5-15-2019

SCOPE FOR USABILITY TESTS IN IS DEVELOPMENT

Kimmo Tarkkanen
Turku University of Applied Sciences, kimmo.tarkkanen@gmail.com

Ville Harkke
University of Turku, ville.harkke@utu.fi

SCOPE FOR USABILITY TESTS IN IS DEVELOPMENT

Research paper

Tarkkanen, Kimmo, Turku University of Applied Sciences, Turku, Finland, kimmo.tarkkanen@turkuamk.fi

Harkke, Ville, University of Turku, Turku, Finland, ville.harkke@utu.fi

Abstract

Despite being a common, established concept in wide usage, usability tests can vary greatly in goals, techniques and results. A usability test purchased and performed for a specific software product, may result in either minor user interface improvements or radical U-turns in the development. Such variation has been discussed as a problem of the scientific reliability and validity of the testing method. In practice it is more important what 'kind of data' one can expect of the selected method than whether it is reliably always the same data. This expectation of information content or 'scope' is of importance for evaluators, who select and conduct usability tests for a specific purpose. However, the scope is not explicitly stated or even discussed: Too often the premise is that, because a usability test involves users, it brings the (necessary) user-centeredness to the design i.e. takes socio-technical fundamentals as inherently given. Through a literature review of testing practices and analytical considerations, we search for the scope of a usability test, which could deliberately approach the socio-technical tradition and equally develop both the system and the user organization. A case example represents a possible realization of the extended scope of usability test.

Keywords: Usability testing, Scope, IS evaluation and development, Socio-technical approach.

1 Introduction

Technology users in a professional work setting still run into situations where information systems are inappropriate and unusable for their work tasks at hand. Each poor use experience and usability problem of the system results in a loss of something: Our personal productivity and efficiency at work decreases, the software company misses a potential user and a paying customer, or we may endanger the patient safety in health care settings. Each poor user experience and usability problem at personal, community and organizational level is in need of professional usability evaluation and technology redesign. Apparently, when a problem manifests itself, the user research, UX and usability efforts have failed, or did not even take place before the software product entered the market or got implemented into an organization and finally evoked such poor experiences in users.

Information system researchers, developers and software designers know that efforts to design and implement new technology for a professional and complex work domain will be most successful when built on a firm knowledge about how users actually accomplish work in their everyday practices (Suchman, 1985). This ideology of user-centered development (UCD) is widely supported in the software industry. However, the actual implementation of UCD into everyday development processes is (a) difficult and laborious, (b) depends on developer's personal attitudes and (c) scarce organizational resources, which impair the breadth and depth of user focus in the system development (e.g. Steen, 2008; Bødker, 2006; Iivari, 2006).

The poor usability of professional information systems arises equally from the usability methods and the evaluation activity itself. Usability evaluation methods lack design relevance and persuasiveness among IT developers and managers (e.g. Rajanen et al., 2011) while having persistent reliability and validity problems themselves (Hertzum et al., 2014). It is not always clear how the results of usability evaluations should be interpreted (Hornbæk, 2008), injected back to development process (Bernhaupt et al., 2016) and to what extent these results are reliable and generalizable to other contexts, users and products (Reijonen and Tarkkanen, 2015). Usability problems found with usability evaluation methods may only confirm earlier impressions of the system developers (Hornbæk and Frøkjær, 2005). Thus, developers neither fix nor react on these (Molich and Dumas, 2008).

Of the individual usability evaluation methods, usability testing is the most popular and widely used among UCD practitioners and probably the best known method among non-professionals. The main characteristics of the usability testing method are: 1) involving prospective users, who carry out tasks with the product, and 2) involving evaluators who observe and record users' behaviour in a short session during which users provide feedback about the product. The most defining characteristic of usability testing is the concrete use of the system based on the test tasks and not the conversation as with interviews (Hertzum, 2016).

Within the usability testing method, the problem is that when put into the context of complex problem domains, the scope and the focus of usability testing are too often traditional and narrow, not aiming at reviewing users' actual work in these contexts (Redish, 2007). To have an impact on design, usability tests will have to mirror the complexity of the problem domain in the planning and aim at revealing issues that bring developers closer to a solution to the wicked problem. In practice, this means questioning all that is known in the design process that far – testing the unknowns – by setting the focus on acquiring user knowledge for the development with a scope that covers not only the design artefact, but also the whole spheres of contexts of use and beyond to the sources of value (see Cockton, 2004, 2006).

In order to apply any method, an evaluator needs to know what type of results one can expect from the method: What problems a method is and is not good for finding (Blandford et al., 2008 p. 283). Such goodness can be evaluated in terms of *the scope of the method* (ibid.). In this paper, we define the scope as *the extent to which a usability testing method uncovers problems in the context of development*[1]. The

[1] Our definition is based on the Oxford Dictionary of English, which describes the scope as 1) the extent of the subject matter that something deals with or to which it is relevant 2) the opportunity or possibility to do or deal with something. We could equally comply with a simpler notion "the scope of 'what is being tested' or 'what the test is really about' " by Reeves (2019).

scope determines whether the method fits or does not fit to the evaluation case and its objectives at hand – whether the evaluator should select a certain method over another. From usability practitioners' perspective, method validity is conveyed as the scope. It is construct validity: What 'kind of data' can one expect from the selected method? And does the method measure usability as it is required and understood in this particular development project? Usability interventions aim to improve the product step by step, usually without a need to replicate, compare or search for similar results among different evaluators. Thus, practical reliability is about the predictable behaviour of the selected method in different evaluation contexts and products, such that designers and developers can understand and trust the data the method produces. The concept and definition of the scope aims to hold such understanding for evaluators – what becomes included and excluded due to methodological choices – and help them select and conduct usability tests for the specific purpose.

However, the scope of usability testing in its various forms is not well articulated and hardly even discussed. The scope is too often viewed as highly abstract and 'black-boxed': Usability tests involve users, who give their (best) contribution and bring the (necessary) user-centeredness to the design and design process. Many software development projects in industry may assume that the institutionalized prescription of the usability testing method acts as a guarantor of the success in the design process (cf. Gray, 2016), and when rigorously followed, will lead to automatic identification of certain types of usability problems and design flaws (cf. Hornbaek, 2010). Thus, socio-technical design fundamentals are taken as inherently given in usability testing due to its institutionalized status, even if the scope of testing is not deliberately put on the social or technical. User organizations who source and outsource evaluation activities cannot rely a well-established understanding of the testing scope in the market. This may mean standardized test procedures and unsuitability in the current development and product requirements, which turns the user-centered design toward discontinued and fragmented direction (e.g. Eshet and Bowman, 2015). In scientific literature, and design science studies in particular, when validating the design usability testing often embodies a very limited scope resulting in e.g. terminology mismatches, structural complexities and redesign recommendations for individual UI elements (as an example see e.g. Guay et al., 2019). In the project management literature, poor scope definition is well recognized leading to project failure and increase in costs and schedule (Cho and Gibson, 2001). In usability research literature, the discussion about the scope of usability testing is dispersed to different analytical arguments for scope change (e.g. works by Cockton 2004; 2006) and different method modifications to apply these changes in different circumstances (see Chapter 3), while the scope itself has remained undefined, unpresented and underrated. Recently for example, Reeves (2019) refers to scope when UX practitioners dissipate the found problem *"by treating it as not in the scope of the usability test"*. Yet, Reeves (2019) does not discuss specifically the scope, although he creditably describes in detail how UX practitioners look for troubles (usability problems) and how they produce findings in usability testing i.e. how they construct the scope of the method through their collaborative actions.

