

# Three Contexts for Evaluating Organizational Usability

**Morten Hertzum**

Professor

University of Copenhagen

Njalsgade 76, Bldg. 4

Copenhagen

Denmark

[hertzum@hum.ku.dk](mailto:hertzum@hum.ku.dk)

## Abstract

Organizational usability is about the match between the user and the system, between the organization and the system, and between the environment and the system. While the first of these matches can, to a large extent, be evaluated in the lab, the two others cannot. Organizational usability must instead be evaluated *in situ*, that is, while the system is used for real work. We propose three contexts for such evaluation: pilot implementation, technochange, and design in use. Pilot implementation aims to inform the finalization of a system on the basis of testing it in the field prior to go-live.

Technochange focuses on shaking down a system during go-live to realize the benefits it was developed to help achieve. Design in use is the tailoring performed by users after go-live to fit a system and its use to their local and emerging needs. For each evaluation context we describe its aim and scope, provide a brief example, and discuss the challenges it presents. To strengthen the focus on evaluation in the three contexts, we propose the measurement of specified effects, combined with a sensitivity toward emergent effects.

Incorporating effects in the evaluation of organizational usability makes for working systematically toward realizing benefits from the use of a system.

## Keywords

Usability, usability evaluation, organizational usability, usage effects, pilot implementation, technochange, design in use



## Introduction

Common methods for usability evaluation include usability testing (Rubin & Chisnell, 2008), heuristic evaluation (Nielsen, 1994), and cognitive walkthrough (Wharton, Rieman, Lewis, & Polson, 1994). An extensive body of research has investigated these methods with respect to, for example, the relative merits of the methods (Virzi, Sorce, & Herbert, 1993), the types of usability problems detected (Cuomo & Bowen, 1994), the number of users needed (Lewis, 1994), the importance of task differences (Cordes, 2001), the effect of the evaluator (Hertzum, Molich, & Jacobsen, 2014), the reactivity of thinking aloud (Hertzum, Hansen, & Andersen, 2009), the prospects of remote evaluation (McFadden, Hager, Elie, & Blackwell, 2002), and the shortcomings of usability-evaluation research (Gray & Salzman, 1998). However, most of this research has construed the use situation as a single user interacting with the system in a setting unaffected by other users, competing tasks, and the larger organizational context. This lab-like approach leaves out aspects pertinent to the usability of many systems. The present study adopts an organizational definition of usability and discusses its implications for usability evaluation.

Elliott and Kling (1997) defined organizational usability as “the match between a computer system and the structure and practices of an organization, such that the system can be effectively integrated into the work practices of the organization’s members” (p. 1024). In addition, they presented a three-level framework according to which organizational usability encompasses the user-system fit, the organization-system fit, and the environment-system fit. The user-system fit is the level that most resembles the focus of conventional usability evaluation. It is about the match between the user’s psychological characteristics and the system, that is, about the social acceptability of the system and its integrability into the user’s work. The organization-system fit is the match between the characteristics of the organization and the attributes of the system. Key characteristics of the organization are its structure, power distribution, institutional norms, and the social organization of computing. The environment-system fit is the match between the environment of the organization and the attributes of the system. Elliott and Kling (1997) emphasized two aspects of this level: the environment structure and the home/work ecology. Organizational usability sensitizes the usability practitioner to a different, but no less relevant, set of issues than, for example, the ISO 9241 (2010) definition of usability. The difference shows the potential for the usability practitioner of shifting between different conceptions of usability, and it points toward different methods of usability evaluation (Hertzum, 2010). Organizational usability has similarities with macroergonomics (Hendrick, 2005; Kleiner, 2006) and brings forth issues that go well beyond what can be evaluated in lab-like settings. Thus, the starting point of this study is that organizational usability necessitates in-situ evaluation.

While we acknowledge that some previous work has addressed organizational usability, we contend that such work is vastly under represented. Taking the *Journal of Usability Studies* as an example, less than 10% of the research papers in the first 12 volumes of the journal (2005-2017) involve in-situ evaluation and only about half of these papers are about organizational aspects of usability (e.g., Pan, Komandur, & Finken, 2015; Selker, Rosenzweig, & Pandolfo, 2006; Smelcer, Miller-Jacobs, & Kantrovich, 2009). In addition, there has been considerable research discussion over the past decade about whether in-situ evaluation is worth the hassle (Kjeldskov & Skov, 2014). Several studies in this discussion report that properly conducted lab evaluations identify essentially the same usability problems as evaluations in the field (e.g., Barnard, Yi, Jacko, & Sears, 2005; Kaikkonen, Kallio, Kekäläinen, Kankainen, & Cankar, 2005; Kjeldskov, Skov, Als, & Høegh, 2004). However, the discussion also reveals widely different views on what it means to set an evaluation in the field (Kjeldskov & Skov, 2014) and surprisingly little consideration of how the definition of usability may influence whether usability problems can or cannot be identified in the lab. The present study addresses these two issues by explicitly adopting an organizational definition of usability and by proposing three contexts for evaluating it in the field. Thereby, we aim to illustrate the prospects and challenges of in-situ usability evaluation.

