to measure the quality of design activities involving design cards (Mueller et al. 2014). We found that evaluations of design cards mainly have been conducted by looking into perceived usefulness and idea counting, e.g. (Kwiatkowska et al. 2014; Chung & Liang 2015; Mueller et al. 2014; Hornecker 2010).

METHOD

During three workshops, we investigated the usefulness of design cards as part of a redesign activity. We based the group work on recent redesign (Bornoe et al. 2014; Bruun et al. 2014), and design workshops using cards (Mueller et al. 2014; Lucero & Arrasvuori 2010; Halskov & Dalsgård 2006). Group work is reported to support a redesign process (Smith & Dunckley 2002; Buur & Soendergaard 2000). Our review of design cards showed that most decks are used during workshops with small groups. We chose group collaboration in *real groups* (*"i.e. face-to-face interacting groups"*) as opposed to *nominal groups ("i.e. individuals working on their own and then collating their outputs to form a cumulative output".)* This approach has been found to feed creativity and idea refinement (Warr & O'Neill 2006).

Afterward, we arranged a quality assessment workshop together with three developers of the webshop.

Participants

Redesign: 44 undergraduate students enrolled in an informatics program (programming/design) following a class about designing and evaluating user interfaces participated. As compensation, they received course credit. To avoid established routines, and a body of design and domain knowledge, we decided to include novice designers.

Quality assessment: two programmers/designers, and a project manager (all called developers, D1-3), involved in the development, assessed the quality of the redesign suggestions individually and afterward in plenum. Note that project management, programming and designing is intertwined in this project. We included developers for the quality assessment because essentially they are the decision makers needing and using the suggestions in the development process (Hornbæk & Frøkjær 2005).

System

The system is a webshop of a soccer team offering match tickets and merchandise. Because domain knowledge is required to design for complex domains (Chilana et al. 2010; Uldall-Espersen et al. 2008), this webshop was chosen as all participants have used mainstream webshops in the past. The moderate complexity of the webshop made it possible to get an understanding during a short period.

Materials

Usability problem list: Beforehand a think-aloud evaluation was conducted. The evaluation resulted in a problem list consisting of 36 usability problems. Each problem was described with the location, a short description, severity, and the number of users encountering the problem.

Demo webshop: The groups had access to an online demo webshop and printed screen shots of most screens.

Design cards: We used three different types of cards in the categories *abstract*, *detailed*, and *customized*. The intention was to get a more general understanding of several types of cards. In relation to the quality assessment of the redesigns, we also wanted to investigate if it was possible to find a correlation between the different card types and the quality.

MethodKit (abstract): a commercial deck of cards for web development, developed by the company MethodKit. This deck is aimed at web developers and is consisting of 53 cards. The overall philosophe behind is that web development is a demanding task, and successful projects have several requirements that need to be taken into account. Each card presents a requirement that should be considered and discussed. What makes this deck abstract is the generality of the themes presented. For example, one card has the title: *"Checkout flow"* followed by the description: *"A smooth payment experience."* (See figure 1.) This deck has no strict method attached.
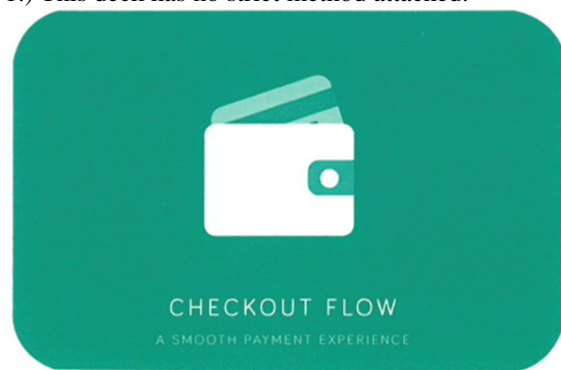

Figure 1: Example of a MethodKit card.