In this paper, we raise our concern in accordance to the theme of this track, which is how to deliberately consider both social and technical aspects of IS design when conducting usability tests. Specifically, we ask, what is the scope of the usability testing and, how and why the scope is extended? The paper begins with setting the stage for the scope change in usability tests; starting from participatory ISD fundamentals, we move towards positioning usability tests within ISD and discussing where the scope of usability test is possible, and many times required, to shift. Then, based again on literature, we take a look at different method modifications, which have broadened the scope of usability testing according to the values of socio-technical approaches. Last, we discuss our empirical usability test case where the equal development of both the system and the user organization can be captured. The case study exemplifies those minor changes in the usability test protocol that can extend the scope of the method to deal with, for example, physical context limitations, complex social relations and more traditional system deficiencies. Findings of the case study act as a one possible representation of the scope of usability testing, which discussion we attempt to stimulate here as well.

2 Searching for the Scope of Usability Tests in IS Development

2.1 Involving users and organizations in ISD

Product design in participation with users has a long tradition and the ideas of participatory design tradition have evolved under different names and concepts throughout the years (e.g. Bjørn-Andersen and Clemmensen, 2017). All these seek to develop richer understanding about the contexts and purposes of users and build them into technology design. A Scandinavian tradition is often referred to in the IS literature when addressing the roots and the first projects of designing information systems in participation with users (e.g. Kensing and Blomberg, 1998). Practices at the workplace have been a core concern of the participatory design community to date (see Kuutti and Bannon, 2014). A legacy from Scandinavian participatory design in user-centered design (UCD) is evident. In the UCD methods, though, the idealistic picture of equal power of stakeholders has diminished to a power of the system developer, who decides what it means that a system is well-designed, thus institutionalizing the participation under the logic of technology development (Holmlid, 2009). In other words, in user-centered design, the user is an information source and a subject instead of an equal partner (Sanders and Stappers, 2008). This artifact-oriented point of view does not seriously consider what sort of activities humans are involved in when they use artifacts (Kuutti, 2011). In the field of design, *co-design* and *co-creation* try to maintain the original idea of equal partnership by changing and mixing the roles of users and designers. Users are working together in the design development process although they are not trained in design (Sanders and Stappers, 2008). The role of users varies from being informants through consultants to equal partners and designers themselves (Kujala, 2003). Kujala (2003) distinguishes ethnography and contextual design as two main approaches in involving users in systems design and development. Ethnographic studies aim to achieve such a shared view on the work and provide insights into the unarticulated aspects of work by applying open-ended (contextual) interviews and participant observations (Kensing and Blomberg, 1998). However, ethnography appears too expensive and too slow in an effective requirements capturing for design purposes that require direct contribution of users to requirements specifications and development decisions (Stewart and Williams, 2005). Moreover, it is impossible to collect, even with ethnographic inquiries, a perfect knowledge base for IT design that addresses all intricacies of use contexts and users' work practices. Ethnography is therefore more a resource to other methods than a primary data gathering method (Stewart and Williams, 2005.) The problem of practice-oriented user research methods, such as contextual inquiry, is that those tend "*to overlook the interaction and knowledge-sharing in user–producer relations"* and are only "eye openers" (Heiskanen et al., 2010).

2.2 About selecting a proper usability evaluation method

Multiple ways exist to involve and collaborate with users during the system development and at the later deployment phase (Johnson et al., 2014). The problem of selecting an appropriate method is emergent and applying a wrong method is waste of money and resources (Hyysalo, 2015). The selected method should fit the particular case, the type of the product designed, the skills of the designers to use the method, the availability of users and knowledge of developers about users and their context. In choosing an appropriate usability testing method, Bødker and Madsen (1998) advise to "*bring test situations closer to the nature of the future situation of use*". According to them, the method choice depends on a number of characteristics of the evaluation situation: (a) what is the purpose of the evaluation, (b) what is known about the context, (c) can the workplace [the intended context of use] accessed, (d) how much resources are available, and (e) what kind of prototypes or other design artifacts are available. The purpose of evaluation may include understanding the current or future practice and context, getting alternative ideas or getting proof of existence for a particular artifact, testing a particular solution and showing which contextual issues are of concern (Bødker and Madsen, 1998). According to Blandford et al. (2008), in the essence of method selection are the costs and benefits of applying any particular usability evaluation method. Costs include time and effort to learn and apply the method, whereas benefits are valued by the insights obtained from using the method (Blandford, 2008). Blandford et al. (2008) use the concept of scope as determining the potential benefits of method usage. The scope is *"what kinds of*

problems a method is and is not good for finding" (ibid. p. 283). The scope should not be confused with the scale and the extent of an individual usability problem, which usually describe its local and global appearance in the system (cf. Dumas and Redish, 1999). The scope is about understanding what type of results one can expect from the method that is an essential determinant in selecting a method for a specific evaluation task and in understanding effects of our choices as evaluators on the evaluation results.

2.3 Positioning usability tests within IS development

A usability test conducted during the software development process is one type of knowledge elicitation intervention with the future users and user organizations. In its classical form, usability testing does not focus on eliciting users' conceptual models or their activities, but on evaluating the system against the set usability goals, detecting the usability problems of the software product and recommending correspondent changes to the system design (Wixon and Wilson, 1997).

Sanders (2006) positions usability testing under UCD methods that emphasize expert's mindset over participatory mindset (Figure 1, diagonal axis). In these methods, designers try their best to understand the world of users and "design for people" (e.g. contextual design, applied ethnography). Thus, designers move towards users (Steen, 2008). Designers do not consider users as partners, but subjects and reactive informers, who are not much empowered in the process (Sanders, 2006). Traditional usability tests are representatives of the expert mindset approach in many ways. Usability experts organize and coordinate a study in time and place, define and recruit an appropriate group of target users, determine the goal-oriented tasks to be performed with a product, and investigate and interpret the results (Sullivan, 1989). Thus, the experts are fully in control. In another end, in methods that emphasize participatory mindset, designers "design with people" who are empowered partners and active co-creators (e.g. lead user approach, co-design, Scandinavian participatory design) i.e. users move towards designers and the communication is intimate and originated by users (Steen, 2008).