In this paper, we present three contexts for the in-situ evaluation of organizational usability: pilot implementation, technochange, and design in use. Table 1 provides an initial overview of the three evaluation contexts that cover the period before, during, and after go-live,

respectively. Compared to the two other contexts, design in use may to a larger extent be an opportunity for working with organizational usability than a genuine opportunity for evaluating it. To strengthen the evaluation element in all three contexts, we propose the specification and measurement of usage effects. This proposal is elaborated in the discussion after the evaluation contexts have been presented. We conclude the paper by summarizing takeaways for usability practitioners.

**Table 1.** Three Contexts for Evaluating Organizational Usability

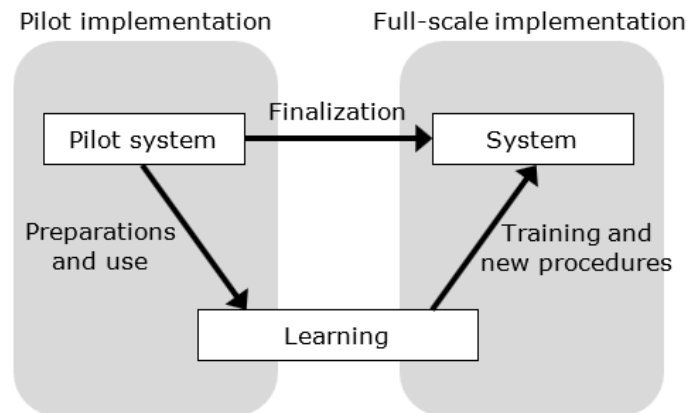
	<b>Pilot implementation</b>	<b>Technochange</b>	<b>Design in use</b>
Time	Before go-live	During go-live	After go-live
Purpose	To learn about the fit between the system and its context in order to explore the value of the system, improve its design, and reduce implementation risk	To transition from old practices to the new system and start realizing the benefits that motivated its introduction	To appropriate—tailor—the system for local and emergent needs when opportunities for such appropriation are seen and seized
People	Usability practitioners plan and facilitate the process	Usability practitioners facilitate the process	Users drive the process, facilitated by usability practitioners
Setting	In the field, i.e. during real work, but limited to a pilot site	In the field but while the system is still new and unfamiliar	In the field during regular use of the system for real work
System	Pilot system, i.e. a properly engineered yet unfinished system	Finished system but not yet error-free and not yet fully configured	Finished system, yet with possibilities for reconfiguration
Process	Used in situ for a limited period of time and with special precautions against errors	Used in situ by inexperienced users, possibly with extra support during go-live	Used in situ by regular users, some of whom occasionally engage in design-in-use activities
Benefit focus	Specified benefits dominate; other benefits may emerge	Specified benefits dominate; other benefits may emerge	Emergent benefits are likely to dominate
Duration	Temporary, typically weeks or months	Temporary, typically months	Continuous, typically years or decades
Main challenge	Boundary between pilot site and organization at large	Premature congealment of the process to ensure benefit from new system	Insufficient capability to make or disseminate changes

### **Pilot Implementation**

A pilot implementation is “a field test of a properly engineered, yet unfinished system in its intended environment, using real data, and aiming – through real-use experience – to explore the value of the system, improve or assess its design, and reduce implementation risk” (Hertzum, Bansler, Havn, & Simonsen, 2012, p. 314). Contrary to prototypes, which are often evaluated in the lab, pilot systems are sufficiently complete to enable in-situ evaluation. However, their design has not yet been finalized; rather the pilot implementation generates feedback that informs the finalization of the system and the conduct of its subsequent full-scale implementation at go-live. Pilot implementations are restricted in both scope and time. That is, one or a few sites are selected for the pilot implementation, and these pilot sites use the system for a specified period of time. In this way, change is managed by trying out the system on a small scale prior to full-scale implementation.

According to Hertzum et al. (2012), a pilot implementation consists of five activities: planning and design, technical configuration, organizational adaptation, pilot use, and learning. The first three activities are preparations. During the preparations, the focus of the pilot implementation is defined, the boundary between the pilot site and the organization at large is determined, the

system is configured for the pilot site, interfaces to other systems are established, work procedures at the pilot site are adjusted, users receive training, safeguards against breakdowns are set up, and so forth. The extent and complexity of the preparations demonstrate that a pilot implementation is not just the period of pilot use. During the period of pilot use, the staff at the pilot site has the double task of conducting their work with the system and providing input to the evaluation. The final activity, learning, is the reason for conducting pilot implementations, as illustrated in Figure 1. Importantly, learning about the organizational usability of the system occurs during the preparations as well as during the period of pilot use.



