Usability Evaluations – User Experiences  
– Usability Evidence

Geir Karsten Hansen  
Norwegian University of Science and Technology  
geir.hansen@ntnu.no

Nils O. E. Olsson  
Norwegian University of Science and Technology  
nils.olsson@ntnu.no

Siri Hunnes Blakstad  
Norwegian University of Science and Technology  
siri.blakstad@ntnu.no

Abstract

In this paper, we elaborate and reflect on usability evaluations and the applicability of results for other projects. Usability evaluations are based on different user’s experiences and assessments on how well the buildings perform. Evaluations are based on different methods and aspects, depending on objective, purpose, focus, competence and resources. Usability is defined as “the extent, to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.” (NS-EN ISO 9241/1998). The paper is based on theory on evaluations methods and qualitative research. The empirical data is mainly from several case studies mainly from two Norwegian usability projects undertaken in the last 8 years. The case studies have covered a broad range of methods used in order to test their suitability related to the concept of usability and within a common methodological framework USEtool. The paper argues that usability evidence must be related to contextual understanding. Usability evaluations have to look on the interaction between architecture, technology and organisation. Previous studies have shown that in order to assess usability one has to focus on the effect of the building on the user organizations fulfilment of goals, as well as the end users satisfaction and experience. Theory affirms that context-dependent knowledge from case studies is not less valid than general theoretical (and context-independent) knowledge.

Usability evaluations give feedback on user experiences of their workplace. The value of such evaluations for feed forward to new projects or improving existing facilities mainly lies in the ability to understand the user experiences and to translate those into adequate products and solutions. Consequently, the results of research into evaluation needs to be translated into a form which will be quickly and easily accessible to clients, designers, decision makers and others involved in the building process.

Keywords: Usability, Buildings in use, Evaluation methods, Applicability
1. Introduction

Usability evaluations are based on different user’s experiences and assessments on how well workplaces or buildings perform and are related to fulfilment of user needs. The concept of usability is recently adapted to buildings through the work in the CiB W111 Usability of workplaces, but has its origin in relation to applications within product design, information technology and web-design (Alexander, 2004). According to ISO 9241-11, Usability is defined as “the extent to which a system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO, 1998). An important approach to usability is that a product has value only insofar as it is used and whether it contributes to and supports the owners and users activities. By itself a product has no value. In this perspective buildings can be regarded as products or premises to achieve strategic goals such as enhanced productivity, reduced operating costs, increased innovation, knowledge exchange, or attractiveness to clients and qualified workers to mention some. This is also in line with Joroff et al. (1993) who describes the real estate as the fifth resource for production. The approach of looking at buildings as a way to fulfil their strategic objectives and not only as a way to house people and activities is quite new and is supported by work of Becker and Steele (1995), Horgen et al. (1999) and Grantham (2000).

Traditional, most owners and occupants, as well as designers and constructors, seldom perform evaluations of how well their buildings perform related to usability aspects. A systematic evaluation of buildings in use should be an effective way not only to collect information, but more important, to produce knowledge both in relation to the planning of new buildings and for the development and change of existing buildings. According to Jensø et al. (2004), Blakstad et al. (2008) and Blakstad et al. (2010), evaluations should be based on different methods and aspects, depending on objective, purpose, focus, competence and resources. There are already many concepts, definitions and methods that are relevant to a buildings condition, standard and quality. Most of these look at the building as a physical object and do not associate the building with its usability. This view is supported by Alexander (2004) that claims that conventional approaches to building performance often focus on technical, functional, and operational aspects of their use. Testing functionality means making sure that the product works according to given specifications, while more important testing usability means making sure that people can find and work with the functions to meet their demands and needs. To be successful in use, a product’s functionality will therefore be critical, but not sufficient.

Our main perspective is on usability’s relation to organizational goals and output and the effect of the building. This means that methods for evaluation of usability will have to explore the user experience of buildings, the relationship between buildings and multiple users, the process and time, as well as the culture, work style, and goals of the organization. In this paper, we elaborate and reflect on usability evaluations and the applicability of results for other projects. The aim of the study has been to develop and test a methodology that the owners and users of buildings can use to assess usability in their own building portfolio. An important discussion is whether the results or findings from those evaluations can be considered as valid and reliable, and whether
context dependent knowledge from usability evaluations can be feed forwarded to new projects or be generalized and added to a more generally applicable body of knowledge.

