ORIGINAL ARTICLE

Systems usability framework for evaluating tools in safety–critical work

Paula Savioja · Leena Norros

Abstract This paper presents a methodology with which evaluation of control rooms in safety–critical industry has been carried out. The results of two cases of evaluation in nuclear power plants (NPP) are presented to elaborate the method thoroughly. The evaluation methodology draws from conventions of human factors and usability evaluation methods but adds an activity theoretical viewpoint by considering different control room functions in a purposeful activity system. A control room has three functions in an activity: instrument, psychological tool, and a communicative tool. The introduced evaluation approach allows understanding of quality of the control room on the level of outcomes it produces, work practices that it consolidates, and the user experience it creates. The evaluation approach is demonstrated by presenting baseline evaluations of two hybrid NPP control rooms that were carried out before major upgrade projects were implemented. Hybrid control room refers to a state in which modern and original hardwired technological solutions are in use simultaneously.

Keywords Control room · Evaluation · Activity theory · Contextual approach · Process control

1 Introduction

Main control room of an industrial plant is the culmination point in which the humans and the technologies function jointly to achieve and maintain the objectives of the respective process. In a good control room, the process control activity, including collaboration of and interaction between the humans and the technology, is purposeful and smooth: Both planned and unexpected process phenomena are addressed with adequate effective responses enabling the achievement of the general operational objectives of safety, production, and health (Vicente 1999). The operating crew performs well, the way of achieving outcomes is sustainable, and the operators feel that the tools are appropriate for the work and support relevant aspects of it.

Unfortunately, control rooms are not always appropriate as described above. There might be problems in many levels of the work system starting from the organization and culture. This paper addresses the issue of *evaluation* of the appropriateness of *tools* in process control work. The main tool is the control room comprising of automation and control system user interfaces (UI), procedures and the entire physical environment. Poor design of control rooms is one of the latent conditions acknowledged by the modern safety research to lead to threats to system safety and finally accidents (Stanton et al. 2010). When a control room is being redesigned or modernized, it is especially important to understand the effect of the changes to the overall process control activity.

The assessment of the appropriateness of control room in safety–critical industry calls for a methodology that is able to take into account the whole work system. There are several aspects to safety of operation that must be considered when evaluating the control room. In this paper, we describe a methodology that has been designed to assess the quality of a control room comprehensively. The methodology is based on two basic assumptions: the quality of control room becomes evident only in a real usage situation, and the control room must support good outcomes, way of working, and users' experience.

P. Savioja (✉) · L. Norros
VTT Technical Research Centre of Finland,
PL 1000, 02044 VTT, Finland
e-mail: paula.savioja@vtt.fi
URL: http://www.vtt.fi

The methodology has been developed in a series of studies of work within complex sociotechnical systems (Norros 2004; see also Norros and Nuutinen 2005; Savioja et al. 2008). The current form of the methodology that is presented in this paper, roots to the two studies presented also in this paper, in which the current control room solutions of two NPPs were evaluated in the intention of formulating a baseline data set prior to major changes in the main control room. The baseline data constitute a reference that can then be used in comparative studies during the design process and finally before implementation in the integrated system validation. The main emphasis of the paper is on the methodical description of the evaluation approach. Results of the evaluation studies, including the development needs identified with regard to hybrid control rooms, are presented to illustrate the benefit of the methodology. The detailed results have been communicated to the respective plants in non-disclosure conditions.

2 Background

The overall quality of a control room can most profoundly be understood through usage of the system (O'Hara 1999). Another possible approach, reviews and inspections conducted against relevant standards can surely give valuable insight in the individual technological solutions within the control room, but in order to assess the totality of the control room, comprising of different systems, UIs, and multiple users, the contemplation must concern real usage in order to maintain validity. Hence, the development of a control room evaluation approach has concentrated on test methods employing professional users in usage experiments. In a usage experiment, the system is used in as realistic as possible settings: real users, realistic scenarios, high fidelity–simulated process situation, etc.

