Usability Evaluation: Past, Present and Future

Johanna Viitanen

Software Business and Engineering Laboratory
Helsinki University of Technology
{johanna.viitanen@tkk.fi}

(Initial version, April 19, 2007)

Abstract. Usability evaluation was introduced in late 1980s with two tools available for making judgments about the product usability: usability testing and inspection methods. In the past decade usability evaluation rose to popularity: many handbooks regarding how to conduct usability testing in laboratory settings were published and within industry usability tests were seen as a way to measure and maximize the “ease of use”. Research literature on usability evaluation was largely about comparing different methodologies and variety of evaluation methodologies applied in different phases of user-centered design processes. In the beginning of 21st century surveys concerning the current user-centered design practices in industry indicated that usability evaluation methods were commonly used, and formative methods like usability testing were found highly effective for influencing decision making in organizations. However, concerns of insufficient usability evaluation methods have been proposed recently: new technology, product integration and shortened product timescales are raising challenges, as well as limited budgets and resources in industry. Within the industry integrated user-centered design continues to be enhanced, meanwhile cost-effective, quick and clean usability evaluation methods are needed.

Keywords: Usability evaluation, usability testing, evaluation methods, evolution of usability evaluation, usability evaluation in industry, challenges for usability evaluation methods

1 Introduction

‘Usability evaluation with users’ is primarily how the concept of usability and user-centered design is related to product development and understood within people. User-centered design refers to a multidisciplinary design approach based on the active involvement of users for a clear understanding of user and task requirements, and the iteration of design and evaluation (ISO 13407). Usage evaluation is considered as being an essential part of user-centered design process: the first tenet of user-centered design principles is to test early and often. In the context of this paper, the term usability evaluation method is taken to refer to any method or technique used to perform usability evaluation, with
emphasis on formative usability evaluation and an interaction design at any state of its development.

The most common user testing method is a *usability test* in which a participant does given tasks with evaluated system. Usability test is a widely used technique to evaluate user performance and acceptance of products, and is probably the best know method for applying usability and for offering project team information about humans. In the past decade usability testing has proven its worth as a crucial part of user-centered design.

Usability evaluation reaches back to the beginning of human-computer interaction, usability evaluation methods go back more than a decade and studies for comparing usability evaluation methods have been conducted for some time too. However, in a broad historical view, the area is still relatively new and incomplete as both a research topics and as an applied body of knowledge.

There is tremendous emphasis in industry on collecting user data concerning products as one enters the 21st century. This paper discusses the status of usability evaluation as it is known today with reference to future trends. The evolution of the field is considered within the history of usability evaluation, the growth of usability and evaluation, usability in industry and usability from business point of view. This paper does not attempt to identify or detail the origins of the field, but instead focus on how it has changed during last 20 years.

Objectives of the paper: With help of literacy available make a review about the history and evolution of usability evaluation and thereby constitute insights about the future of the usability evaluation as a part of ever growing user-centered design approach.

Contents and short descriptions: The remainder of this document is structured as follows: …

2 Usability Evaluation – Evolution and Revolution

2.1 History of Usability Evaluation

Usability evaluation was introduced in the late 1980s. At that time user-centered design was usually thought as being merely usability testing or software engineering.

In the late 1980s journals and conferences feature papers on usability issues. Evaluation papers at the early human-computer interaction conferences were largely about measuring and broadly comparing different interface components. The focus of the evaluation shifted from product assurance testing to integrated product design and development (Karat, 2003).

In the end of the 1980s it seemed that usability testing was becoming a common tool for evaluating and improving user interface and documentation for complex, computer-based products. Within the industry software and documentation managers were trying to budget resources and schedule time to plan and test for usability. The “easy of use” was the guiding objective for evaluation: the more
successfully people could accomplish their objectives, the more usable the product were judged. Two tools were available for making judgments about product usability: usability test and expert evaluation.

Usability testing quickly became the primary usability evaluation method for examining a new or modified interface. Traditional usability testing was conducted in carefully controlled laboratory settings, in which people were asked to complete tasks with the product under evaluation. Usability testing involved user performance testing to evaluate speed, accuracy, and errors in addition to user subjective evaluations.