2. Research methods and process

The work to apply the usability concept in building design, construction, management and use has been developed through a programme of action research, comprising an intensive series of case studies and associated workshops carried out both within the CiB W111 and in several national case studies representing a variation of companies and disciplines (Alexander et al 2004). Here we present the methodological discussion related to the process of developing and testing different methods to evaluate usability within a common framework named USEtool (Blakstad, 2010). This paper is based mainly on the experience from a number of Norwegian case studies in close collaboration with project partners, each providing a case that has been used for testing and developing methods and tools. Our cases have been university colleges, workplaces (offices) and secondary schools. We have had real users and stakeholders participating such as students, teachers, administration staff, representatives from owner and facilities management staff. The researchers and the project partners have been engaged in participatory workshops to develop the project’s aims and approach to evaluation, as well as to reflect on the results of various tests of methods and tools. The aim of the testing has not been primarily to evaluate the actual workplaces or spaces as such, but to gain experience with use of the methods, tools and indicators as they were developed and refined. Several methods such as interviews, document analysis, structured group interviews, walkthroughs and workshops have been tested in cases and have been included in the USEtool handbook. As a part of the development process, also questionnaires such as ASTM and DQI have been tested. However, questionnaires are not part of the final set of methods in the USEtool handbook.

The presented research is based on an action research or applied research approach (Lewin, 1946). According to Johansson and Lindhult (2008), action research aims at combining research and development, through involvement of practitioners and users. Our work can also be described as a “real world enquiry” focusing on practice with the limitations and challenges this implies (Robson 2002). The validity, reliability and the generalizability of the methods in general will be discussed later in this paper.

3. Theoretical framework

Searching in literature, we find various approaches and understanding of terms associated with the use of buildings and quality of use. According to OECD (2000) an evaluation is defined as a ‘Systematic and objective assessment of an ongoing or completed project, program or policy, its design, implementation and results’. In general, we can distinguish between different approaches in ex post project evaluations. Evaluators, especially those who aim at including a user perspective typically prefer holistic evaluations based on a diverse set of approaches and indicators, typically combinations of quantitative and qualitative evaluations (OECD 2000).
Government agencies and municipalities often are expected to apply a socio-economic perspective, where benefits and costs are being processed to as great extent as can be defended scientifically (Sager 1991, Small 1999), while property investors typically apply the business perspective, which can be supported by a performance measurement approach (Olsson et al 2008).

Since the introduction of Post Occupancy Evaluation in the late 60s, several building evaluation tools have been introduced. According to Preiser et al (1988), a post-occupancy evaluation (POE) is: “the process of evaluating buildings in a systematic and rigorous manner after they have been built and occupied for some time”. Methods within the field of post occupancy evaluation focus on the users and their needs, and ideally include both physical, technical and psychosocial aspects and evaluations. Examples of this are Design Quality Indicators (DQI), ASTM and Buildings Use Studies (Leaman, and Bordass 2001). However, in practice, we have seen that most POE methods have focused on technical aspects, and less on the building’s relation to the users due to the fact that many methods for POE evaluate the building in relation to functional and technical requirements as evaluation parameters. A search in literature on building evaluation gives a number of different evaluation tools that have been developed in order to assess a certain design or building in use. Examples are ASPECT, BREEAN, ST&M, LEED, EPFS etc. Such methods often focus on specific and limited aspects such as building technology, health and safety, working environment or user satisfaction and well-being. Other methods are concerned with evaluation of the direct use of buildings (data of occupancy etc).

According to Steinke et al (2010) there is no industry-accepted definition of building evaluation, nor is there a standardized method for conducting such evaluation. In contrary, as shown in this paper, you find that the different building evaluation methods and tools range from specifically assessing limited aspects to generally assessing whole building design. In a study of Steinke et al (2010) the evaluation tools were categorized according to the dimensions of performance and user groups assessed using the following four categories;

- Service performance. Focus on enhancing the service experience for clients
- Functional performance. Focus on creating a quality work environment for staff
- Physical performance. Focus on physical design and technical performance
- Financial performance. Focus on financial efficiency.