2.1 Evaluation of technologies in the field of human factors

In the planning phase of a usage experiment, many issues are to be resolved about the forthcoming evaluation that can be summarized under the heading of choosing an evaluation approach. Evaluation approach unveils the theoretical principles and assumptions that determine, for example, what kind of data are considered to reveal the quality of the control room. For example, it is interview, questionnaire, performance, or task load data. Neale et al. (2004) recommend a multi-method evaluation approach including observation among other methods to be used in an environment in which multiple users use the system in work setting.

In designing a study that involves observing users using a system, it is important to explicate what are the unit of observation and the related unit of analysis. This is not always evident. In traditional human factors, the unit of observation is often user performance, and the unit of analysis can be the user error, situation awareness, completion time, etc. This means that the reasoning logic goes like this: If the users perform well, meaning that they do not commit errors, have high situation awareness, complete tasks in a timely manner, etc., then the tool must be of good quality.

Human factor methods that utilize human performance methods in a way described above are grounded in well-established scientific method that is based on an idea of testing well-defined hypotheses in controlled experiments. The aim is to prove a theory or an assumption (e.g., safety of a new design), and thus, values such as independency of evaluation from the design process and statistical power are emphasized in the approach.

Without undermining the importance and viability of this approach, concerns may be expressed that the traditional human factors approach cannot truly capture comprehensive quality of a control room due to the restricted scope of analysis reflected in the unit of analysis. For example, the adequacy of prevailing evaluation methods has recently been explicated in the domain of health care Randell et al. (2010). Health care applications are set in a complex context that involves users (personnel), patients, and ever-evolving practice of medicine. The methods that concentrate on clinical outcome measures or that attempt to meet the "gold standard" of randomized controlled trials or that aim for quantitative results by exploiting laboratory user test do not reveal the complexities with which the components of the system—technological, clinical, social, organizational, and professional interact together to produce health care for the society (Randell et al. 2010). Concerns have also been expressed with regard to the adequacy of methods used in the control room technology evaluations in the nuclear power production domain (Braarud and Skraaning 2006; O'Hara 1999), yet, without explicit questioning of the used unit of analysis.

The problem with controlled experiments is that the complexities that actually are a major factor contributing to the possible hazards in the work cannot be properly incorporated. By introducing control to the experiment, the complexities of the real work are excluded. In measuring human performance with the outcome related metrics, the activity itself, the practical work and interaction with the process and the tool is paid less attention to. It is almost as if the actual usage is treated as a black box: The users are introduced a new tool, they use it in their work, and then, the outcome is measured. This whole logic is based on the assumption that the possible problems in the new tool

would always have an effect on the outcome. But this is not true. Professional users have been reported to perform well even in adverse conditions (Norros 2004). Another problem is that outcome-related measures leave very much room for interpretation of the results because it is difficult to know what happened inside the box, that is, what are the mechanisms of usage which produced the outcome.

The unquestionable advantage of the human factors approach is that it takes seriously the objective of safety that is inherent in probably all process control domains and especially in safety–critical industries. But, the problem is that safety in a complex activity is approached with the outcome measures that extract some of the complexes of interacting with the technology away (for the sake of control in the experiment). And who knows, it is possible that it is exactly these complexes that in the end pose the greatest threat to safety. We claim that safety must of course be proved by demonstrating a process control activity that is free of disturbances (errors), and this demonstration must be independent and have statistical proving power to rule out the possibility of chance. But, at the same time, the scope of measures used should be widened (for developments in this direction see also Schraagen et al. (2008)). The outcome measures do not reveal all the potential threats to safety, which might exist in a system, because they are not sensitive to the delicate features of the activity which do not pose an immediate threat but are the possible seeds from which the user practices start to develop into a non-optimal direction. We claim that when a new technological system is introduced to a safety–critical domain, it is not enough to demonstrate that activity is safe when evaluated with outcome measures. In addition, the internal mechanisms with which the outcome is achieved must be considered, and the potential of the tool must be evaluated.

2.2 Evaluation of technologies in human–computer interaction and usability

The human–computer interaction (HCI) or usability approach to evaluation of UIs has developed in parallel with the human factors approaches. Usability's main interest is not safety but rather to understand the mechanisms of interaction and to utilize this information in order to improve design (Carroll 1997). In analyzing usability, the unit of analysis is the human technology interaction[1] (HTI). Typical metrics in studying HTI are usability and user experience. Usability refers to the effectiveness,

efficiency, and satisfaction in use, whereas user experience, although a somewhat controversial and different interpretations bearing concept (Roto et al. 2011), generally refers to the feelings and emotions by which the user is seized in encountering and interacting with the system.