Usually schedules and budgets only permitted one evaluation per project. The evaluation was conducted in the later phases of the development process. It was clearly noted that any sub-project in the development process, which threatens to delay the product release, can meet heavy opposition. (Rosenbaum, 1989)

2.2 The Growth of Usability and Evaluations

In the early 1990s, significant attention was paid to the ergonomics, the literature describing usability testing techniques, labs and applications, as well as methodological considerations.

In the 1990s, several ever since famous handbooks of how to conduct usability evaluations in laboratory were published: “Usability Engineering” by Jakob Nielsen (Nielsen, 1993), “Handbook of Usability Testing” by Rubin (Rubin, 1994), and “A Practical Guide to Usability Testing” by Dumas and Redish (Dumas & Redish, 1999). Usability testing in laboratory settings was seen as a way to minimize the cost of service calls, increase sales through the design of a more competitive product, minimize risk, and create a historical record of usability benchmarks for future releases (Rubin, 1994). Since the importance of human involvement in product development was realized, usability tests, conducted in the later phases of the process, gained industry acceptance.

In the late 1990s, ISO 9241-11 standard (ISO 9241, 1998) presenting a kind of high-level guidance on how to specify and evaluate usability in different contexts was released. The information in part 11 (entitled “Guidance on Usability”) included the definition for the term usability and discussion of how to go about measuring it.

As the ISO 9241-11 standard indicates, the context of use needs to be taken into account in design as well as in evaluation. In the end of the 1990s century, user-centered design approach was not only dealing with software systems nor applications used in controlled environments like office workplaces. It became obvious that laboratory studies could not capture the richness of technology and various social, physical and temporal context of use.

Many developers explored other methods in an attempt to bring down the cost and time requirements of traditional usability testing. In addition, because usability testing often occurs late in the design process, developers were motivated to look at methods that could be used earlier in the development process. As a result, expert-based inspection methods, such as heuristic evaluation
(Nielsen, 1993), grew in popularity because many of them were intended to be used with a relatively early design concept.

Existing variety of evaluation methods gave rise to an interesting challenge for development process: How effective product development process can combine usability evaluation methods, applying methods at different times during development. It was concluded, that an iterative sequence of heuristic evaluation followed by laboratory testing achieves the greatest value from the method (Nielsen, 1993, Kantner & Rosenbaum, 1997): the heuristic evaluation makes a first pass at catching the most visible usability problems, enabling laboratory testing to focus on deeper issues. For the reason that different evaluation methods serve different evaluation purposes and reveal different problems, methods should be used as a complement to each other.

The availability of easy-to-learn software prototyping tools, the resurgence of paper prototyping and a variety of evaluation methods made it possible to push evaluation earlier and earlier in the design lifecycle. Although in 1990s there appeared to be a general awareness among the practitioners of the importance and the basic philosophy of user-centered design, Gould et al. (Gould & al., 1991) admitted that the user-centered design process was still not often used due to both organizational and technical reasons.

During the past decade there have been a number of research studies comparing the strengths and weaknesses of different techniques that can be applied in a usability evaluation (Bailey, 1992, Karat, 1992, Henderson 1995, Molish 1998, Riihiaho 2000). As consequence of variety of alternative approaches and therefore occurred general lack of understanding of the capabilities and limitations of each methods, Hartson et al. conducted a study of criteria for evaluating usability evaluation methods (Hartson et al. 2000).

### 2.3 Usability Evaluation in Industry

The research of Nielsen and Mack (Nielsen and Mack, 1994) revealed that traditional laboratory testing with users was probably the most widely used technique in industry in mid 1990’s. At about the same time Dillon et al. conducted a survey of current practices in design for usability across the European software industry (Dillon et al. 1993). Results indicated that current practices are more informal than formal. Though dedicated facilities and staff are becoming commonplace, typical evaluations are either carried out by the designer or handed over to outside consultants. Respondents felt that lack of time and resources were the major drawbacks to designing for usability although quantifying usability, planning effective evaluations and establishing the cost-benefits of usability evaluations were also raised as issues.