**Figure 1.** Pilot implementation.

As an example, Hertzum et al. (2017) reported on the pilot implementation of an electronic ambulance record in one of the healthcare regions in Denmark. The main purpose of the pilot implementation was to evaluate the match between the regional pre-hospital services, especially the paramedics' work, and the electronic ambulance record. For this purpose, a pilot version of the ambulance record was installed in 17 ambulances and used for all their acute dispatches. With respect to the user-system fit, the pilot implementation immediately revealed problems with the user interface of the ambulance record. Data entry was divided onto more than 20 screens, thereby degrading the paramedics' overview of what information they had already entered and what information they still needed to enter. This problem illustrates that the user-system fit resembles the focus of conventional usability evaluation; the problem could probably, and preferably, have been uncovered in the lab prior to the pilot implementation. With respect to the organization-system fit, the pilot implementation, for example, revealed ambiguity about whether the ambulance record was primarily intended to provide the clinicians at the emergency departments with better information about the patients en route to the hospital or to provide a better record of the paramedics' treatment of the patients for documentation purposes. A concrete manifestation of this issue was when problems occurred during the handover of the patients from the paramedics to the emergency departments because the ambulance record had to be printed and this turned out to be an exceedingly slow process. Consequently, the handovers were often mainly oral. This defeated any positive effect of the ambulance record on the handover and gave the paramedics the impression that their work was considered secondary to that of the emergency-department clinicians. Throughout the pilot implementation it remained unclear whether and, if so, when, the ambulance record would be integrated with the electronic patient record at the hospitals and, thereby, eliminate the need for printing the ambulance record. With respect to the environment-system fit, the pilot implementation served the supplementary purpose of collecting data about the paramedics' work in order to show that a regional decision to remove physicians from the ambulances had no adverse consequences for patients. For the ambulance record to be usable at this level meant feeding into a political process between the region responsible for providing healthcare services and the public relying on these services in the event of illness and injury. It is difficult to imagine that problems with the handover of the patients to the emergency department (organization-system fit) or the removal of physicians from the ambulances (environment-

system fit) could have been evaluated in the lab; these problems are about contextual consequences that do not surface until the ambulance record is used in situ.

The main challenges of pilot implementation concern the boundary between the pilot site and the organization at large. First, it is nontrivial to decide the scope of a pilot implementation because most systems influence the interactions among multiple organizational units. Keeping the pilot site small excludes many of these relevant interactions from the pilot implementation. Sizing up the pilot site includes more interactions but also increases costs in terms of, for example, the number of units to enroll, users to train, work procedures to revise, and errors to safeguard against. Second, when the scope has been decided, the interactions that cross the boundary between the pilot site and the rest of the organization must be handled. This may involve developing technical interfaces for dynamically migrating data between the pilot system and the existing system or introducing temporary manual procedures to interface between the systems. Hertzum et al. (2012) described a pilot implementation in which the interactions that would be electronically supported when the system was fully implemented were simulated in a Wizard-of-Oz manner during the pilot implementation. This way the users got a more realistic impression of the system functionality, but the resources required to run the Wizard-of-Oz simulation 24 hours a day meant that the pilot implementation was restricted to a five-day period. Third, the presence of the boundary will likely introduce some difficulty in telling the particulars of the pilot implementation from generic insights about the system. This difficulty is aggravated by the unfinishedness of the pilot system. If particulars are mistaken for generic insights, or vice versa, then confusion and faulty conclusions will ensue. On this basis Hertzum et al. (2017) concluded that learning from pilot implementations is situated and messy. Finally, the learning objective may become secondary to concerns about getting the daily work done. Pilot implementations involve using the system for real work. While this realism creates the possibilities for evaluating organizational usability, it also incurs the risk that the users focus on their work to the extent of not devoting time to incorporate the pilot system in their routines, not reporting problems they experience, or otherwise not contributing fully to the process of learning about the system.

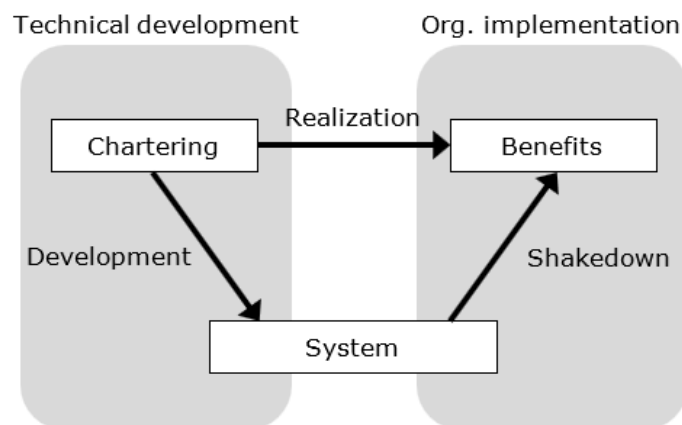
## **Technochange**

Deliberate technochange is “the use of IT to drive improvements in organizational performance” (Markus, 2004, p. 19). That is, technochange combines IT development, which on its own would be purely about technical matters, and organizational change, which on its own would not be driven by IT. Because IT development and organizational change tend to reside on either side of go-live, their combination in technochange has go-live as its focal point. In technochange the idea that motivated the IT development must be brought forward into organizational change and used as a yardstick against which to evaluate the performance improvement obtained by using the system. This requires a sustained focus on the idea during and immediately following go-live. It is through the sustained focus on the idea that change becomes managed rather than messy and unlikely to happen. By linking IT development to organizational improvements, technochange assigns organizational usability a key role in change management.