inSights (detailed): a commercial deck of cards for web development, developed by the company Fabrique. This deck consists of 65 cards divided into 11 categories, for example, *"Affordances"*. The cards present different conventions about what is believed to be both visually and operationally good design. What makes these cards detailed is the focus on specific design elements as opposed to the general themes of the MethodKit cards. All cards contain the elements: "[a title]", "Used for", "Principle". "Effect", and "Example" accompanied by a screenshot. For example, one card is "Overlap" with the principle *"Noticeably overlapping an element with another element."* (See figure 2). The inSights deck has no strict method attached, but is intended for reference.
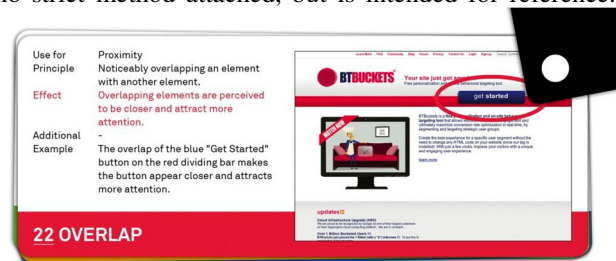

Figure 2: Example of a card in the InSights deck.

Customized cards: This approach was based on the principles of *"Inspiration card workshops"* (Halskov & Dalsgård 2006) and *"The CARD method"* (Tudor et al.

1993). In inspiration cards workshops the idea is to create *'domain cards'*, and *'technology cards'*. Domain cards can represent people, settings, themes, etc. related to the domain in question. For example, a design activity with the theme "Norse mythology" could have a card titled 'blood' because brutality often figures in the tales of Norse mythology. Technology cards can represent a general technology or application, for example, a card can represent a camera tracking movements of an object (Halskov & Dalsgård 2006). Instead of making *'domain cards'*, and *'technology cards'* the groups were asked to make *'problem cards'* and *'flow cards'*. For the *'problem cards'* the groups were asked to divide all problems, from the provided list, into a set of general categories and make a card for each. For the *'flow cards'* they were asked to make cards representing different use flows in the webshop. This idea was based on a method used to analyze high-level task flows (Tudor et al. 1993). The concept was to make it explicit what problem categories users are reporting, and what flows exist. Thereby they could get inspiration towards focusing on fixing a group of common problems and/or changing a flow.

Procedure of redesign workshops
The participants received an introduction to the principles of design cards, an outline for a redesign process, keywords about how to document ideas, and a short introduction before each workshop. We divided the 44 participants into 15 groups and used a between subjects design. Originally we created 18 groups (named G1-18), but three groups left. To get equal groups we used the grades received on an earlier project about analyzing and evaluating an IT system. Each group consisted of two – three participants, one with a high, middle, and low grade. They participated in three workshops, each lasting up to four hours. The participants were allowed to ask questions, but we did not include a facilitator in each group because we wanted to simulate constrains, the fast peace, and the ad-hoc nature related to practical development settings (Nielsen & Madsen 2012).

Workshop 1: During the first workshop, the groups were given the usability problem list and asked to get familiar with both the problem list and the webshop. The purpose was to give the participants a chance to explore and understand the problems, recreate problems, and get a feeling of the flow and functions of the webshop, for example, how to use promotional codes. The groups did not receive design cards during this session.

Workshop 2: During the second workshop, we divided the 15 groups into four different groups clusters each consisting of three or four groups using a between subjects design. Group cluster 1 acted as a control group and did not receive any cards (all in all three groups). Group cluster 2, 3, and 4 used with a set of design cards (all in all 12 groups). Groups in cluster 2 were asked to make a set of customized cards. Groups in cluster 3 were provided the inSights cards, and the groups in cluster 4 were provided the MethodKit cards. Group cluster 3 and 4 were asked to get familiar with and discuss the provided decks by going through each card.

Workshop 3: During the third workshop, all groups were asked to make redesign suggestions of the webshop based on both the problem list and the provided cards (except the control group). They were asked to present the suggestions as sketches and descriptions as this has been suggested as a valuable feedback format (Høegh et al. 2006; Nørgaard & Hornbæk 2009; Hornbæk & Frøkjær 2005; Howarth et al. 2007). The groups were informed that the redesign suggestions would be assessed. A summary of the workshop activities is listed in Table 1.

Workshop	Cluster	Activity
1	All	Review system and problem list.
2	1	- (control group)
	2	Develop customized cards.
	3	Explorer inSights cards.
	4	Explorer MethodKit cards.
3	1	Redesign, no cards.
	2	Redesign, customized cards.
	3	Redesign, inSights cards.
	4	Redesign, MethodKit cards.

Table 1: Summary of workshop activities.

Data collection
After workshop 3 we collected four types of data.

Redesign suggestions: We collected the redesign suggestions from the 15 groups.

Observations: During the workshops, each of the three authors observed a group. We divided the observations so we observed a group in each of the three card categories.