In the beginning of the 21st century, research about the most commonly used user-centered design methods were conducted. All three surveys (Hudson 2000, Gunther et al. 2001, Vredenburg et al. 2002) conducted in early 2000s indicated that same kind of results: Evaluation methods (usability evaluation and formal heuristic evaluation) were considered very important on the basis of their actual impact on product development. Informal expert review was widely used likely
because of its low costs, but it was not considered having a high impact. User-centered design methods were generally considered to have improved product usefulness and usability, although the degree of user-centered method adoption is quite uneven across different organizations.

Taken together, in the beginning of the 21st century, the usability testing was still considered perhaps the most important user-centered design method (Rosenbaum et al., 2002). The adoption of user-centered design methods within industry has followed a similar evolution over past two decades. Typically organizations begin with the implementation of formative usability studies (like usability testing), often accompanied by the building of one or more usability labs, and then add field studies and other user-centered design methods to their repertoire (Rohn et al., 2002).

Studies conducted using formative methods like usability testing are a critical tool in user-centered design toolkit, but insufficient as the sole tool to use. In addition studies are said to be highly effective for influencing decision making in organizations (Rosenbaum et al. 2002). The degree to which usability testing flourished over the past decade had been surprising: many practitioners had assumed the need for usability testing would decline as more companies came to accept the early phases of the usability engineering process, namely development of customer requirements and rapid prototyping design (Wichansky, 2000).

### 2.4 Usability Evaluation and the Business Standpoint

When talking about business strategies, return on investments, decision making, cost-benefit trade-offs and savings, usability evaluation is often mentioned. It is argued, that evidence of ROI for usability in user-interface design is clear. When usability is included into development and production processes, even great returns are possible.

(It is easy to criticize usability evaluation for not being ‘enough’ from the strategic design point of view: Conducting traditional usability tests with real users is expensive. When evaluation is conducted in the later phases of the product development, correcting a problem may cost 10 times as much as fixing the same problem in design (Nielsen). Only one product is usually involved in testing. Evaluation and results are closely related the stage of the development the product is going through. Also worth noting is, that only big companies can afford on their own usability teams, others do have to use consultation. If the company developing a product is not familiar with the usability domain, taking full advantage of the written usability evaluation report, problems found and especially the most important - ideas of enhancements, might be challenging.)

User-centered design methods are generally considered to have improved product usefulness and usability, although the degree of method adoption is quite uneven across different organizations. User-centered design appears to be making an impact across the industry. Cost-benefit tradeoffs are a key consideration in the adoption of user-centered methods. However, measures in user-centered effectiveness are lacking and rarely applied. (Vredenburg, 2002)
Challenges: Communication techniques between the designer and the evaluator. In dealing with the technical, logistical and social challenges presented by evaluating a new collaborative technology, one element of usability is critical: the ability to communicate with the product team.

Challenges: How to influence the product design and to view the testing from broader point of view? Suggestions: Involve designers in test planning, Encourage test participants, Structure the tests report to include long-range issues (Dumas, 1989).

Challenges: Usability specialists as change agents. The initial motivation to accomplish usability testing came from the desire to improve the products that are tested. Nevertheless, as early as in 1989 it was argued that the impact of usability testing can go far beyond influencing a product design. Testing can be used for raising the awareness of product designers about usability issues, improving the technical and managerial skills of product designers and product managers and stimulating cooperation among design team members, such as engineers and writers (Dumas, 1989). To achieve these results usability specialists must broaden their view of testing as a way to simply to improve products. Usability specialists have the opportunity to become a change agent in their organization or in the organization they are conducting the test for.

Challenges: The role and impact of user-centered design in development process: how to introduce user-centered design into development organizations?

2.5 The State of Usability Evaluation Today

Usability testing is alive and well at the turn of the 21st century. It is a vital methodology that evolves to suit the technology of the day. The testing in 21st century should be quick and clean, not quick and dirty. The goal of testing should be to make tests fast, valid, reliable and comparable in results with existing data on competitive products.