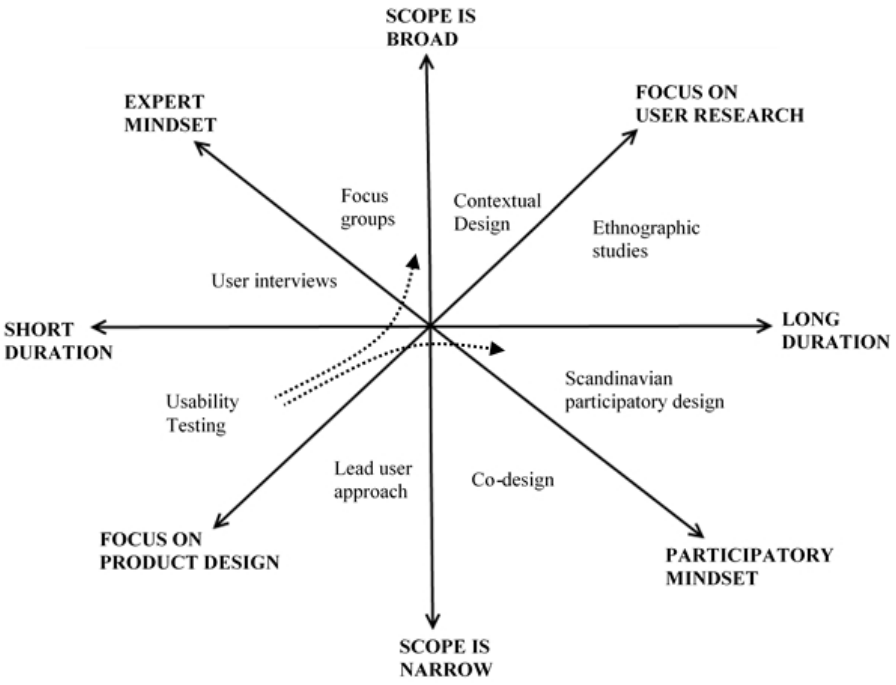

Figure 1. Scope of usability testing among UCD methods. Dimensions and positions are inspired by Sanders (2006), Steen (2008) and Blandford et al. (2008b). Dashed arrows describe the scope expansions discussed in Chapter 3.

Another diagonal dimension of product design versus user research (Figure 1) describes what the method is concerned about (Steen, 2008), and thus reflects the concept of scope as well. The user research -focused methods are concerned with the current situation "as is", whereas the product design -focused methods are concerned with envisioning the future and alternative situations "to be" (Steen, 2008 p. 32). Methods that emphasize the user research orientation carry interest in exploring the current situations of the users and the use contexts in order to first find out what and why some design is needed. In contrast, product design oriented methods (at their purest) consider the technological artefact as an end itself and begin searching "the new thing" without first exploring whether any needs exist for design.

Usability testing has an inherent focus on designing the product rather than an interest on its users. The rationale of a usability test is that evaluators perform it in order to detect and correct the usability problems of the IT artefact. The product vision – the state of to-be – is naturally and tangibly present in the form of new product design tested. The focus on product suppresses the importance of the users to only a few relevant and pre-selected aspects concerning the design. The social and organizational setting, the use context and the activities in which users engage when using the artefact, are not in the essence of usability testing, but usually studied before the usability test takes place, which then can focus on testing the IT artifact only. Moreover, usability evaluations mostly address the fit between the system and the individual user, because usability professionals construe the concept of usability at individual level (Hertzum and Clemmensen, 2012). However, the individual view may not match with "organizational usability" i.e. the socially acceptable and effective integration of the system into the work practices of employees of the organization (Elliott and Kling, 1997). Systems evaluation at work prioritizes the focus on users and their work activities before assessing the potential support from system characteristics. By describing these links between the systemic and contextual characteristics affecting usability, not only the system redesign could take place, but equally learning and change processes in the user organization (Nurminen, 2006, p. 414).

3 Practical Modifications of the Scope of Usability Tests

Various method collections, combinations and modifications expand the scope of traditional usability testing, in order to better understand the user and the work domain beyond the user interface. One of the common drivers has been the notion that conventional, designer-created usability test tasks, which have correct answers and clear endings, are weak in answering usability questions concerning user needs and situated work practices (Redish, 2007). Therefore, to understand and test complex work systems, Redish (2007) suggests using method collections and combinations such as conducting usability studies in the field of users, exploiting multiple evaluators, building simulations, developing situation awareness assessments, implementing unattended long-term data capture and using cued retrospective think aloud method with users.

Similarly, in the search of the indicators of usability of complex work systems, Savioja and Norros (2013) found out that a traditional usability test concerns only with the performance measures of the activities with the tool. Thus, testing lacks in scope: It does not take the perspective of work practices (way of acting) and psychological and communicative functioning of the tool. Based on activity theoretical foundations, they propose a contextual evaluation approach (Contextual Assessment of Systems Usability), which applies tool use simulation according to the scenarios modelled with functional situation representations (a type of an extended task analysis). The scope of the method is on the different perspectives on and levels of work activity as well as tool support when evaluating usability. Similar to Redish (2007), their approach is not leaning only on single data collection technique but for example employs usability questionnaires, interviews, observations, task load measures and expert judgements before and after the simulation sessions.

Følstad and Hornbæk (2010) use a method called cooperative usability test, in order to gain knowledge about the work domain during the test. Several interpretation phases after each task are in the essence of the method. An interpretation phase asks why the user acted in a certain way and thus utilizes user's knowledge of the work-domain to identify and understand usability problems. Thus, the interpretation

phase has similarities with a debriefing phase (Rubin and Chisnell, 2008 p. 229). The basis for interpretation discussions is a task-scenario walkthrough, which aims to bring user comments also on parts of the system that were not used in performing the tasks. As a result, the method's scope expands to system requirements that are outside the responsibility of the developers. Similarly, Spool (2006) exploited more communicative testing approach with users, where test tasks are not assigned to users, but discovered through an interview. The task goals and endings were limited into the web shop context where these interview-based tasks were applied. While these modifications expand the scope of usability testing to "insights about users' domain of interest", it remains rather product design –oriented. The aim of the method is more to identify *"the passionate users"* of the shop, give the users the most realistic test tasks and learn users' wordings, which help in re-organizing the web site content (Spool, 2006).