Interviews: After workshop 3 we interviewed three other groups, one interview with a group in each card category.

Survey: All participants filled out a survey consisting of 40 questions mixed of 27 7-point likert scale questions, and 13 short comments. The survey covered: group cooperation (2), using the usability problem list (8), self-assessment of the redesigns (17) (based on an assessment framework (Dean et al. 2006)), and perceived usefulness of the cards (13). For the last part, we included questions from a survey used by Mueller et al. (Mueller et al. 2014).

Data analysis
For the qualitative data we followed the sequence consisting of: compiling, disassembling, reassembling, interpreting and concluding suggested by Yin (Yin 2010). We based the initial codes on the keywords: *"generating ideas"*, *"improving ideas"*, *"communication"*, *"focus"*, and *"consensus"*. The quantitative data was analyzed for significant differences between the different types of cards using a one-way ANOVA test.

Procedure of the quality assessment workshop
Firstly, an introduction to the study, the cards, and the assessment was given. Secondly, the developers were asked to make an individual assessment of each redesign suggestion. Initially, we asked the developers, for each redesign suggestion, to fill out a survey, consisting of 12 questions, on a 7-point likert scale and provide a short written statement. The questions were written according to a theoretical assessment approach developed by Dean et al. They present four quality dimensions to assess

ideas: *novelty*, *workability*, *relevance*, and *specificity*. In addition, each dimension has a number of sub-dimensions (Dean et al. 2006). Thirdly, the developers participated in a plenary session discussing each redesign suggestion and dividing the redesigns into *good*, *medium*, *poor*, and *"out of category."* The last category was added by the developers and contained the redesign suggestions found useless or when a group clearly had misunderstood the task. For example, one group only created a map linking design cards to specific usability problems without providing any suggestions for a redesign.

FINDINGS

In the first section we present the findings from the redesign workshops, in the second section we present the findings from the quality assessment workshop.

Perceived usefulness of cards

When looking at the quantitative data from the survey (n=35, participants provided with design cards) a one-way ANOVA test revealed no significant differences in perceived usefulness between any of the card types (F=2.8, df=2, p>0.1).

Individually, we asked all participants about the atmosphere of the group cooperation to get a sense of the work setting of the workshops. Overwhelmingly, positive feedback was given: *"I saw good cooperation between group members, all provided input and suggestions during the process"* (MethodKit, G1), and *"It was a iterative and innovative process which was very constructive"* (Customized, G2).

Because we found that the process around the customized cards was different from the inSights and MethodKit cards, we decided to split the findings into two. For each section we will first discuss customized cards, then inSights and MethodKit cards, and finally commonalities for all decks.

Generating and improving ideas

The groups with customized cards did not express that they found the cards helpful for generating and improving ideas. This statement put in plain explains well: *"We did not really use it in practice"* (Customized, G4), which could be related to the confusion mentioned above. How to actually use the created cards during workshop 3 for ide-ation also seemed to be confusing, as the creative process had happened during the creation phase. For example, a participant explained that the group had created a card related to the design of a menu. This helped ideation because: *"The card was made specifically for this purpose"* (Customized, G14). This indicates that during workshop 3 the cards served as a reminder, and the ideas generation had happened earlier.

The groups using the inSights and MethodKit cards expressed that the cards triggered aspects not initially considered: *"Certain things had I not investigated if the cards had not inspired me"* (inSights, G13), and the cards encouraged new viewpoints: *"The cards got one to think in a different way"* (MethodKit, G15).

While not providing explicit ideas the groups expressed how the cards helped broaden the spectrum of ideas and

thereby implicit supporting the ideration process. A participant from a group using the inSights cards said: *"[The cards] gave us an idea to take on the task differently and look for things other than what we first thought"* (inSights, G13) One using MethodKit explained: *"We did not use them specifically to design something, but they provided a good mindset […] one might well forget that there are so many elements to take into account"* (MethodKit, G8) More specifically the cards seemed to facilitate the ideration process by providing something explicit as a foundation for brainstorming and discussion. A participant from a group using MethodKit explained: *"[The cards] were good at generating new ideas we would not have been able to come up with if we had brainstormed without"* (MethodKit, G15). However, a participant felt opposite when asked if the cards supported idearation: *"No, they rather restrained my ideas"* (MethodKit, G1).