From 17 building evaluation tools in this study, 16 tools covered physical performance, and others covered one or two categories within the same tool. Only one tool covered all the four dimensions simultaneously (Shiem-Shin Then, 2005). An overview of different tools can also be found in Baird et al (1996) and Voordt & Wegen (2005).

One of the interesting methods described, has been BPE, building performance evaluation, first presented by Preiser (1989) in the book Building Evaluation. The methods and focus in BPE
have moved towards a more holistic, process-oriented approach. This means that not only facilities, but also other aspects like organizational, economic, social and political aspects are taken into consideration. In 1997, the POE process model was developed into an Integrative Framework for Building Performance Evaluation (Preiser and Schramm, 1997), followed by other publications such as Learning from our Buildings (Federal Facilities Council, 2001), Improving Building Performance and Evaluating Building Performance in Healthcare Facilities; An organizational Perspective (Steinke et al, 2010). The last based on the established multidimensional framework Balanced Scorecard (Kaplan & Norton, 1996), developed to a conceptual model; The Building Performance Evaluation (BPE) Scorecard. The BPE Scorecard views facilities from four perspectives or performance dimensions, and allows a variety of existing evaluation tools to be incorporated into the methodology.

Chan, Beckman & Lawrence (2007) argues that in despite of a long tradition on studies on organizations, there has been relatively little systematic work linking the built environment with organization theory or vice versa. It seems that organizations rarely see the physical environments as an important factor to their business strategies. On the other hand we find literature that strongly underline the effect the built environment has on several aspects of organizational functioning (Becker, 1981), and the fact that the built environment defines the context in which work processes, services, social interactions and outcomes is taken place (Becker, 1981; Bitner, 1992).

Within the work of the CiB W111 Usability of workplaces, there has been developed a theoretical framework describing the concept of usability, and different methods and tools have been studied and examined in several case studies the last 6 – 8 years (Jensø et al 1994; Hansen et al, 2005; Hansen et al 2006, Alexander, 2004; Blakstad et al, 2008). Alexander (2008) underlines that seen in an organizational context; Buildings usually will be part of a portfolio of buildings and are evaluated in terms of their asset value. He argues that the tools and metrics for considering the use value of buildings are less well developed and understood. This is in line with Granath and Gilliard (2008) that state that “usability cannot be evaluated simply on the product alone but also with respect to how the product is perceived by and interacts with the user”. Fenker (2008) relates usability to user experiences and social relations between users and facilities and describes usability as a process that can only be understood as a social construction where the building act as a sort of stage. According to Fenker, “...the artefacts are bearers of a set of possibilities and constraints as well as, most importantly, activity and social practices models.”

Usability evaluations are based on different user’s experiences and assessments on how well the buildings perform regarding different parameters. A building’s performance can never be seen or understood isolated from an organisational and technical perspective, as those aspects interact and influence each other. Usability has hence a complex nature and can be described as a “wicked problem” Blakstad et al (2008). Such problems are characterized by no definitive formulation of solutions, and they are open to multiple interpretations (Rittel and Webber 1973). According to Blakstad, an adequate approach to “wicked problems” will require multi-method strategies using a triangulation of methods and evaluations with multiple perspectives.
This is in line with findings from previous studies showing that evaluations work best when they are based on several methods and aspects, depending on objective, purpose, focus, competence and resources (Frechtling, 2002). All this implies that usability evaluations are complex, that there is a need for simplification and that the evaluator possesses both theoretical and practical knowledge and skills (Baird et al 1996). Blakstad et al (2008) describes how different methods and tools were explored and tested according to their relevance and validity for usability in several Norwegian cases. As pointed out earlier, few of the available methodologies aim directly at evaluation of usability related to organizational objectives. However, they found that many traditional research and evaluation methods had potential to be developed for the purpose of usability evaluation.