Kankainen (2002) experienced that traditional usability testing with predefined tasks did not work in getting user feedback when evaluating an early product design. Users evaluated only the interface and said very little about the overall product concept, which made it a torture to users and designers alike. As a method modification, Kankainen (2002) modified the testing into co-discovery exploration, and presented the new design concept with a storyboard and a blank model with accessories. In consequence, the scope, and the focus of users turned towards the overall product concept instead of the interface as such, which was much more useful and inspiring for later design. Similarly, Still and Morris (2010) applied a blank-page technique when testing the usability of paper prototypes. They allowed users to navigate to non-existing pages and dead ends, while they encouraged users to create and design the content for these empty spots. The technique expanded the scope by 1) giving insights into users' mental models and 2) how they conceptualized information encountered.

Blandford et al. (2008) developed 'a concept-based analysis of surface and structural misfits' (CASSM) method, which was motivated by a finding that compromising between a fully naturalistic study and a conventional lab-based study protocol could not identify mismatches between user requirements and system representations i.e. evaluate enough the utility. People are required to work with the concepts implemented into the system, which, when poorly fitted, may place a high workload on the user (Blandford, 2013). For example, in the context of booking flights, test participants are more interested in operating with journeys between places than with flights between airports (Blandford et. al, 2008). The CASSM method then helps in identifying users' conceptualizations of a domain prior to system implementation: It extracts and compares concepts that a user uses to the ones the system implements. The user data is collected with a think-aloud protocol or similar, which provides knowledge of user's procedures for completing the tasks. Thus, CASSM is more an analytical usability evaluation tool than a method for empirical testing and data collecting. However, it broadens the scope of think-aloud protocol and usability evaluation to look at profound misfits of the underlying structures, which, when found, represent typically new design opportunities for the product. Similarly, Johannessen and Hornbæk (2014) expanded the analytical usability evaluation by creating an expert inspection method that focuses on finding utility issues and problems in the context of use of the system.

Concentrating on utility issues on the empirical side, Juurmaa et al. (2013) modified a visual walkthrough method, in order to find elements in the user interface that users consider important or useless. Among other flexible modifications of usability testing methods, Riihiaho (2015; 2009) introduced two more walkthroughs, informal and contextual, where the latter intends for evaluating the usefulness of the system in a professional work setting. In the essence of the contextual walkthrough is to conduct evaluation in the real use context with real data as well as letting the test tasks arise from the users. Similarly, Bødker and Grønbæk (1991) added realism to the evaluation by having everyday materials and tools available to users alongside the product tested. Users were able to demonstrate their current role when going through a typical work task, because of the work material brought into session (Bødker and Madsen, 1998). According to Riihiaho (2015), contextual and informal walkthrough methods can tackle the bias of predefined test tasks. For example, in a call-center work, Riihiaho (2009) found that other unintegrated applications affected to the use situation of the new application and that the physical locations of the other applications became the biggest problem instead of the usability of the new application. In other test cases, informal and contextual walkthroughs have revealed 1) terminology mismatches, 2) technical infrastructure problems, 3) discontinuities of task flows, 4) missing

functions, 5) user misunderstandings and 6) concerns about the post-usage behavior, which all probably would have been out of the scope of the method if only traditional usability test tasks were applied (Riihiaho, 2009).

Similar to the contextual walkthrough above, McDonald et al. (2006) collected data with a contextual interview asking participants to carry out their normal work tasks with the system while thinking aloud. In contrast to the informal walkthrough above, the moderator occasionally interrupted asking questions for clarification purposes. The data analysis was more traditional to contextual design where an affinity wall was exploited to extract usability problems from the overall data. As a result, about two third of found problems were not within the evaluated system itself, but were related to 1) other applications in use (email etc.), 2) lack of user training, 3) lack of documentation and 4) technical and 5) physical environment. Based on the extended scope of found problems, McDonald et al. (2006) conclude that studies in the laboratory premises iterate our understanding about the design artifact, whereas testing in the field iterates our understanding about the use context and the intended value of the product. Without questioning the conclusion, this is a generalization on the favour of the testing environment that may not notify the effect of unstructured test protocol and work-originated tasks on the results (cf. Reijonen and Tarkkanen, 2015).

In their ethnographic model of field usability testing, Rosenbaum and Kantner (2007) also borrow from contextual inquiry practices. They applied both predefined high-level tasks that were same for each participant as well as lookup tasks that "*were of the participant's own choosing and thus were unique from session to session"*. The benefit of latter tasks was that users were more interested in the task and the outcome, and the results were so fundamental that an entire iterative program of field studies were begun. Instead of usability metrics, the results of their tests of online banking for vision-impaired people took the form of "cases" and "scenarios". These addressed 1) the variety of use strategies, 2) behavioral trends and 3) utility issues. According to Rosenbaum and Kantner (2007) field testing is best suited to exploratory objectives where evaluators "*want to learn what problems users encounter as they follow their own work processes"*. However, they suggest using contextual inquiry and ethnographical interview when the primary goal is to understand what people really do with the products or to explore which new features to add. In this manner, Viitanen and Nieminen (2011) pre-explore the work practices of users with the contextual inquiry method before combining an interaction sequence analysis to their usability test. A user research method called 'guerrilla testing' involves the artifact in the pre-exploration and represents a quick way to validate how effective the design is among its intended users and whether the design works in the way supposed.

An approach called ADA (Åborg et al., 2003) was built around users' ordinary work tasks and natural test settings to address both usability and work environment aspects at the same time. Central is viewing the work and tasks as larger units. Although the task assignment is not pre-defined, the evaluator needs to be very familiar with the "aspect" list, which defines in detail what is observed during the session. The list emphasizes user interface issues, but is also exhaustive concerning users and systems in use (e.g. user's role, tasks, competence, system functionality, manuals). However, the predefined list of observable aspects limits the scope and can be irrelevant when evaluating early prototypes. Thus, the method primarily fits evaluating the daily use of systems.

When testing software prototypes in the work, which itself is of open-ended nature (e.g. artistic, creative and knowledge-intensive domains of work), Sy (2006; 2007) pre-explored workflows for a future design by interviewing users on the telephone and began test sessions with contextual investigation. In open work domains, she found that *"scripted usability tasks often set unrealistic constraints on user behaviour that don't match the open-ended nature of the task."* (Sy, 2006 p. 18), and suggests to use open-ended test tasks for more realistic results. Her technique is "*a way to sneak contextual inquiry into a usability testing"* (Sy, 2006, p. 21), where closed test tasks are used only for non-workflow-specific design goals. Open-ended test tasks, which start from a high-level activity that would cover all the tasks to be validated, are based on pre-interviews. In the test session, users are in lead and evaluators direct users only if a certain design goal needs a validation. The scope of tests with open tasks in open work domains include 1) contextual information about users' workflows, especially unexpected uses of the

product; 2) examples of users' work in the application; 3) feature requests; 4) major usability problems; 5) bugs and 6) successes with the design prototype (Sy, 2007).