As mentioned, ideas for redesigns are generated based on several sources. The cards were considered one source among the ones available rather than the main driver. *"Ideas were generated from both the cards, the problem list, and inspiration from other sites"* (inSights, G7). The feedback clearly showed that all groups highly relayed on inspiration from other sources. Particularly they received inspiration from similar webshops of other soccer teams and webshops known and used by the participants.

Focus, communication, and consensus

During workshop 3 the groups with customized cards barely used them during the design process for communication and discussion. Instead, it seems as if they served their main purpose during the creation phase in workshop 2. A participant noted about using the customized cards for communication in the group that: *"That's not really how we used the cards."* (Customized, G4), and: *"We could very well remember the ideas since we made the cards ourselves"* (Customized, G14).

This was moderately different regarding the groups with the inSights and MethodKit cards. A participant explained that it could be useful to have an artifact during the interactions to keep the focus: *"Since we had something physically at hand, it was easy to get the group unified"* (inSights, G7), as an artifact could foster discussion: *"[The cards] provided the opportunity of discussing various cards while having something visually at hand"* (inSights, G7). The cards could help avoiding procrastination since they were something concrete to work with as noted by a group: *"When the group lost focus, we turned a new card to regain the focus"* (MethodKit, G15).

Because of the detailed design details provided by the inSights cards they could to some extent be used to communicate ideas: *"If one of my ideas was supported by a card, I could use it as an example"* (inSights, G7). When it comes to the MethodKit cards, none of the groups expressed that they used the cards to communicating ideas between the group members. The feedback indicates that the abstraction of the cards does not make them suitable for specific communication between group members. For example, a group explained

that they would discuss and try to find consensus about the meaning of certain cards due to the abstraction and generality of this deck. *"It was not always clear what the cards meant. Clarifying this spawned discussion"* (MethodKit, G8). It was also clear that the groups did not use the cards for finding consensus about the suggestions during the process. None of the groups indicated that this had been the case.

Use and value
For the groups creating customized cards, the process turned out to be complex and abstract. Understand what they were supposed to do and how to make the cards was challenging. This comment covers our impression well: *"It was a confusing process"* (Customized, G14). Creating the cards was an activity of its own and the groups spend hours making the cards. For these groups, creating the cards was more a reflective activity. Using the cards during workshop 3 to developed redesign suggestions was not the case. *"Only limited work was based on the cards"* (Customized, G14).

The groups using the inSights and MethodKit cards also reported challenges when figuring out how to include the cards in the process. *"It was unclear how they should be used. [We] used them somewhat randomly"* (MethodKit, G1). The cards added an extra layer of complexity to the process. For example, how the groups should go about using the cards in the first place. *"[I] do not think the cards made things easier, rather more confusing in the beginning..."* (inSights, G12).

During the redesign activity in workshop 3, we observed that the groups mainly used the cards in the beginning. Several groups conducted brainstorming based on the selected subset of cards. Afterward, the cards were put away for the rest of the session. Instead, the focus was about maturing and documenting specific ideas.

In summary, all the groups found it challenging to use the cards and implement them in the process. The missing facilitation, the minimal introduction to the cards, and the fact that the participants are novice designers made the process somewhat unstructured and disorganized as the participants both had to get familiar with the usability problem list, the cards, and being creative.

Seven participants explicit expressed some dissatisfaction with the value of the cards. They were not certain about how to use the cards or what they would get from the cards. *"We did not really use the cards as the principles generally were self-evident"* (inSights, G10). Several claimed that the cards provided *"common sense"*. For example, it was stated about the inSights deck that: *"They are very intuitive and therefore not necessary"* (inSights, G7), and a participated commented about the MethodKit deck that: *"The cards helped a bit, but we had probably reached the same result without"* (MethodKit, G1). Another group member added: *"[The cards] were too much 'of cause'"* (MethodKit, G1).

While most groups did express advantages and positive experiences with the cards, it was also clear that they found it difficult to explain advantage in explicit terms. As a result, positive experiences were often expressed in weak or generic terms. For example, *"Some cards inspired new things to think about"* (MethodKit, G9).

We observed that making redesign suggestions is a tough discipline and cannot be conducted for more than a few hours at a time. Being creative and having to come up with ideas and actually transform these ideas into documentation created fatigue among the participants. After the first couple of hours at workshop 3, we observed impatience and restlessness. To deal with this fatigue, we encouraged the participants to take a few short breaks during the process.