4. Usability evaluations in Norwegian case studies

In a number of Norwegian case studies, several methods and ways of doing usability evaluations have been studied. The purpose of the Norwegian work has been to provide building owners, users and Facility Managers with knowledge of usability in order to support continuous improvements. As a result, a common usability framework or methodology named USEtool has been developed. The recommended process for mapping usability consists of five stages (figure 1). For each stage there is a description of the goals, the methods and tools used and the expected results from each method and stage (Blakstad et al, 2010).

![Figure 1: USEtool framework. The evaluation process (Blakstad et al., 2010)](image)

Methods used in the USEtool methodology have been; Document analysis, interviews, structured group interviews, walkthrough and workshop. The framework reflects the importance of understanding and taking in consideration the contextual conditions that may determine the outcome of the user experiences with the building or workplace (Hansen et al. 2006; Fenker, 2008).

5. Applicability of results from usability evaluations

The purpose of this paper is to discuss important aspects of evaluation of usability, and the applicability of the results for other projects. Usability evaluation within the USEtool context, can be regarded as a methodology (Silverman, 2007) referring to the choices one makes about cases to study, methods of data gathering, forms of data analysis. Most of the methods proposed in USEtool are quantitative methods like interviews, focus groups, document or textual analysis, observation etc. mainly conducted within a case study context. USEtool complements other methods used in more traditional post occupancy evaluations or building performance studies.
Quantitative methods like questionnaires, statistics etc will give some idea of how much, how usual or unusual a given phenomena occurs. Those methods will give some background information, but not necessary any explanation.

Usability evaluations are not only concerned about how, but also on the question of why. They study meanings as well as causes (Hammersley, 1992). In literature we find a number of different methods for evaluation of efficiency both related to buildings and to organizations. We also have quite a few methods for evaluation of effectiveness and satisfaction, but effectiveness is more difficult to assess. Related to our perspective on buildings as means of production, effectiveness will be the most important. Of the methods we have used so far, interviews, participatory methods, and walkthroughs are methods that will be well suited for further development in order to evaluate effectiveness, but we will also need to develop criteria that can be used in quantitative studies such as questionnaires.

The value of usability evaluations for feed forward to new projects mainly lies in the ability to understand the user experiences and to translate those into adequate products and solutions. An important discussion is whether the results or findings from usability evaluations can be considered as valid and reliable, and whether context dependent knowledge from usability evaluations can be fed forward to new projects or be generalized and added to a more generally applicable body of knowledge. The idea behind evidence based design is to have significant number of results from several case studies or project to say something about the validity for other projects (Hamilton et al., 2009). From a usability perspective this is not necessary a universal truth isolated from any given context. Back to the problem of copy and past thinking, rather than learn and translate. This is more in line with Kroll (2005) that underlines that “evidence-based design means not using a cookbook approach to building design”. Every facility is built within a specific set of requirements and constraints, and consequently each requires a tailored approach. The aim should be to intelligently adopt or adapt the research and apply it to new projects. One big discussion about qualitative research is contextual sensitivity (Silverman 2007) and the awareness of the context in understanding the studied phenomenon. The question of reliability and validity is therefore a main issue discussing usability evaluation, user experiences and usability evidence.

In the following we relate the discussion of usability evaluation methods to the questions of reliability, validity and generalizability.

### 5.1 Reliability

Reliability is related to consistency of a measure. Hammersley (1992) defines reliability to the degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions. That means that it deals with replicability or the question of whether the research could be repeated with the same results, interpretations and claims. As argued earlier in this paper, usability is strongly depending on context, situation and perspectives, and hence could not be reliability at all. Moisander and
Valtonen (2006) suggest two ways to improve reliability in non-quantitative work. One is to make the research process transparent, and the other is to pay attention to theoretical transparency. In this type of research, reliability cannot be ensured through large, representative samples of research material. The methods to summarize information may be affected by judgmental subjectivity. To a certain extent, the subjectivity is intended, as we want to map the perceptions of different user representatives. The problem of reliability may therefore be considerable in each of the sub-studies. To compensate for this, several studies have to be made.

5.2 Validity

Validity concerns how well a measure does in fact measure what it is intended to measure. The findings could be valid for other contexts if;

- Statistics enough case studies giving the same findings

- There are some similarities between case studies carried out and a given project.