Traditionally, field studies are mostly conducted within large organizations that can invest in research for long-term product design improvements. In contrast, usability testing methods are regularly used in short-term data collection projects. The literature on “discount usability” and many other published case histories describe the successful application of usability testing to achieve immediate commercial goals. Usability testing – especially iterative usability testing – is easy to justify and highly productive. (Rosenbaum, 2002)

It is not possible to test usability effectively from the back end of the development process. However, careful gathering of requirements and small iterative tests on prototypes may be sufficient to meet customer needs without lengthy usability testing at the end of the development process.

Usability evaluation – Topics of today (each of these will be described shortly):
- The focus of the research and practice has moved from interface to interaction.
- The amount of test users needed for usability tests (Nielsen, Bailey, Spool and Schroeder)
- Comparative usability evaluation (Molich)
- Automating usability evaluation
- Subjective evaluation, evaluator effect
- Cost effective evaluation methods and processes
- Remote testing and international testing
- Common reporting formats and methods for industry
- Rapid development processes and short product lifecycles
- Extending the concept of evaluation: 1) Inspections methods, including heuristic evaluation, are not only an effective means to improve the usability of a product, but they are also effective as a means to educate colleagues and improve process. 2) Every usability evaluation should be analyzed not only for what is right and wrong with the product, but also for what is right and wrong with the development process. Thus, every usability evaluation should address improvements in both the product and the process. (Redish)

Usability evaluation – Challenges in the 21st century (each of these will be described shortly):
- Testing internet applications and mobile handheld devices
- Conducting tests in natural environments. Textbook methods like traditional usability testing often have to be adapted to different contexts.
- Usability methods applied to the evaluation have to adapt to match the target user groups as they are redefined.
- New technology and product integration: There is a trend toward product integration increased complexity.
- Consumer experience instead of ease of use: Businesses are struggling with unwieldy usability techniques in environments where the usability engineering issues are increasingly demanding and complex and are better conceived of as being about “consumer experience” than “ease of use”. (Thomas & Macredie, 2002)

→ Are current evaluation methods enough anymore?

3 Conclusions

Alustavia ajattelua:
- Evaluoinnin vaikutukset helposti todennettavissa → yksi syy menestykseen?
- Toisaalta, arvioinnin hyötyjen todistaminen numeroilla arveluttavaa, harhaanjohtavaa ja lyhytnäköistä toimintaa. Arvioinnin hyötyjen todentamiseen ja ulottuvuuksiin liittyvät useita epävarmuustekijöitä: hyödynnetäänkö arvioinnin tuloksia vain yhden tuotteen kehityksessä ja kuinka suuri osa arvioinnin tuottamasta tiedosta todellisuudessa tulee suunnittelijoiden tietoon ja hyödynnettäväksi.
- Muutokset UCD menetelmissä ja tuotekehitysprosesseissa: yritykset tiedostavat UCD:n tärkeyden → enemmän huomiota ja resursseja tuotekehityksen alkuvaiheisiin ja ”kattavuuteen” yksittäisten
arviointitapausten sijaan (field studyt) → käytettävyyssyksiköiden perustaminen ja systemaattinen toiminta → miten heijastuu arviointiin?

- Yritysmaailmassa käytettävyyssarvioinnin onnistuminen on monen tekijän summa: resurssit tuotekehityksessä, testaus oikeassa ympäristössä, testitulosten uskottavuus ja luotettavuus, yhteistyö arvioijien ja kehittäjien välillä, arviointiraportin hyödynnettävyys, ...

- Vaikka perinteistä arviointia kohtaan on esitetty kritiikkiä sen riittämättömyydestä ja nykyajan mukanaan tuomista haasteista ja vaatimuksista, voi käytettävyyden arviointi menetelmänä edelleenkin hyvin. Suurissa yrityksissä arvioinnin merkitystä pidetään edelleen korkeana, tuotteita ei haluta päästää tuotantoon ennen kuin ne on arvioitu käyttäjien kanssa. Kirjoista löytynnen menetelmien ja niiden tarkan toteuttamisen sijaan käytettävyyssasiantuntijoilta vaaditaan kuitenkin nykyään mielikuvitusta ja syvempää ymmärrystä aiheesta: miten vähillä resursseilla saadaan hyviä ja luotettavia tuloksia, missä kontekstissa tuotetta tulee testata, mitkä ovat keskeiset testattavat ominaisuudet jne.

References