Open-ended and user-initiated tasks are probably more widely applied in the user experience research than usability research (cf. Bargas-Avila and Hornbæk, 2012). This is because the exploration of true experiences requires systems that can be "let loose" into the everyday practices and lives of people where detailed instructions cannot be given or controlled (Buchenau and Fulton Suri, 2000). The core dimension in user experience research is in the hedonic qualities of the system, such as in emotions and affect, enjoyment and aesthetics (Bargas-Avila and Hornbæk, 2012), which may not always involve goal-oriented activity in contrast to business application domains and professional work settings, on which we concentrate. One of the objectives of using open-ended test tasks in the business application domain, as demonstrated by Sy (2007), is to understand users' goals and means at work as well as systems' applicability in these.

All method modifications introduced in this chapter have their foundations in empirical usability evaluation methods with users as well as in the problems and challenges traditional evaluation practices may confront in the product development. A common denominator of the introduced methods seems to be a need for a shift from highly expert-minded usability evaluation towards more user participative evaluation practice, which gives room for users to explore the system based on their needs, wants, expertise and experience (cf. previous Chapter 2 and the dashed arrows in Figure 1 showing the direction of the extended scope). Emergent is also a need to apply goals that are more user research oriented and "ethnographic" than focusing on the designing and the artefact as the only frames of reference. Consequently, all these methods more or less end up being methods for evaluating systems quality in context due to their extended scope and focus beyond usability as well as valuable, wide-ranging results provided for the subsequent development process. Thus, a question may arise whether these methods are about usability and usability testing – or more about IS quality evaluation in general. A third common factor to these methods is that modifications are created not only to expand the scope per se but to achieve a better fit with the current design process and its challenges at hand, as well as to increase the design value of the results for the specific project.

In summary, we conclude that the evaluation scope is required to expand to areas outside the interaction between the system and the individual user. Another conclusion is that although all scope extensions seem to be complementing and validating the user research efforts in the development, the representations of scope are varied and dispersed into a mixture of case-specific result descriptions. Not to mention the vast majority of usability tests, which are reported as part of the design science activities in scientific literature without the specific goal to develop the evaluation method nor discuss its scope.

As a continuation to the presented methods above, we introduce the findings of our empirical usability test case in the next chapter. The usability test applies an open-ended test task as a test protocol. The protocol aims at broaden the scope of usability testing beyond the technical to the system value in human and social context, likewise the methods introduced in this section. Here, the open-ended test task (or shortly: open task) means *a task assigned to the user containing only a request to use the system with a minimal explanation of the context and the purpose the system.* In its shortest form, the task is just "a please", while in the other cases the intended purpose or the low fidelity of the system requires defining a starting point for use. Many of the usability evaluations and testing methods presented in this chapter integrate other methods in parallel, are analytical or expert evaluations. In contrast, the open test task is a modification *within* the usability testing method only, and unlike Sy (2006), we apply the open task in the complex professional health care domain instead of unregulated and creative work domain. The case study is an introduction to the extended scope of usability testing with the open test task method. It is also one possible realization of representing the scope of usability testing in general.

4 A Case Example on the Extended Scope

4.1 Method description

Our open task test took place in an IS development project, the purpose of which was to design a mobile application called Round for nurses at hospital wards. There are not many studies investigating the impact of mobile EPR tablets on clinical routines at hospital wards and the underlying mechanisms leading to the savings of time remain unclear (Fleischmann et al., 2015). The Round application provided an interface to the current electronic patient record (EPR) system in-use, which then, thanks to the mobility of application, could be instantly accessed during the patient work taking place in the patient rooms. We conducted the first usability test of the application with six (6) nurses at one hospital ward. In the test, Round was a fully functional demo only at the interface level, operated with a mobile tablet device. Two developers of Round (a UX designer and a system architect) followed test sessions and were able to intervene. The participating nurses used the application 1-1.5 hours in front of the table in the hospital premises. The participants were given only an open-ended test task: *"You have just arrived at your workplace and you begin to prepare your work shift. Round is a new application that you can use during your work. (You have already logged in)."* There is relative lack of mobile usability research involving open and unstructured test tasks of this kind (Coursaris and Kim, 2011). The test sessions of two participants expanded to measuring a blood pressure and monitoring a heart rate of a real patient. After the test, we arranged a short meeting with the developer representatives to discuss the first insights and initial results. A full report delivered one week after introduced 57 usability problems. In addition to a list of problems and recommendations, the report included a description of the common phases of nurses' working day i.e. what they do, why, when and what is the result of the work at the hospital ward.

In representing and articulating the scope of usability tests with open test tasks, and the results of this particular case, we utilize the problem classification schemes similar to UAF by Andre et al. (2001) and CUP by Vilbergsdottir et al. (2014). However, we are not following any pre-existing problem classification or values of failure qualifiers, in order to keep the origins of the analysis purely in our empirical data, and to go deeper into the subject of scope. In the data analysis, we reviewed and grouped the found usability problems i.e. gave each problem a category, which abstracts a group of similar problems. In determining the categories, we concentrated on analysing "how something (in the system) is a problem from the users' point of view". The names of the categories in the next section reflect this view. Here, each problem category is a representation of the scope, i.e. the extent to which our method uncovers problems in the IS development, although we like to note that it is not the only possible representation. In the data analysis, both researchers coded and grouped problems independently. As a result an inter-rater reliability showed 78.0% agreement between the two researchers and Cohen's Kappa 0.742 (Cohen, 1960). The result means a moderate agreement in coding and the agreement percentage value is reliable as well, since no categorization was made randomly. (McHugh, 2012). In the next section, we discuss these problem categories as a representation of the scope of the method and provide examples of findings in each category.

4.2 Results

The usability test produced a lot of information about the work practices at the hospital ward. The most important findings of the test consist of identification of the missing, the inadequate and the problematic functionality of the system as well as highlighting previously unexplored design options, which would bring value for users and induce positive changes in their current work practices. Below we provide practical examples in each category of problems that as a whole form the scope for our test in the case.

***Previously unexplored design option in the context*:** We identified during the test that major proportions of care actions and their documentation needs are fundamentally similar. The current EPR at the desktop PC does not support such a unified view on documentation, but diversifies the care documentation into separate system modules and dialog windows each aimed at different care actions. Despite the slight differences in care documentation between different care tasks, the work on the ward and the

use of the Round application would become more effective, if Round supported similar and consistent design patterns for all care tasks as possible.

***Problematic change of work practice due to system implementation*:** With Round, care tasks can be assigned to specific nurses and the task completion rate is visible to other users. Despite the possible benefits of the task structuration, the care tasks that are failed, neglected, delayed or remain undone may lead to unpleasant social pressures within the community and discrimination among nurses. The nurses considered as well that assigning work tasks might hinder and impair their personal autonomy in planning different work duties (which is currently high). For example, how they want to coordinate the tasks between their co-workers and how they personally want to perform these tasks in a certain order and at the specific time of a day (e.g. measuring blood pressure in the morning shift).