According to Markus (2004) technochange has four phases: chartering, project, shakedown, and benefit capture. Figure 2 illustrates the process. During chartering the technochange idea is proposed and approved. The project phase could also be termed development; it is the phase during which the technochange system is designed and developed. Development ends when the system goes live. Go-live also marks the start of the shakedown phase. During shakedown the organization starts using the new system and troubleshoots the problems associated with the system and the new ways of working. The goal of this phase is “normal operations”; achieving that goal requires a good match between the system and all three levels of organizational usability. Insufficient attention to shakedown incurs the risk that work practices will congeal before normal operations have been reached, resulting instead in mismatches between the system and work practices. During the final phase, benefit capture, the organization derives benefits from the new ways of working through its day-to-day use of the system. If shakedown was unsuccessful then this phase may never occur or the benefits may only be partially realized.

As an example, Aarts and Berg (2006) analyzed the implementation of a computerized physician order entry (CPOE) system in a Dutch regional hospital. Prime examples of the

planned benefits of the system were improved completeness and legibility of the medical orders. Because these benefits mainly concerned the physicians' work, the implementation of the system was aimed at the physicians. After go-live, the system was gradually introduced in all clinical departments. The gradual introduction meant that experiences from the shakedown of the system in early departments could be utilized in its shakedown in subsequent departments. With respect to the user-system fit, the physicians resisted the CPOE system because they perceived order entry as clerical work not to be performed by physicians. They were used to signing medical orders on paper or giving them verbally to nurses. Efforts to have the physicians adopt the system failed, including efforts by the hospital board of directors. With respect to the organization-system fit, the physicians' rejection of the system was a major problem. However, the nurses, unexpectedly, adopted the system to document nursing care because they found the system useful in their work. The nurses' use of the system included entering information about medical orders. Medical orders had to be authorized by a physician and the CPOE system enforced this rule by prompting the nurses for information about the physician responsible for each medical order. The prompting complicated the nurses' use of the system because it presumed a workflow in which the physicians entered their orders in the system, contrary to how the system was used in practice. When it became apparent that the physicians would never adopt the system, it was revised to accommodate the nurses' use of it. Nurses would enter medical orders without being prompted, and it was agreed that the physicians could authorize the orders by signing printouts of them. This way the physicians' ways of working remained virtually unchanged, and the system was rendered usable because the nurses stepped in as its actual users. With respect to the environment-system fit, it remained an issue that about 60% of the medical orders were not authorized even after the physicians could perform the authorization by signing printouts. That is, the hospital did not formally provide the quality of service it was required to provide to the citizens in its catchment area. As in the pilot-implementation example, it is possible that the user-system fit could to a larger extent have been evaluated in the lab prior to go-live. The events that followed after the physicians' rejection of the CPOE system could, however, not have been uncovered in the lab because they were contrary to the planned benefits and did not emerge until the physicians' rejection of the system provided the nurses with an opportunity they perceived as attractive.



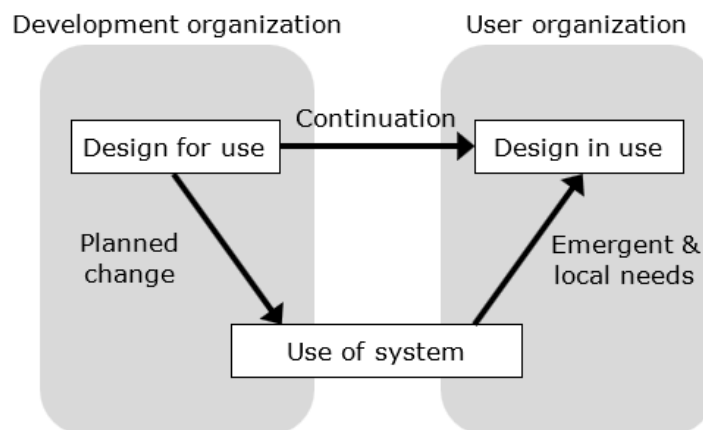
**Figure 2.** Technochange.

The main challenge in technochange is that shakedown stops short of accomplishing the changes necessary to derive benefit from the new system. First, shakedown is situated at a critical point in the process. In the period leading up to shakedown, the IT vendor has been responsible for progress; now the development phase has ended and with it the vendor's responsibility for driving the process. Instead, the user organization assumes responsibility for the process. This shift may create a reorientation or loss of momentum at a point in time where it is important to respond clearly and quickly to problems and user reactions. Second, some groups of users may be unconvinced that the changes associated with the system are beneficial

or they may be fully occupied with their primary work and therefore lacking the time to work out how well-known tasks are to be performed with the new system. These users will likely be reluctant to abandon old ways of working, and when they find a way to accomplish their tasks with the new system, they will be unlikely to explore it further in search of smarter ways of working. Tyre and Orlikowski (1994) found that work practices with a new system tend to congeal quickly, leaving a brief window of opportunity for shakedown. Third, while the problems that emerge during shakedown may be obvious once experienced, their solution may not be obvious. If the troubleshooting becomes a prolonged process, the users will need a temporary solution while they, for example, wait for a system upgrade that fixes the error. The temporary solution may lower benefits realization and be difficult to unlearn when it, at some point, is no longer needed because the system has been upgraded. Finally, shakedown becomes a less directed effort if the pursued benefits are unclear. The benefits should ideally be specified during chartering; leaving them unclear exports their clarification to later phases. Without clarity about the pursued benefits the entire technochange process may be wasted because obtainable benefits are not systematically pursued and, as a consequence, remain fully or partly unrealized.