The indisputable merit or the usability approach is that the user and the mechanisms of interaction have been brought to the center of attention. This has meant a significant shift of focus from the early days of computing and human factors even. This development has also lead to more profound understanding of the overall complexity of the usage situation and the contextual factors affecting it. Perhaps, the most influential addition that usability approach has introduced is its inherent orientation toward design, evident especially in improvement of the processes of design, by including participatory aspects.

There are several other domains that share some important characteristics with NPP process control and thus have had to address usability evaluation with novel approaches as new technological tools have been designed and implemented. For example, other industrial process control (Nachreiner et al. 2006; Nickel and Nachreiner 2008), health care (Dahl et al. 2010; Stevenson et al. 2010), aviation and space (Huang et al. 2006), maritime and rail transportation (Lützhöft 2004; Lepreux et al. 2003;) are domains that lay special requirements for UI evaluation. In all these domains, the potential threat for human life and environment in the case of abnormal system behavior is great, and thus, the domains are all safety critical. In addition, users are professionals with niche expertise and several years of training, and the object to be controlled is complex, dynamic and uncertain, and the manipulation of the object is indirect.

In the domain of air traffic control (ATC), development and evaluation of new technological tools has been addressed by Twidale et al. (1994) who describe an evolving evaluation approach that is bound to the overall design process of a novel ATC tool. They emphasize the role of an informal evaluation procedure in guiding design and argue that anecdotal evidence gathered in ethnographic studies can actually contribute to knowledge about general issues of collaborative working and the potential role of new tools. The mechanism of generalizing from anecdotal evidence is not discussed by Twidale et al., but they claim that even novices and students can spot apparent usability and functionality issues in low-fidelity user tests.

The so called third paradigm (Harrison et al. in press), or third wave (Bødker 2006) of HCI, is enlarging the scope of the field to cover the entirety of technologies included in the modern life. What is characteristic for the third paradigm is an epistemological orientation that embraces the social, cultural, and physical situatedness of both users and analysts and that treats interaction as a form of embodied

---

[1] Probably the most commonly used term is human computer interaction (HCI) but by replacing the "computer" with a more general "technology" the scope of the research field widens slightly and also technologies not involving computers can be addressed.

meaning making in which the artifact, its context, and its study are mutually defining and subject to multiple interpretations (Harrison et al. in press). According to Harrison (in press), the third paradigm consists of the variety of approaches such as participatory design, value-sensitive design, user experience design, ethnomethodology, embodied interaction, interaction analysis, and critical design. As the new developments in HCI research are taking place also the general understanding of the mechanisms guiding human activity and living with technology takes new forms such as described in, for example, Dourish (2001) or McCarthy and Wright (2004). How the users experience the technology, the user experience, has been brought to the center of analysis to enable the analysts understand the rich reality in which technology is evaluated.

2.3 From usability toward activity oriented approach

The usability approach has developed first with the development of the personal computer and later with the internet and different kinds of consumer electronics. Because of these roots, safety has not been the first quality objective, and thus, the approach is not suitable for control room evaluation directly as such.

Usability evaluation methods most often deal with a single user action whereas process control work is almost always collaborative. Although measures for collaboration do of course exist (Burkhardt et al. 2009; Mattsson 2011), operationalization of the usability of a collaborative system, that is, what system characteristics affect the collaboration, is difficult to find. In a literature view of 41 papers on evaluation of collaboration in computer supported cooperative work (CSCW) systems, Mattsson (2011) found that usability (in addition to collective measures) was the least studied category. CSCW research methodologies aim at understanding particularly the collaborative aspects of using computer systems in work. The aim in CSCW is in design, and thus, evaluation also belongs to the scope of research activities constituting CSCW. The challenges of evaluating CSCW systems have been elaborated by Neale et al. (Neale et al. 2004) who advocate a mixed-method evaluation in order to tackle the variety of challenges in evaluating collaborative systems.