Quality assessment workshop
While assessing the quality of ideas is difficult it is argued that making a quality assessment of ideas inspired by design cards might provide additional insights about the usefulness (Mueller et al. 2014).

Inter-rater reliability of quantitative results
A Fleiss' kappa analysis of the inter-rater reliability of the quantitative survey data resulted in 0.042. Since this is low we instead present the qualitative data given both individually and in plenum by the developers. We suspect that the structure of the analysis process has resulted in what in usability engineering is known as the evaluator effect (Hertzum & Jacobsen 2003).

Communicating ideas
While the developers had received no specific instructions to do so, they independently commented on the presentation for almost all redesign suggestions. Text combined with a graphical representation (screenshots, mockups, etc.) and a clear connection between the two was highly valued. Text by itself is not good at explaining an idea or flow. *"Good description, showing a solution with as a mockup in relation of the specific page"* (D1, G18). This makes the redesigns accessible since it's possible in a short time to get an overview and understand the context of the suggestion.

Inspiration and concepts
During the plenary session, the developers explained that they are looking for inspiration for further development and not necessarily highly detailed suggestions ready for implementation. They preferred depth and conceptual ideas instead of suggestions fixing as many usability problems as possible. The redesign suggestions will not be implemented directly, but will act as a source of inspiration when specific plans are made for design changes. For example, a group received the comment: *"Very fine that you have selected a few problems and has worked with them"* (D3, G18). Ideas are interesting if they are based in real problems and can influence and mature a future design. The comment *"Very innovative and well graphically illustrated. Good solutions that creates a personal experience"* (D2, G8) was given to a group presenting new idea about the user login flow and presentation of a search field and shopping basket. This suggestion both dealt with reported problems, and provided specific ideas related to the flow and presentation of information to the users. Specifically the group had targeted new users of the webshop, this focus was appreciated: *"Good description and good thoughts about the experience for new users"* (D1, G8). The idea

was something the developer had not considered. While it is uncertain if the idea actually will be implemented in some form or another, the idea was still found useful as it seeded new inspiration.

The developers wanted focus on design aspects such as use flows and arrangement of elements, and what to include in different screens, not specific content or visual aesthetics. A few groups had made suggestions that mainly focused on content. For example, one group had suggested that a news section was added. Another group had suggested labels for a menu. Rather the developers were interested in how to present a menu and less on the labels. Content is often up to the client. Potentially content suggestions could be valuable with other projects. Example of feedback given to a group focusing on content: *"…focused on category structure that we do not manage - it's the customers who decide (i.e. photos, texts, etc.) This solution is not workable, but might be fine to pass on to the customer"* (D3, G10).

Technicalities such as browser support were not considered useful. This is about the specific implementation and therefore not relevant when it comes to the design and form of the system, which is what the developers wanted. Here it would have been an advantage if the groups had received instructions from the developers beforehand.

Group assessment of the redesign suggestions
In the final part, the developers as a group were asked to index the redesign suggestions according to the four categories 'good', 'medium', 'poor', and "out of category".

'Good': Three redesign suggestions received the labels: focused and understandable, clear, descriptive, and show courage to correct complex problems. *"…clear description, with good suggestions for improvement!"* (D1, G17). Interestingly it turned out that the redesign suggestions were made by, a control group, a group using customized cards, and a group using the MethodKit deck (i.e., there was no obvious connection between card type and the quality of the redesign suggestion.) This mix of conditions was true for all the classification categories.

'Medium': Three redesign suggestions received the labels: decent redesign suggestions, not optimal design solutions, imperfect but innovative suggestions, suggestions that not necessarily create value. *"The handling of discount codes is probably not the optimal solution"* (D3, G1).

'Poor': Four redesign suggestions received the labels: looks more like a requirements specification, confusing, missing information, not optimal design solutions. *"Too much focus on problems which are 'self-invented', and not those from the list of problems"* (D2, G4).

"Out of category": Five redesign suggestions received the labels: no pictures, unspecific, poor design solutions, ambitions rather than concrete redesign suggestions, purpose of the assignment misunderstood. *"Adding the two buttons 'Do you have a ticket code?' and 'Do you have a discount coupon?' will be even more confusion. How do I know what I have?"* (D3, G7).