### Design in Use

Design in use encompasses the “practices of interpretation, appropriation, assembly, tailoring and further development of computer support in what is normally regarded as deployment or use” (Dittrich, Eriksén, & Hansson, 2002, p. 125). This definition emphasizes that the boundary between design and use is permeable in that design, in some form, continues during use, see Figure 3. The continuation of design during use is a user-driven process but may, to varying extents, be organizationally supported. Users may, for example, revise their ways of working as they gradually learn how best to utilize the system or work around its shortcomings, they may configure the system for local needs, and they may share revised practices and system reconfigurations. Over time such design-in-use activities may transform the users’ work and have a substantial impact on the organizational usability of a system. Due to the pervasive and emergent character of design-in-use activities, Orlikowski (1996) described change management as a predominantly improvisational process of responding to these activities in an opportunity-based manner.



**Figure 3.** Design in use.

According to Hartswood et al. (2002) design in use involves four activities: being there, lightweight design, picking up on small problems, and committing to a long-term engagement. Being there implies that IT facilitators, such as usability practitioners, should spend time in the user organization after the system has gone live. The aim of this activity is to achieve a prolonged dialogue in which users and usability practitioners can spontaneously shift their attention between design and use. Lightweight design and picking up on small problems imply a

here-and-now focus. That is, design in use should move smoothly and quickly between using a system to accomplish work, grappling with the problems of applying the system, adjusting the system or its use, and evaluating the outcome of the adjustments. The here-and-now focus ensures that design in use stays relevant to the current needs of local users. Finally, the commitment to a long-term engagement is necessary because the lifetime of most systems is years or decades. To remain usable throughout this period, systems must continue to evolve with the changes in user needs and organizational context.

As an example, Aanestad et al. (2017) described the design in use of videoconferencing at a Norwegian rehabilitation hospital. When videoconferencing went live at the hospital, it was to support administrative and educational tasks internal to the hospital, which was spread across multiple locations. Over an eight-year period, the use of videoconferencing has gradually embraced the follow-up activities with patients. Throughout this period the hospital has relied completely on off-the-shelf technologies. The use of videoconferencing has evolved by facilitating clinicians in their experimentation with new uses of videoconferencing, turning successful experiments into services for routine use across the hospital, and assembling and evolving an up-to-date technical infrastructure for these services. With respect to the user-system fit, considerable resources have gone into establishing ready-to-use videoconferencing rooms and writing user guides to make the videoconferencing services usable for clinicians beyond those with a special interest in technology. With respect to the organization-system fit, hospital management has expressed its long-term commitment to videoconferencing by assigning an annual budget to the expansion of the technical videoconferencing infrastructure and by ensuring that each department has a team coordinator who is the go-to person for issues about videoconferencing. An example of the continuous design-in-use activities concerned an isolation patient who wished to participate in patient teaching sessions. The clinicians arranged for the patient to have an iPad with a videoconference connection to the teaching room. Later, the patient also used the videoconference connection to communicate with the clinicians outside of the isolation room. This solution has potential to become a routine service because still more patients have conditions that require isolation. The largest design-in-use initiatives have, however, concerned the environment-system fit. Videoconferencing has been introduced in the assessment meetings with the local care workers in the patients' municipality and, lately, in the recurrent follow-up meetings with patients after they have returned to their homes. Follow-up meetings with patients via videoconference are especially attractive because many of the patients have reduced mobility; thus, even short-distance travel is exhausting. A cost-benefit analysis of the videoconferencing services showed a net benefit for society, although not for the hospital because reimbursement fees were not defined for these services. It would have been impossible to evaluate the organizational usability of the videoconferencing services in the lab prior to go-live because they did not exist at the time. Rather, the videoconferencing infrastructure and services evolved over a multi-year period when opportunities emerged during local use and led to design-in-use activities.

The main challenge of design in use is insufficient competences in making or disseminating changes. First, local users will often lack the technical knowledge or inclination necessary to experiment with new ways of utilizing a system. Thus, they are dependent on IT facilitators being there. While such facilitators were present in the videoconferencing case, they are often not. Multiple studies propose the introduction of IT facilitators as a much needed way of supporting design in use (e.g., Dittrich et al., 2002). Second, design-in-use activities may improve the usability of a system considerably for the users involved in the activities, but unless their solution is spread to other users in the organization, it will be of limited organizational value. Sometimes idiosyncratic hacks diffuse widely but most often organizational dissemination requires that solutions are developed into operational services. Usability practitioners could make valuable contributions by evaluating the potential of design-in-use solutions to become services that improve organizational usability. Finally, the substantial duration of design in use—years to decades—makes it important but also difficult to maintain some sort of direction in the continued design of a system. Aanestad et al. (2017) recommended distinguishing between time-boxed efforts with specified goals and long-term commitments with room for evolution. Relatedly, they recommended balancing here-and-now considerations against later-and-larger considerations in decisions about the direction and evolution of design-in-use work. These recommendations appear to presuppose periodic evaluations, formal or informal, of the current match between the system and its use as well as between the realized and pursued benefits.