If the domains in which usability approaches are mainly applied are compared with process control work perhaps the main difference is that in process control work the manipulation of the object of work is indirect. With this we mean that the control room, as a tool of the users, has a mediating role in the process control activity. The user is not hands on interacting with the real world live process (Vicente 1999) whereas in, for example, office work the object of manipulation, even though abstract in nature, is directly in the reach of the user. In process control work,

the user manipulates the UI, which communicates with the control system sending process control signals (e.g., close a valve, control the flow) to the actual process. This means that if the user wants to change something about the process s/he has to both understand the task in terms of the controlled production process and understand it in terms of the control system in order to give a correct command to the control system. This indirect nature of interaction in process control work induces some changes to the prevailing usability principles and measures. These changes have been elaborated by Savioja and Norros (2008) in outlining in addition to the instrumental functions of the tool also the psychological and communicative.

Yet, another major aspect of the usability approach which is not directly suitable for control room evaluation is that of learning and training. Where traditional usability measures emphasize learnability and "out of the box" immediate usability, neither of the measures is really relevant in process control work. In control rooms, the case is almost the opposite: Systems that are considered very usable by the operators seem almost incomprehensible on the first look for an outsider. This is explained by the extensive training that the operators go through. For example, in nuclear domain, the training of operators (who have as a background an engineering degree) takes more than 3 years. In this time, the seemingly incomprehensible control room UI becomes an effective tool as the new signs and meanings used in the UI are learned. Béguin and Rabardel (2000) have, based on notions of activity theory theoretically elaborated this process of the technology becoming a tool in use. They have labeled the process instrumental genesis.

3 Contextual approach to evaluation of control rooms

In order to build a methodology for evaluating appropriateness of control rooms, the human factors evaluation methods and the state of the art usability evaluation have been reviewed and combined to constitute an approach which is meant to both assess the effects of the technology on safety and understand the mechanisms of interaction on a level that is able to identify the potential hazards that are maybe not evident on the level of outcome but which may be an early sign of degradation in the sociotechnical system. All of the above described approaches have provided inspiration for the method of evaluating control rooms described in this section. In our own approach, we have weaved in together the "empirical-positivist approaches" arising from cognitive systems engineering and the empirical-hermeneutic approaches exerting from post-modern understanding human conduct (Upton et al. 2010).

In addition, the fundamental starting point for understanding the appropriateness of a control room comes from

cultural historical theory of activity, in short activity theory (AT), in which the role of tools in activity has been thoroughly examined. We have labeled our approach contextual. This characterization denotes that the activity, in which the technological tool is used, constitutes the context of use for the specific technology.

AT has been suggested as a promising theoretical framework for HCI already in the early 90s (Kuutti 1991; Bannon 1991), and AT inspired design approaches (Gay and Hembrooke 2004; Hyysalo 2002; Kaptelinin and Nardi 2006) have since been developed and advocated (Norman 2006). The theoretical concept of activity has also been widely adopted in the French developmental research (see e.g., Daniellou (2005) for review of the French approaches). The significant contribution of the above mentioned approaches is to consider the design and use of technology in a socio-cultural context. As a result the unit of analysis of tool usage is a multilayered system of activity. Usage of tools is seen to be constructed as an interaction between users of different user groups and designers (see e.g., Gay and Hembrooke 2004, pp. 2–14). We share this elaboration of the unit of analysis and the idea of the interactive construction of the design product. Furthermore, the mediating role of tools in activity is one of the basic theoretical assumptions in AT. Consequently, in the AT inspired design the role of tools is of course acknowledged. For example, Gay and Hembrooke (2004) discuss the bidirectional nature of tool mediation, that is, how perceptions, motivations, culture, and actions shape the tool and simultaneously are also shaped by the tool (pp. 5–6). Béguin and Rabardel (2000) for their part draw on the bidirectional mediation and maintain that to be defined a tool technology must have been appropriated by human actors. In this paper, the mediating role of tools is elaborated and the different roles of tools are built into a new way of *evaluating* the appropriateness of tools in particular contexts. Detailed means for the evaluation of tools is something that the earlier appliers of AT in HCI have not proposed even though evaluation has been identified as an important function of an activity-centered design (Gay and Hembrooke 2004, p. 12).