A few redesign suggestions received conflicting comments in the individual assessment. This made the collaborative assessment challenging. For example, G2 presented their redesign by adding comments as text fields directly on a set of screenshots. This yielded two opposite comments: *"Very inconsistent and confusing…"* (D1, G2), and: *"A nice clear format to present suggestions"* (D2, G2). This illustrates, along with the low inter-rater reliability, how difficult it is to make a quality assessment of something as abstract as a redesign suggestion based on each developer's preference.

DISCUSSION
We discuss the findings in relation to related literature.

Perceived usefulness of cards
Overall we received neutral feedback regarding the perceived usefulness. On the positive side the cards diversified the range of system aspects that novices considered and kept discussion/work going, one group even asked if they could keep the MethodKit cards. While the participants had trouble explaining exactly how, it does seem as the MethodKit, and InSights cards provoked participants to consider aspects other than personal knowledge and preferences: *"[The cards] were good at generating new ideas we would not have been able to come up with if we had brainstormed without"* (MethodKit, G15). Similarly, a study reports that design cards supported structure during idea generation and steered the group work. It's highlighted that the most important aspect of using cards was that the cards could introduce new contexts to a given problem, thereby offering the designers to expand on personal experiences (Kwiatkowska et al. 2014). Some groups report that the cards helped to keep a focus and the process going when a "vacuum" in the discussion occurred.

Of problems, especially understanding the value and how to use them in the process was challenging. Understanding and interpreting design cards have been indicated as somewhat problematic in several other studies (Mueller et al. 2014; Kwiatkowska et al. 2014). A study reports that when the frequency of use increases so does the understanding of the cards (Chung & Liang 2015). Our findings indicate that using design cards, like other design techniques, require experience to understand and judge the potential, benefits, and disadvantages of specific decks.

Since the groups found it challenging to use the cards and implement them in the process, it is tempting to suggest some form of directions for use. Studies report that some participants requested more guidance (Mueller et al. 2014; Bekker & Antle 2011), and Hornecker concludes that guidance is important for the flow of a workshop based on design cards (Hornecker 2010). One study providing guidance report mixed responses about the usefulness (Lucero & Arrasvuori 2010). Especially regarding the creation and use of the customized cards approach, guidance beyond what was provided, or even facilitation seems essential. In comparison to the premade

MethodKit and InSights cards, initially creating cards is a much more demanding and reflective process. From a workshop, it is reported that the creation of customized cards was done in collaboration between stakeholders and experienced designers (Halskov & Dalsgård 2006).

Quality assessment workshop

Several studies evaluating the effectiveness and usefulness of design cards have counted ideas, e.g. (Kwiatkowska et al. 2014; Chung & Liang 2015). We do not believe this will be a good approach in our case. Counting ideas will be just as counting problems when comparing usability evaluation methods to decide which one is the best. Such numbers do not by themselves support problem fixing (Hornbæk 2010; Wixon 2003). As mentioned by the developers the number of suggestions was not essential. Arguably, redesign emphasizes focus on detailed design and improving the existing design based on feedback and less on a number of suggestions. We found it relevant to evaluate the design cards according to the output. Here we looked at related studies of redesigns. Hornbæk and Frøkjær had developers assessing the usefulness of redesigns as development input according to the four criteria: problem severity, problem frequency, problem persistence, and usefulness (Hornbæk & Frøkjær 2005). One study rated redesigns according to the four criteria: number of users affected, problem occurrence, business impact, and user impact (Uldall-Espersen et al. 2008), and another study rated according to the self-developed 5-point rating scales 'usefulness' and 'usability' (Molich et al. 2007). We decided to go a different direction, and rate according to idea assessment rather than according to usability values. To get comparable assessments and a recognized theoretical foundation, we used an idea assessment framework (Dean et al. 2006). Due to low inter-rater reliability, it was difficult to conduct a firm quantitative quality rating. Several redesign suggestions received opposite opinions from the developers making the assessment process challenging. While we aimed at conducting systematic assessments of the redesign suggestions, this is still very subjective by nature since the three developers have different roles and personal preferences. Finding a firmer common ground and assessment process beforehand by reducing the evaluator effect (Hertzum & Jacobsen 2003) might improve the inter-rater reliability and make it clear what is requested of the designers and what defines good quality.