## Discussion

In-situ evaluation occurs when a system has started to affect the users' daily lives and require them to change their ways of working. Wagner and Piccoli (2007) contended that design projects do not really become salient to users until this point in time. This contention suggests that it makes more sense for users to engage in pilot implementation, technochange shakedown, and design in use than in lab-based usability tests. For example, it was when the CPOE system started to affect the physicians' daily work that they rejected to use it. Up to that point they might have expressed reservations toward the system but until it was deployed in situ it remained unknown, probably also to the physicians, that they had the resolve and organizational power to reject the system in spite of hospital management's efforts to the contrary. In addition, it is important to keep in mind that the users' primary interest will likely be to get their work done with the new system rather than to evaluate its organizational usability. To instill a focus on evaluation during pilot implementation, technochange, and design in use, we propose specifying the benefits pursued with the introduction of the new system and, subsequently, evaluating whether they are realized. If they are not, yet, realized then work is needed to revise the system, change the work practices, or both—followed by a renewed evaluation.

A focus on benefits and their realization differs from conventional usability evaluation in at least two ways. First, evaluating whether specified benefits are realized is different from evaluating systems to identify usability problems. The former implies an up-front effort to formulate the pursued benefits; the latter relies on uncovering problems post hoc. An advantage of specifying benefits early on is that once specified they provide a means for guiding the subsequent process. Second, a benefit is an improvement of the work system. A work system comprises people working together and interacting with technology in an organizational setting to make products or provide services. The focus on the work system follows from the macroergonomic scope of organizational usability. Pursued benefits may, for example, concern the efficiency and profitability of the work system, the quality of its products or services, its innovative capacity, and the work-life quality for the people in the work system (Kleiner, 2006). The changes required to achieve a benefit may involve revision of procedures, collaborative practices, organizational norms, or the division of labor in addition to changes in the technological systems.

Elliott and Kling's (1997) definition of organizational usability talks explicitly about the effective integration of the system into the work practices and is, thereby, in line with a focus on benefits. Benefits realization is a strong indicator of effective integration. At the user-system level, specified benefits resemble the usability specifications proposed by Whiteside et al. (1988). For the initial use of a conferencing system, they, for example, specified an effect stating that the user should be able to successfully complete three to four interactions in 30 minutes. This effect could be assessed by observing a sample of users when they start using the system for performing their work. At the organization-system level, Granlien and Hertzum (2009) analyzed an intervention to pursue the effect that all information about the patients' medication was recorded in the electronic medication record of a hospital. The realization of this effect was assessed through audits of the medical records to ascertain whether they contained violations of the requirement to record all medication information in the electronic medication system. By repeated audits at monthly intervals, the intervention sought to improve the cross-disciplinary collaboration between physicians and nurses, who previously documented their work in discipline-specific systems. At the environment-system level, the electronic medication record served to realize the effect that the right medication was given to the right patient at the right time. This effect states an, otherwise implicit, expectation of the patients and general public toward the hospital that introduced the electronic medication record.

The specified effects should capture the purpose of a system in terms that are both measurable and meaningful to the organization (Hertzum & Simonsen, 2011). For pilot implementation, the effects should be specified during the preparations, for technochange during chartering. For design in use, they may not be specified until during use. The users who do the work necessary to realize an effect may differ from those who benefit from its realization (Grudin, 1994). This asymmetry is a result of the division of labor in the organization and most visible at the organization-system and environment-system levels; it is a defining characteristic of organizational usability. Because the benefits associated with a system may be specific to user

groups, it is important to involve all user groups in the specification of the effects to be pursued. Along with the effect itself, the specification should describe how the effect is to be measured. Effects measurement may require the capture of some additional information during use; the means of capturing this information must be set up. The specification and measurement of effects provide for evaluating organizational usability in a manner that can guide system finalization (pilot implementation), benefits realization (technochange), and continued design (design in use). However, evaluation may not merely reveal whether specified effects have slipped or been realized; evaluation may also lead to the recognition that a specified effect is not attractive after all or to the identification of hitherto unspecified, yet attractive, effects (Hertzum & Simonsen, 2011). It is important that the evaluation remains sensitive to such emergent insights and effects, which may surface because the evaluation is performed in situ. While the measurement of specified effects can, and should, be preplanned, the identification of emergent effects requires an open-ended approach that is more about being there—not least in the sense of being attentive. In macroergonomic terms, this means that the work with effects should combine top-down and bottom-up approaches (Kleiner, 2006). With its focus on shakedown, technochange is mainly a top-down approach to change management. In contrast, emergent effects are likely to dominate during design in use, thereby suggesting bottom-up approaches—possibly followed by top-down efforts to disseminate effects that emerged locally to the organization at large. Pilot implementation may be the evaluation context most suited to a combination of top-down and bottom-up approaches.