3.1 Conceptualization of activity

In this paper, it is assumed that the reader is familiar with the central concepts of AT, such as activity, subject, object, and tool. For a detailed elaboration of the basic concepts, we refer to the original writings (Leontjev 1978; Vygotsky 1978) and the later interpretations and developments (Engeström 1987; Kaptelinin and Nardi 2006; Norros 2004).

The central concept of AT is *activity*. Activity is understood as historically and culturally developed. Hence,

the central aim in the analysis is to find out the current state of affairs but also its historical roots and possible trends toward which development is proceeding. The approach suites well the needs of control room evaluation in a state in which hybrid technologies have been implemented and a more profound modernizations are under design. In order to understand whether the development of tools is proceeding in a good direction, the wider historical context of tools must be understood.

In the following section, the aim is to highlight two major principles of AT that we have effectively used in developing our contextual approach to evaluation of technologies. These principles are the *object-orientedness* of activity and the *mediated nature* of activity.

3.1.1 Object oriented activity: content of work is elaborated

Activity is a process that denotes to the continuous interaction of the human with the environment. Activity is always directed to some parts of the environment that may provide possibilities to maintain and develop human existence. The material or conceptual entity toward which the activity is directed to forms the object of activity. The environmental possibilities and the expected outcomes that may be reached while acting on the object are central determinants in structuring the activity. The object of activity is considered as the motive of activity, and it is the purpose of the organized conduct of a community of people that collaborate in the domain. Hence, motive does not refer to an individual driver of action (motivation). To be successful in their activity, it is necessary that actors take into account the actual and real possibilities and constraints of the domain to fulfill their purpose. We have developed a framework to identify and model the possibilities and constraints of the domain and the demands that they set on appropriate activity. These demands define the core task of the community of actors (Norros 2004).

In the analysis of actual actions of people, it is now possible to observe how actors in real singular situations take into account the constraints and possibilities of the domain and according to which logic they do so. We observe how the core-task demands are portrayed in specific situations. In the analysis, three perspectives to activity are included:

1. Analysis of *performance* sequence. It is necessary first to identify the sequence of actions and the operations the actors accomplish and the outcomes they reach while performing certain tasks. This aspect of analysis answers the question what was done in a situation and with which outcome. The analysis of performance sequence corresponds to those accomplished in most

human factors or end-user studies, in which the attempt is to analyze the activity in the light of the performance outcome such as task completeness or time.

2. Analysis of *way of acting*. In safety–critical and complex domain, the outcome is, however, not sufficient as the only measure. This is due to the fact that numerous, for example, technical, training-related or organizational (e.g., procedures) barriers have been designed to neutralize the effect of possible performance variance on the outcome. All actors provide required actions and reach the outcome sufficiently. Yet, in our studies, we have found (SAFIR 2011) that the actors themselves and the trainers report that there are clear differences in the ways of accomplishing the work. We are interested to identify these differences and their significance to safety or other targets of activity. Hence, in addition to outcome, the way of achieving the outcome should be measured. We develop behavioral patterns and use them as markers to signify good ways of acting. These markers are different depending on the type of work, but in all cases, they have a connection to the core-task demands of the particular work and to the way they are taken into account in actual situations. The behavioral markers can be identified by a thorough work analysis.

3. Analysis of *experience* in action. In AT and in philosophically oriented analyses of activity within pragmatism (Dewey 1999; Määttänen 2009; Peirce 1998a) or phenomenology (Kestenbaum 1977; Merleau-Ponty 1986) human–environment interaction is considered to take place as continuous embodied action-perception cycles, during which potential is built to anticipate results of action, and to create adaptive methods to respond to the contingencies of the environment (see further Norros 2004). Perceptions of the environment and technologies embedded in the environment and used in acting on the environment are accumulated in the experience the actors. We are interested in the qualities of experience and awareness that is accumulated in action, because it reveals inherent features of action that cannot be reached by observation from outside (see also Norros et al. 2011a).

In the issue of experience, we also draw on the AT that considers the outcome of activity to be twofold. For the first, there is the material outcome, for example, continuous and safe production of energy into the electrical grid. There is also a second aspect of outcome which relates to the development potential created via the activity to improve the production activity, for example, to deliver expertise and good practices of safe implementation of nuclear production technology. When people learn to use particular tools and technologies successfully, a new mediation is created into the complex activity and a positive emotion emerges (Koski-Jännes 1999; Vygotsky 1978).