Providing the designers with directions for what the developers' need and request might be useful. Hornbæk and Frøkjær found business goals of a system to be a useful resource for usability evaluators since these goals can support locating usability problems with higher business relevance and overall utility (Hornbæk & Frøkjær 2008). In relation, we speculate that setting up some redesign goals can improve the utility of the generated suggestions. We found that ideas for content, visual aesthetics, and technical aspects were of no or less interest in this particular case. Instead, ideas for flows and concept changes were wanted. Providing designers with goals about what the developer of the system is requesting could support the process of making redesign suggestions with higher utility and direct the designers.

CONCLUSION

In this paper, we have reported from a study in which 44 novice designers participated in creating redesign suggestions of a webshop. We provided 35 participants with design cards, as a source of inspiration, to facilitate the process. Afterward, three developers of the webshop assessed the quality of the redesign suggestions.

Overall we found no significant differences in perceived usefulness of the different types of cards used. There was no obvious connection between card type and the quality of the redesign suggestion. Even a control group with no cards provided one of the best redesign suggestions. The findings from our study indicate that the design cards worked for generating ideas, keeping focus, and pushing the discussion and flow during the idearation phase. The cards showed to be less helpful for supporting communication and consensus in a group. The cards did not compensate the limited design experience Especially, understanding the value of the cards and how they could be implemented in the process was found to be an explicit challenge.

Getting the redesign suggestions assessed turned out to be difficult due to the subjective nature of what good and bad design is. Here we believe a more strict assessment process is required to avoid the evaluator effect and to increase the inter-rater reliability. One essential aspect is to define clearly what good quality is, and have the raters agree about the specific focus of a given assessment process. This could support both the designers and the raters. Future work could further explorer opportunities for conducting systematic quality assessments of designs based on design cards.

A limitation of this study is the chosen cards and procedure of the workshops. It can be argued that other design cards potentially could lead to different results. To partly avoid this we have worked with three substantially different types of cards. As discussed, the quality assessment of the redesigns has some challenges and could have been conducted in a number of different ways.

REFERENCES

Bekker, T. and Antle, A. N. Developmentally Situated Design (DSD): Making Theoretical Knowledge Accessible to Designers of Children's Technology. *Proc. CHI '11*. ACM (2011), 2531–2540.

Bornoe, N., Billestrup, J., Andersen, J. L., Stage, J., and Bruun, A. Redesign Workshop: Involving Software Developers Actively in Usability Engineering. *Proc. NordiCHI '14*. ACM (2014), 1113–1118.

Bruun, A., Jensen J. J., Skov, M., and Stage, J. Active Collaborative Learning: Supporting Software

Developers in Creating Redesign Proposals. In S. Sauer et al., eds. *HCSE 2014*. Lecture Notes in Computer Science. Springer, Heidelberg, (2014), 1–18.

Buur, J. and Soendergaard, A. Video Card Game: An Augmented Environment for User Centred Design Discussions. *Proc. DARE '00*. ACM (2000), 63–69.

Chilana, P. K., Wobbrock, J. O., and Ko, A. J. Understanding Usability Practices in Complex Domains. *Proc. CHI '10*. ACM (2010), 2337–2346.

Chung, D. and Liang, R.-H. Understanding the Usefulness of Ideation Tools with the Grounding Lenses. *Proc. Chinese CHI '15*. ACM (2015), 13–22.

Dean, D. L. Hender, J. M., Rodgers, T. L., and Santanen, E. L. Identifying quality, novel, and creative Ideas: Constructs and scales for idea evaluation. *Journal of the Association for Information Systems*, 7, 1 (2006), Article 30.

Deng, Y., Antle, A. N., and Neustaedter, C. Tango Cards: A Card-based Design Tool for Informing the Design of Tangible Learning Games. *Proc. DIS '14*. ACM (2014), 695–704.

Garnik, I., Sikorski, M., and Cockton, G. Creative Sprints: An Unplanned Broad Agile Evaluation and Redesign Process. *Proc. NordiCHI '14*. ACM (2014), 1125–1130.

Halskov, K. and Dalsgård, P. Inspiration Card Workshops. *Proc. DIS '06*. ACM (2006), 2–11.

Hertzum, M. and Jacobsen, N.E. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15, 1 (2003), 183–204.

Høegh, R. T., Nielsen, C. M., Overgaard, M., Pedersen, M. B., and Stage, J. The Impact of Usability Reports and User Test Observations on Developers' Understanding of Usability Data: An Exploratory Study. *International Journal of Human-Computer Interaction*, 21, 2 (2006), 173–196.