Pilot implementation, technochange, and design in use are three distinct contexts for evaluating organizational usability. Each of these contexts frames the evaluation differently and provides the usability practitioner with different possibilities. Consequently, the choice of evaluation context matters and the characteristics of the context in which an evaluation is conducted should be kept in mind. Otherwise a pilot implementation may, for example, drift into technochange without clearly appreciating the possibility to cancel or postpone full-scale implementation after the pilot implementation if the evaluation yields negative results or shows that considerable work is needed to finalize the system. Or shakedown may extend into design in use without clearly recognizing that the benefits that motivated the introduction of the system are not being realized; instead emergent effects are being pursued in an opportunity-based manner. These examples also show that not heeding the results of evaluating organizational usability will cause the export of problems from pilot implementation and technochange toward design in use. From an optimistic point of view, this shows the importance of pilot implementation and technochange as contexts for evaluating organizational usability—to avoid exporting problems to later stages. From a more cynical point of view, it may rather show the need for evaluating organizational usability during design in use—to capture problems exported from earlier stages.

## Conclusion

The way in which usability is defined influences whether usability evaluation can or cannot be adequately performed in the lab. Organizational usability calls for embracing the hassle of in-situ evaluation in order to assess the match between a system and its surroundings at three interrelated levels: user, organization, and environment. Thereby, the organizational definition of usability sensitizes the usability practitioner to new aspects of usability and requires knowledge of contexts for in-situ usability evaluation. As contexts for evaluating organizational usability in situ, we propose pilot implementation, technochange, and design in use. We also propose to incorporate the specification and measurement of usage effects in these three evaluation contexts. Collectively, the three contexts provide for evaluating organizational usability before, during, and after a system goes live. The main challenges involved in working with organizational usability are managing the boundary between the pilot site and the organization at large (pilot implementation), avoiding premature congealment of the shakedown process (technochange), and having insufficient competences in making or disseminating change (design in use).

We would welcome more research on the evaluation of organizational usability to answer questions such as: How can the challenges be countered? What are the concrete steps involved in evaluating usability in each of the contexts? How should prespecified effects be balanced against emergent effects? How long must an evaluation be to assess the match between

organization and system? Under what conditions is one of the evaluation contexts to be preferred over the two others? A lot of research is still outstanding before the evaluation of organizational usability is as well-described as conventional usability tests.

### Tips for Usability Practitioners

To evaluate organizational usability, it is necessary to leave the laboratory. This study presents three contexts for the in-situ evaluation of organizational usability. Usability practitioners should consider the following:

- Exploit pilot implementation, technochange, and design in use as opportunities for in-situ usability evaluation. Evaluating organizational usability early (i.e., during pilot implementation) reduces the risk of exporting problems to after go-live.
- Evaluate all three levels of organizational usability. Leaving out, say, the environment-system fit amounts to the risky assumption that problems at this level are inconsequential or will also surface at one of the two other levels.
- Include the specification and measurement of effects in evaluations. This requires knowledge of the practices and performance prior to the new system to be able to set targets. Assessing whether specifications are met is understood by project managers.
- Be sensitive to emergent effects. They may reveal a need for redirecting efforts to retain a positive emergent effect or for redefining planned effects to avoid their negative side effects. Sensitivity to emergent effects requires a bottom-up approach.

Finally, we encourage usability practitioners who have experience with evaluating organizational usability to consider reporting their experiences in *Journal of Usability Studies* or similar outlets.

### References

- Aanestad, M., Driveklepp, A. M., Sørli, H., & Hertzum, M. (2017). Participatory continuing design: "Living with" videoconferencing in rehabilitation. In A. M. Kanstrup, A. Bygholm, P. Bertelsen and C. Nøhr (Eds.), *Participatory Design and Health Information Technology* (pp. 45–59). Amsterdam: IOS Press.
- Aarts, J., & Berg, M. (2006). Same systems, different outcomes: Comparing the implementation of computerized physician order entry in two Dutch hospitals. *Methods of Information in Medicine*, 45(1), 53–61.
- Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2005). An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies*, 62(4), 487–520.
- Cordes, R. E. (2001). Task-selection bias: A case for user-defined tasks. *International Journal of Human-Computer Interaction*, 13(4), 411–419.
- Cuomo, D. L., & Bowen, C. D. (1994). Understanding usability issues addressed by three user-system interface evaluation techniques. *Interacting with Computers*, 6(1), 86–108.
- Dittrich, Y., Eriksén, S., & Hansson, C. (2002). PD in the wild: Evolving practices of design in use. In T. Binder, J. Gregory and I. Wagner (Eds.), *PDC2002: Proceedings of the Seventh Conference on Participatory Design* (pp. 124–134). Palo Alto, CA: CPSR.
- Elliott, M., & 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.
- Granlien, M. S., & Hertzum, M. (2009). Implementing new ways of working: Interventions and their effect on the use of an electronic medication record. In *Proceedings of the GROUP 2009 Conference on Supporting Group Work* (pp. 321–330). New York: ACM Press.
- Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13(3), 203–261.
- Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1), 92–105.