***Missing functionality*:** In the beginning of the test and the nurses' imaginary work shift, they would have liked to print on the paper the free form notes entered earlier into the PC-based desktop EPR system. These notes concern patient's health condition, physical abilities, reasons for admission etc. Nurses were used to carry these notes in their pocket along the day. During the patient work, they write new notes on the paper, which they update to the desktop EPR later. In the very beginning of the test, the nurses were incapable of simulating their work with Round due to the missing annotation field, which correspond to these notes on paper. Round aims to replace such a manual task and overlapping documentation by decentralizing the notes under specific care tasks in a structured representation form. On the one hand, nurses thought they wanted to decrease double-documentation, but on the other, they considered the centralized overview on patients' health status important as well. In the test report, we could ponder both design options of Round with and without the notes field, their effects on nurses' practices and organizational system implementation as well as identify the types of notes missing from the current prototype.

***Inadequate functionality*:** Nurses were able to filter the list of patients by the ward and by the nursing team with Round, but most of all they needed filtering "by the pairs" as a third option, because the work on the ward was organized together with a colleague in the same shift. Thus, the filtering function, although implemented, lacked a proper fit with the needs and work practices of the community.

Unfinished physical use context: In addition, this case concretized some of the limitations of applying open tasks in a simulated environment compared to testing in real patient work with real patient data. Two out of six participants took the application into clinical work partly on their own initiative (possible due to open task). Therefore, we could observe the limitations of the physical device and environment that we would not have observed when sitting at the table without a contact with patients. First, the tablet-based device was not very feasible to carry with other care equipment. Inpatients at hospital wards vary in their physical health condition that restricts their abilities to move, which naturally implies that nurses and physicians are in the constant move from room to room while caring the patients. In the studied hospital ward, the room doors are closed and rather heavy to open. Therefore, the nurse could hardly open the door with the tablet on the one hand, and a blood glucose meter on the other. She opened the door with a little finger. *Second*, when the nurse began to measure the blood glucose of the patient she did not find a proper place to put the tablet device down. Tables next to the beds are for patients' personal use and may not be free for the tablet device. Clearly, usability problems related to the physical appearance of the mobile device are difficult to find in a test at the desk even when using the open task approach. The possible solutions to these problems are not only in the hands of software developers if at all. Software developers' option is to implement the application in a smaller device (e.g. smart phone version), yet more comprehensive design and involvement of user organization are required, in order to provide more table space next to patient beds, sewing larger pockets for nursing jackets, purchasing carrying bags for equipment or keeping doors open at the ward etc. The user organization with its practices and policies is equally a key stakeholder in improving the usability of information systems.

Unexpectancies of situated work: *Third*, while measuring the blood pressure, the nurse finds out that the application does not support recording the saturation and CRP values. Although the other participants of the test notified the lack of features, it is notable that simulating the work and personal practices in the lab-like premises without the real patient contact cannot always be exhaustive. *Fourth*, even when

testing with the open task approach, work tasks easily become treated as separate entities, which follow the order, which is in the test subject's mind at the time of the session. In any service type of work, the flow of tasks depends on the client side as well and sometimes it consists of unanticipated turns and 'jumps' that may not become visible in the test performed in lab-like premises. This happened also in the patient contact: Due to discussion with the patient, the blood pressure measurement task expanded to two other tasks, specifically tasks related to patient medication. The nurse could not anticipate (or remember to ask in the first place) that the patient needed both a painkiller and a digestive medication. Both of these new tasks need official medication record entries, a note for the nurse herself and possibly a note for the colleagues sharing the care tasks with her (because somebody needs to remember to give the medicine). However, the nurse did not turn to Round application even though it would have supported the tasks. Clearly, participants are not always ready to use the new design in the familiar situation, which is of course very natural as participants cannot know all the possibilities of the new software and the open-ended task approach does not offer only doable test tasks or give hints what actions are needed next. Thus, participants' behaviour requires careful attention from the evaluator, possible intervention and correction after observing and recording their initial actions and aims with the new system.

5 Discussion

Usability testing is a well-known concept in software product evaluation. Due to its institutionalized status as the must-be method, its scope – the extent to which it covers problems in the development – would appear to be well defined. Our literature review, however, clearly indicates that usability testing is not a singular method with one well-defined scope, which evaluators could always lean on, or even know in advance, when applying it in practice. If we observed the UX industry practitioners at work, we would find many forms of usability testing practices and different scopes. Organizations who develop and purchase information systems, and outsource evaluation activities, cannot afford to bypass problems that are out of the scope of usability testing, nor be relying on a vague understanding of the scope.

The methods brought up from the literature also show that the scope of a method is affected by many other elements than the test task or scenarios provided during the test, which was the main modification in the case study presented here. However, as the case study example shows, diminutive changes in the test protocol, such as the open-ended task itself and the subsequent possibility to use the system in a real interaction, can radically expand the scope towards new areas. In contrast to other methods discussed in Chapter 3, both our modifications were applied *within* usability testing without additional pre- or post-phases or deliberately attached methods. The former modification, use of the open test task, set the basis for findings, which were not only human-computer interaction specific, but expanded to cover complex social relations of users and their concerns related to possible social changes due to the system implementation. The latter modification was a subordinate to the open task and took the scope of the case method to deal with physical limitations set by the device and the environment (opening the door and lack of space in the table). Thus, it is a representation of a scope which is not usually achievable with tests in lab premises. Both modifications brought findings, which the future design nor the user organization can bypass, in order to make the system effectively, efficiently and satisfactory usable for the nurses in the hospital ward.

Based on the case study, the scope of usability testing with open task seem to cover equally well the system utility aspects and the more traditional usability problems. Open task method has a scope which finds problems that 1) render doing the job with the system impossible (missing and inadequate functions), 2) are prone to cause unfavourable and uncontrollable consequences in the users' work (problematic social changes) 3) require more user research and context exploration for more benefits (unexplored design opportunities) as well as problems which 4) can prevent users' task completion, cause inefficiency and unsatisfied users (physical limitations and unexpectancies in the service work). The scope covers problems which comply with the classical definition of usability, whereas other problems are in contradiction to the goals and tasks of the users. With the former type of problems, designers can produce alternative solutions without challenging their understanding about the context of use and the

collected requirements. Problems related to the utility-scope usually require a deeper user research. It is vital that a practical usability test can address these both aspects of system usefulness, because possible solutions to the problems are different.

Here, the scope is tightly intertwined with the expected outcomes – practically a synonym (due to our data analysis practices). One may want to represent the scope of the method in other terms as well. For example, Reeves (2019), although not speaking with the term 'scope', observes findings emerging in a usability test from four different 'relevancy devices': Some user troubles become insights, other are issues or recommendations in the final report whereas some troubles become dissipated through the discussions in the observation room. Following this, we see the scope not only as a result of method prescriptiveness but also as a result of such a collaborative work of stakeholders towards producing findings i.e. as a result of a positive evaluator effect.