Hornbæk, K. Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology*, 29, 1 (2010), 97–111.

Hornbæk, K. Usability Evaluation as Idea Generation. In E.-C. Law, E. Hvannberg, and G. Cockton, eds. *Maturing Usability: Quality in Software, Interaction and Value*. Human-Computer Interaction Series. Springer London (2008), 267–286.

Hornbæk, K. and Frøkjær, E. Comparing usability problems and redesign proposals as input to practical systems development. *Proc. CHI '05*. ACM (2005), 391–400.

Hornbæk, K. and Frøkjær, E. Making use of business goals in usability evaluation: an experiment with novice evaluators. *Proc. CHI '08*. ACM (2008), 903–912.

Hornecker, E. Creative Idea Exploration Within the Structure of a Guiding Framework: The Card Brainstorming Game. *Proc. TEI '10*. ACM (2010), 101–108.

Howarth, J., Andre, T. S., and Hartson, R. A structured process for transforming usability data into usability information. *Journal of usability studies*, 3, 1 (2007), 7–23.

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

Kwiatkowska, J., Szóstek, A., and Lamas, D. (Un)Structured Sources of Inspiration: Comparing the Effects of Game-like Cards and Design Cards on Creativity in Co-design Process. *Proc. PDC '14*. ACM (2014), 31–39.

Law, E. L.-C. Evaluating the downstream utility of user tests and examining the developer effect: A case study. *International Journal of Human-Computer Interaction*, 21, 2 (2006), 147–172.

Lewis, J. R. Usability: Lessons Learned … and Yet to Be Learned. *International Journal of Human-Computer Interaction*, 30, 9 (2014), 663–684.

Lottridge, D. and Mackay, W. E. Generative Walkthroughs: To Support Creative Redesign. *Proc. C&C '09*. ACM (2009), 175–184.

Lucero, A. and Arrasvuori, J. PLEX Cards: A Source of Inspiration when Designing for Playfulness. *Proc. Fun and Games '10*. ACM (2010), 28–37.

Molich, R., Jeffries, R., and Dumas, J. S. Making usability recommendations useful and usable. *Journal of Usability Studies*, 2, 4 (2007), 162–179.

Mueller, F., Gibbs, M. R., Vetere, F., and Edge, D. Supporting the Creative Game Design Process with Exertion Cards. *Proc. CHI '14*. ACM (2014), 2211–2220.

Nielsen, L. and Madsen, S. The Usability Expert's Fear of Agility: An Empirical Study of Global Trends and Emerging Practices. *Proc. NordiCHI '12*. ACM (2012), 261–264.

Nørgaard, M. and Hornbæk, K. Exploring the Value of Usability Feedback Formats. *International Journal of Human-Computer Interaction*, 25, 1 (2009), 49–74.

Sawyer, P., Flanders, A., and Wixon, D. Making a Difference - the Impact of Inspections. *Proc. CHI '96*. ACM (1996), 376–382.

Skov, M. B. and Stage, J. Training software developers and designers to conduct usability evaluations. *Behaviour & Information Technology*, 31, 4 (2012), 425–435.

Smith, A. and Dunckley, L. Prototype evaluation and redesign: structuring the design space through contextual techniques. *Interacting with Computers*, 14, 6 (2002), 821–843.

Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. Getting the Right Design and the Design Right. *Proc. CHI '06*. ACM (2006), 1243–1252.

Tudor, L. G., Muller, M. J., Dayton, T., and Root R. W. A Participatory Design Technique for High-Level Task Analysis, Critique, and Redesign: The CARD Method.

Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 37, 4 (1993), 295–299.

Uldall-Espersen, T., Frøkjær, E., and Hornbæk, K. Tracing impact in a usability improvement process. *Interacting with Computers*, 20, 1 (2008), 48–63.

Warr, A. and O'Neill, E. The effect of group composition on divergent thinking in an interaction design activity. *Proc. DIS '06*. ACM (2006), 122.

Wixon, D. Evaluating Usability Methods: Why the Current Literature Fails the Practitioner. *interactions*, 10, 4 (2003), 28–34.

Wölfel, C. and Merritt, T. Method Card Design Dimensions: A Survey of Card-Based Design Tools. In P. Kotzé et al., eds. *Human-Computer Interaction – INTERACT 2013 SE - 34*. Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013), 479–486.

Yin, R.K. *Qualitative research from start to finish*, Guilford Press (2010).