- Hartswood, M., Proctor, R., Slack, R., Voss, A., Büscher, M., Rouncefield, M., & Rouchy, P. (2002). Co-realisation: Towards a principled synthesis of ethnomethodology and participatory design. *Scandinavian Journal of Information Systems*, *14*(2), 9–30.
- Hendrick, H. W. (2005). Macroergonomic methods. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas and H. Hendrick (Eds.), *Handbook of Human Factors and Ergonomics Methods* (pp. 75:1–75:4). Boca Raton, FL: CRC Press.
- Hertzum, M. (2010). Images of usability. *International Journal of Human-Computer Interaction*, *26*(6), 567–600.
- Hertzum, M., Bansler, J. P., Havn, E., & Simonsen, J. (2012). Pilot implementation: Learning from field tests in IS development. *Communications of the Association for Information Systems*, *30*(1), 313–328.
- Hertzum, M., Hansen, K. D., & Andersen, H. H. K. (2009). Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, *28*(2), 165–181.
- Hertzum, M., Manikas, M. I., & Torkilsheyggi, A. (2017). Grappling with the future: The messiness of pilot implementation in information systems design. *Health Informatics Journal*, (online first), 1–17.
- Hertzum, M., Molich, R., & Jacobsen, N. E. (2014). What you get is what you see: Revisiting the evaluator effect in usability tests. *Behaviour & Information Technology*, *33*(2), 143–161.
- Hertzum, M., & Simonsen, J. (2011). Effects-driven IT development: Specifying, realizing, and assessing usage effects. *Scandinavian Journal of Information Systems*, *23*(1), 3–28.
- ISO 9241. (2010). *Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems*. Geneva, CH: International Standard Organization.
- Kaikkonen, A., Kallio, T., Kekäläinen, A., Kankainen, A., & Cankar, M. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies*, *1*(1), 4–17.
- Kjeldskov, J., & Skov, M. B. (2014). Was it worth the hassle? Ten years of mobile HCI research discussions on lab and field evaluations. In *Proceedings of the MobileHCI 2014 Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 43–52). New York: ACM Press.
- Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of the MobileHCI 2004 International Conference on Mobile Human-Computer Interaction* (Vol. LNCS 3160, pp. 61–73). Berlin: Springer.
- Kleiner, B. M. (2006). Macroergonomics: Analysis and design of work systems. *Applied Ergonomics*, *37*(1), 81–89.
- Lewis, J. R. (1994). Sample sizes for usability studies: Additional considerations. *Human Factors*, *36*(2), 368–378.
- Markus, M. L. (2004). Technochange management: Using IT to drive organizational change. *Journal of Information Technology*, *19*(1), 4–20.
- McFadden, E., Hager, D. R., Elie, C. J., & Blackwell, J. M. (2002). Remote usability evaluation: Overview and case studies. *International Journal of Human-Computer Interaction*, *14*(3&4), 489–502.
- Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen and R. L. Mack (Eds.), *Usability Inspection Methods* (pp. 25–62). New York: Wiley.
- Orlikowski, W. J. (1996). Improvising organizational transformation over time: A situated change perspective. *Information Systems Research*, *7*(1), 63–92.
- Pan, Y., Komandur, S., & Finken, S. (2015). Complex systems, cooperative work, and usability. *Journal of Usability Studies*, *10*(3), 100–112.

- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests* (2nd ed.). Indianapolis, IN: Wiley.
- Selker, T., Rosenzweig, E., & Pandolfo, A. (2006). A methodology for testing voting systems. *Journal of Usability Studies*, 2(1), 7–21.
- Smelcer, J. B., Miller-Jacobs, H., & Kantrovich, L. (2009). Usability of electronic medical records. *Journal of Usability Studies*, 4(2), 70–84.
- Tyre, M. J., & Orlikowski, W. J. (1994). Windows of opportunity: Temporal patterns of technological adaptation in organizations. *Organization Science*, 5(1), 98–118.
- Virzi, R. A., Sorce, J. F., & Herbert, L. B. (1993). A comparison of three usability evaluation methods: Heuristic, think-aloud, and performance testing. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 309–313). Santa Monica, CA: HFES.
- Wagner, E. L., & Piccoli, G. (2007). Moving beyond user participation to achieve successful IS design. *Communications of the ACM*, 50(12), 51–55.
- Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In J. Nielsen and R. L. Mack (Eds.), *Usability Inspection Methods* (pp. 105–140). New York: Wiley.
- Whiteside, J., Bennett, J., & Holtzblatt, K. (1988). Usability engineering: Our experience and evolution. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 791–817). Amsterdam: Elsevier.

### About the Author



#### **Morten Hertzum**

Dr. Hertzum is a professor of Information Science at University of Copenhagen. His research interests include human-computer interaction, computer supported cooperative work, information seeking, and healthcare informatics. He is co-editor of the book *Situated Design Methods* (MIT Press, 2014) and has published a series of papers about usability evaluation methods.