Our case would easily stretch also to a more abstract materialization of scope: The scope could be presented as (the number of) usability findings targeted at the technical system, the social context and the physical environment. Important is that these targets are recognized and articulated, regardless of the classifications and formats used to represent and describe the scope of the method. In any case, identifying the scope of a method calls for more than a basic method description containing the name of the method, its strengths and weaknesses and possible usage phases. Every evaluator and researcher can take similar analysis effort after the studies they conduct, and eventually this work serves the community of evaluators. However, we acknowledge that the scope of a method, the method's outcomes and boundaries, cannot be thoroughly described – the work will never finish – and it would rather lead us into a situation described by Gray (2016) where *"a designer would have to make decisions about the limits of the method in situations that are explicitly coded for."* Despite our representation of the scope here as problem categories, following Gray (2016), we also consider the methods as merely tools and players in a design game instead of an objective set of outcomes. Even if methods are considered prescriptive but situated, the discussion about the scope is relevant, because each test outcome is a beginning of a new design iteration and an opportunity to learn.

6 Conclusions

This paper discusses the scope of usability testing method in IS development. The scope is considered important in understanding the method validity and effectiveness, and subsequently, for selecting the right method for the evaluation case at hand. This paper was an attempt to give rise to a discussion about differences in the scope of usability tests and how the elements and applied protocols of usability testing effect the scope. From a theoretical perspective, the study contributes to shifting the scope of usability testing towards more participatory and user research centric direction, which are fundamentally required, yet easily dismissed in IS evaluation and IS development practices. Literature has shown for a long time that usability testing does not naturally implement the view of users and organizations including wider socio-technical design dimensions. Therefore, these should not be taken for granted in usability testing, but deliberately attached to the method performance requirements when needed. The practical contribution of the paper lies in introducing the scope-broadening method modifications of the literature, the usage and value of which the presented case study further exemplifies. Although, based on this study, it is impossible to find only one scope for usability testing, the implication is that the discovered and experimented shift in the method scope is both possible and valuable in practice. By cutting the link between the design and the evaluation process, the presented methods and the method of the case specifically, serve both the technology developers and the end-user organizations equally and rather cost-effectively. As the case study shows, only a moderate change in the test task towards openness allows user control and freedom in the test session, which further reveals for example unexplored design options and problematic future changes at organizational and community levels of work. Thus, the open-ended test task introduces one possible realization of usability testing, which considers organizational and social factors beyond individual users.

References

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

Bargas-Avila, J., and Hornbæk, K. (2012). Foci and blind spots in user experience research. *Interactions, 19(6), 24*

Bernhaupt, R. Palanque, P. Manciet, F. and Martinie, C. (2016) User-Test Results Injection into Task-Based Design Process for the Assessment and Improvement of Both Usability and User Experience. *Human-Centered and Error-Resilient Systems Development.* Volume 9856. Lecture Notes in Computer Science, pp. 56-72.

Bjørn-Andersen, N., and Clemmensen, T. (2017). The Shaping of the Scandinavian Socio-Technical IS Research Tradition. Confessions of an accomplice. *Scandinavian Journal of Information Systems,* 29(1), 4.

Blandford, A. (2013). Eliciting People's Conceptual Models of Activities and Systems. *International Journal of Conceptual Structures and Smart Applications* (IJCSSA), 1(1), 1-17.

Blandford, A. E., Hyde, J. K., Green, T. R., and Connell, I. (2008). Scoping analytical usability evaluation methods: a case study. *Human–Computer Interaction*, 23(3), 278-327.

Bødker, S. (2006). When second wave HCI meets third wave challenges. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (pp. 1-8). ACM.

Bødker, S. and Grønbæk, K. (1991). Cooperative Prototyping: Users and Designers in Mutual Ac-tivity. *International Journal of Man-Machine Studies, 34, Special Issue on CSCW* (453-478).

Bødker, S., and Madsen, K. H. (1998). Context: an active choice in usability work. *Interactions*, 5(4), 17-25.

Buchenau, M., and Fulton Suri, J. (2000). Experience prototyping. *In Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 424-433). ACM.

Cho, C.-S., and Gibson Jr., G. E. (2001). Building Project Scope Definition Using Project Definition Rating Index. *Journal of Architectural Engineering*, 7(4), 115.

Cockton G. (2006). Focus, Fit, and Fervor: Future Factors Beyond Play With the Interplay. *International Journal of Human-Computer Interaction*, 21:2, 239-250

Cockton, G. (2004). From quality in use to value in the world. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1287-1290). ACM.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement 20(1)*, 37–46.

Coursaris, C. K., and Kim, D. J. (2011). A meta-analytical review of empirical mobile usability studies. *Journal of Usability Studies 6, 3, 117-171.*

Dumas, J. and Redish, J. (1999). A practical guide to usability testing (revised ed.). Intellect Ltd, Exeter

Elliott, M., and Kling, R. (1997). Organizational usability of digital libraries: Case study of legal research in civil and criminal courts. *Journal of the American Society for Information Science*, 48(11), 1023-1035.

Eshet, E., and Bouwman, H. (2015). Approaching Users and Context of Use in the Design and Development of Mobile Systems. In: Marcus A. (ed.) *Design, User Experience, and Usability: Users and Interactions. DUXU 2015.* LNCS, vol. 9187. Springer, pp. 508-519.

Fleischmann, R., Duhm, J., Hupperts, H. and Brandt S.A. (2015). Tablet computers with mobile electronic medical records enhance clinical routine and promote bedside time: a controlled prospective crossover study. *J Neurol Mar 262(3):532-540*

Følstad, A., and Hornbæk, K. (2010). Work-domain knowledge in usability evaluation: Experiences with Cooperative Usability Testing. *Journal of systems and software*, 83(11), 2019-2030.

Gray, C. M. (2016). What is the nature and intended use of design methods. *Proceedings of the Design Research Society.* 27-30 June 2016. Brighton, UK.

Guay, S., Rudin, L., and Reynolds, S. (2019). Testing, testing: a usability case study at University of Toronto Scarborough Library, *Library Management 40(1)*, 88-97.

Heiskanen, E., Hyysalo, S., Kotro, T., and Repo, P. (2010). Constructing innovative users and user-inclusive innovation communities. *Technology Analysis & Strategic Management,* 22(4), 495-511.

Hertzum, M. (2016). A Usability Test Is Not an Interview. *Interactions,* March-April pp. 82-84. ACM.

Hertzum, M., and Clemmensen, T. (2012). How do usability professionals construe usability? *International Journal of Human-Computer Studies,* 70(1), 26-42.