The question is if usability evaluations and user experiences could be generalized and be valid for other contexts and situations? To address validity in our cases, several evaluation methods are used. This means a method triangulation. The combined use of methods gives a better measure than each of the methods or tools independently. According to Yin (2003), case studies using multiple sources of evidence are generally rated as having a higher quality, compared to those that rely only on single sources of information. In contrary, Flyvbjerg (2004) argues that the key advantage of qualitative research is its ability to give insight into local practices and is important for the development of a nuanced view of reality. According to Flyvbjerg, this is one of the strongest characteristic of qualitative research, the possibility of falsify ability as an excellent way to test the validity of any research finding.

5.3 Generalizability

Generalizability is a main aim in quantitative research and is normally achieved by statistical sampling methods or procedures. In qualitative analysis, generalizability is a very important and debated issue. Generalizability refers to the extent to which findings from a study apply to a wider population or to different contexts. Generalizability can be discussed along two dimensions: generalizability of experiences from the use of the proposed method, and generalizability of the results from the studies. The results from each study may have limited generalizability. The experiences from use of usability evaluation methods are likely to have a higher degree of generalizability than the actual evaluation results, even though both are context dependent. A common argument against case studies and other qualitative studies has been the lack of generalizability, and indirectly against reliability. Seen in isolation, case studies are context dependent. It is typically said that it is up to the reader to judge if a previously conducted case study is relevant in the context he or she studies. However, more generalizable
results may be established through a series of replications and validations. When the number of studies with consistent results grows, the confidence in the findings should increase. As the number of case studies increases, it is possible to perform meta-analyses of the cases, a similar way as meta studies (studies of studies) are common in for example medicine. Especially if the case studies, and other qualitative research, are conducted in a uniform way, the possibilities to observe general trends across a number of studies increases. Through our work with usability, we have developed a common framework for usability evaluations, named USEtool (Blakstad et al, 2010). This framework has been used both in student projects and in real life, giving us a good basis for meta-analysis. Related to our proposed tool and method for evaluating usability, a uniform way of conducting usability evaluations will enable us to perform meta-analyses of our evaluations and hence improve the question of generalizability.

5.4 Experiences

One may always discuss the external validity of qualitative methods. According to Halvorsen (2008) the main question is not if results may be generalized but if knowledge can be transferred to other settings. The validity and reliability of the methods in general can be discussed. One important aim for the development of the methods have been their ability to produce and/or obtain relevant information and experience relating to the defined topics for evaluation, understanding the situation and context, and obtaining differences in interests and opinions, rather than focusing on consensus. In one case we used both qualitative methods and questionnaire in the evaluation. The results were very concurrent and gave the same picture, but the qualitative methods gave a much more richer and useful material to understand why and where when discussing possible strategizes and activities to improve building performance and usability for the users.

From what we have seen in the case studies and tests, the described methods and tools really assess usability within the given context, with special focus on the effectiveness of the facilities and their ability to support value creation in the user organization. We acknowledge the fact that one cannot generalize directly from the results of highly context dependent evaluations such as USEtool. In fact, the Usability concept is context-dependent in nature. One of the case studies showed a university college with a very high score on building performance and coloration between program and completed building, but still showed a lack of usability due to change in pedagogic, increased student number, lack of changing culture among the department (Hansen et al., 2006). In another case we found a high degree of pride and high academic score among the pupils, in spite for a building performing really badly.

One may argue that the contextual knowledge gained from applying the methods described, is as important as the generic results for building performance. The main contribution is the way these methods are combined in a structured framework with process descriptions and easy-to-use guidelines, as well as the operationalized relation to effectiveness and usability. Further testing carried out by our project partners will reveal the method’s usefulness, simplicity, and the necessary amount of resources to carry out evaluations.
6. Acknowledgements

The work has been supported by Sør-Trøndelag County, Statsbygg and Statoil. We also thank users and other stakeholders for active and important participation in the different cases, providing us with necessary feedback and input for developing the USEtool.

References


Granath JÅ, Gilleard J (2008) Usability in the Workplace: Case study of Pamela Youde Eastern Hospital Hong Kong. In Usability of Workplaces, Phase 2. CIB W111 research report.