Hertzum, M., Molich, R., and Jacobsen, N. E. (2014). What you get is what you see: revisiting the evaluator effect in usability tests. *Behaviour & Information Technology,* 33(2), 144-162.

Holmlid, S. (2009). Participative; co-operative; emancipatory: From participatory design to service design. *In Conference Proceedings ServDes. DeThinking Service ReThinking Design*. Oslo Norway, pp. 105-118. Linköping University Electronic Press.

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

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

Hornbæk, K. and Frøkjær, E. (2005). Comparing usability problems and redesign proposals as input to practical systems development. In: *Proceedings of ACM conference on human factors in computing systems*. New York: ACM Press, 391–400.

Hyysalo, S. (2015). Redrawing the landscape of designing for, with and by users. *Keynote in SCIS 2015 - System design for, with and by users,* Oulu, Finland. August 9-12, 2015.

Iivari, N. (2006). *Discourses on 'culture' and 'usability work' in software product development*. Doctoral dissertation, University of Oulu, Finland.

Johannessen, G. H. J., and Hornbæk, K. (2014). Must evaluation methods be about usability? Devising and assessing the utility inspection method. *Behaviour & Information Technology,* 33(2), 195-206.

Johnson, M., Mozaffar, H., Campagnolo, G. M., Hyysalo, S., Pollock, N., and Williams, R. (2014). The managed prosumer: evolving knowledge strategies in the design of information infrastructures. *Information, Communication & Society*, 17(7), 795-813.

Juurmaa, K., Pitkänen, J., and Riihiaho, S. (2013). Visual walkthrough as a tool for utility assessment in a usability test. *In Proceedings of the 27th International BCS Human Computer Interaction Conference* (p. 30). British Computer Society.

Kankainen, A. (2002). *Thinking model and tools for understanding user experience related to information appliance product concepts.* Espoo. Acta Polytechnico Scandinavica. Mathematics and Computing Series, 2002, No. 1 18

Kensing, F. and Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)*, 7(3-4), 167-185.

Kujala, S. (2003). User involvement: a review of the benefits and challenges. *Behaviour & information technology*, 22(1), 1-16.

Kuutti, K. (2011). Out of the shadow of Simon: artifacts, practices, and history in design research. *In Proceedings of the Doctoral Education in Design Conference,* Hong Kong 23-25.5.2011.

Kuutti, K., and Bannon, L. J. (2014). The turn to practice in HCI: towards a research agenda. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 3543-3552). ACM.

McDonald, S., Monahan, K., and Cockton, G. (2006). Modified contextual design as a field evaluation method. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (pp. 437-440). ACM.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica 22(3)*, 276–282.

Molich, R. and Dumas, J.S. (2008). Comparative Usability Evaluation (CUE-4). *Behaviour & Information Technology* 27(3) 263-281.

Nurminen, M. I. (2006). Work Informatics—An Operationalisation of Social Informatics. *In IFIP International Conference on Human Choice and Computers,* pp. 407-416. Springer, Boston, MA.

Rajanen, M., Iivari, N., and Anttila, K. (2011). Introducing usability activities into open source software development projects–searching for a suitable approach. *Journal of Information Technology Theory and Application,* 12(4), 5-26.

Redish, J. (2007) Expanding Usability Testing to Evaluate Complex Systems. Journal of Usability Studies 2, pp. 102-111.

Reeves, S. (2019). How UX Practitioners Produce Findings in Usability Testing. *ACM Transactions on Computer-Human Interaction 26(1)*, 3.

Reijonen, P. and Tarkkanen, K. (2015). Artifacts, Tools and Generalizing Usability Test Results. *In Scandinavian Conference on Information Systems* (pp. 121-134). Springer.

Riihiaho, S. (2009) User testing when test tasks are not appropriate. In: L. Norros, H. Koskinen, L. Salo and P. Savioja (Eds.) European Conference on Cognitive Ergonomics, pp. 228-235. VTT Technical Research Centre of Finland.

Riihiaho, S. (2015). *Experiences with usability testing: Effects of thinking aloud and moderator presence.* Doctoral dissertation. Aalto University publication series.

Rosenbaum, S., and Kantner, L. (2007). Field usability testing: method, not compromise. *In IEEE International Professional Communication Conference*. (pp. 1-7). IEEE.

Rubin, J. and Chisnell, D. (2008). *Handbook of usability testing (2nd ed.): How to plan, design and conduct effective tests.* Wiley Publishing Inc., Indianapolis.

Sanders, L. (2006). Design Research 2006. *Design Research Quarterly* 1. 11 of Nov. 2006

Sanders, E. B. N. and Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-design,* 4(1), 5-18.

Savioja, P., and Norros, L. (2008). Systems usability—promoting core-task oriented work practices. In *Maturing Usability: Quality in software, Interaction and Value* (pp. 123-143). Springer London.

Spool, J. (2006). *Interview-Based Tasks: Learning from Leonardo DiCaprio.* Blogpost. Retrieved 19.10.2017 from https://articles.uie.com/interview_based_tasks/

Steen, M. (2008). *The fragility of human-centred design.* Thesis PhD. Delft University of Technology

Stewart, J., and Williams, R. (2005). The wrong trousers? Beyond the design fallacy: social learning and the user. *Handbook of Critical Information Systems Research,* 195.

Still, B., and Morris, J. (2010). The blank-page technique: Reinvigorating paper prototyping in usability testing. *IEEE Transactions on Professional Communication,* 53(2), 144-157.

Suchman, L. (1985). *Plans and Situated Actions. The Problem of Human-Machine Communication* (Thesis). Xerox PARC. ISL-6.

Sullivan, P. (1989). Beyond a narrow conception of usability testing. *IEEE Transactions on Professional Communication*, 32(4), 256-264.

Sy, D. (2006). Formative usability investigations for open-ended tasks. *Proceedings of UPA* 2006.

Sy, D. (2007). Adapting usability investigations for agile user-centered design. J*ournal of Usability Studies, 2(3),* 112-132.

Viitanen, J., and Nieminen, M. (2011). Usability evaluation of digital dictation procedure - an interaction analysis approach. *In Symposium of the Austrian HCI and Usability Engineering Group* (pp. 133-149). Springer, Berlin, Heidelberg.

Vilbergsdottir, S. G., Hvannberg, E. T., and Law, E. L. C. (2014). Assessing the reliability, validity and acceptance of a classification scheme of usability problems (CUP). *Journal of Systems and Software*, 87, 18-37.

Wixon, D. and Wilson, C. (1997). The usability engineering framework for product design and evaluation. *Handbook of human-computer interaction*, *2*, 653-68.

Åborg, C., Sandblad, B., Gulliksen, J., and Lif, M. (2003). Integrating work environment considerations into usability evaluation methods—the ADA approach. *Interacting with Computers,* 15(3), 453-471.