In summary, we may state that the first principle of the contextual evaluation approach is that all the three perspectives to activity must be addressed when evaluating technologies.

3.1.2 Mediated activity: functions of a tool in an activity

A second central concept of AT that we draw attention to is mediation. This notion refers to the nature of the relationship between main components of an activity system. For the relationship between subject and object, mediation means that human interventions with the object take place through historically and culturally developed artifacts: tools and instruments. Thus, the relationship of subject and object is mediated by the tools. Tools provide the object to the subject and provide the possibilities to act upon it. In the context of control rooms, this means that a control room (consisting of several different UI solutions) is a tool with which the operating crew controls the object of work, the power production process.

Vygotsky (1978) elaborated mediation by making a distinction between two different functions of tools in an activity: instrument and a psychological tool. Instrument refers to the tool's ability to produce the intended effect in the environment whereas the psychological refers to the comprehension of the tool's potential by the human and the tool's capability to function as an external control of human action. The distinction is exemplified by describing the use of a common tool, a hammer. The instrumental function of a hammer is the ability to slam down nails, to make the nail dig into the surface. The psychological function of the hammer means that the human understands what hammering is and knows how to do it with the tool. For example, immediately when a person sees a hammer they understand that now they have a possibility to hammer nails, that is, the human has learnt his/her possibilities to manipulate and control objects with the tool. The psychological function is very important in activity because it enables reflection of activity which is the basic prerequisite for learning and development of the activity. Fairly recently, Georg Rückriem (2003) proposed a third general function for a tool in an activity that of communication. The idea is based on media theory and the new aspects which the digital tools bring into tool functions (Rückriem 2009). Communicative function of a tool denotes to the social aspects of using a tool. For example, a selection of a particular tool communicates intentions and purposes of action within a community. The communicative function

addresses the issues of sense making in action and the meaning of action in a wider cultural and societal perspective.

The concept of Systems Usability (Savioja and Norros 2008) has been coined to refer to a system's overall meaningful role in an activity system which can be completed by fulfilling each of the above described three functions. Systems Usability emphasizes the systemic and mediating role of information technology in an activity system. Savioja and Norros (2008) claim that for the prevailing approaches of usability evaluation it is inherent to use the concept of action to refer to human conduct in human technology interaction. This emphasizes the analysis of usability on the level of the instrumental function of the tool because action as such does not refer to the more global, societally defined purpose and objective which are the reasons for the action and which explain the psychological and the communicative functions of the tool.

The second principle of the contextual evaluation approach is that all three functions of a tool in an activity must be addressed.

3.2 Framework for analysis of tools in activity

By combining the principles concerning the tool and the activity, that is, by stating that the tool must fulfill all three tool functions in an activity and that the activity must be addressed from three perspectives, we get a two-dimensional space (Fig. 1) that denotes the frame of the evaluation approach.

Below (Fig. 1), there are nine classes of metrics that constitute the evaluation framework for assessing systems usability of a complex tool. The framework is a conceptual tool which aids in finding relevant methods to be used in comprehensive evaluation. In the following section, we

describe the measures of good performance in the context of control room evaluation. At the same time we are building a characterization of good process control work.

3.2.1 Measures of performance

Measures of performance describe the outcomes of work and reflect the first perspective to activity. The measures aim at being objective and characterize the work as it can be seen from the outside by an external observer. In control room evaluation, instrumental outcome refers to the operators being able to carry out their tasks with the tool. This means that the tasks are completed in a manner that does not endanger safety or production related goals of the activity. Psychological performance refers to the measures which capture the cognitive performance of the operators such as cognitive load. Communicative performance refers to the collaborative aspects of process control activity, for example, the amount to which process information is communicated out loud by the members of the operating crew.

3.2.2 Measures of way of acting

Measures of way of acting supplement the measures of performance by revealing intricacies of how the performance is achieved by the operating crew. The overbearing quality of way of acting is its orientation to the core task. This analysis is important as it provides a way of considering the underlying mechanisms of human conduct which produce the outcome. It reveals reasons for the outcome. The question of "how" can be answered by studying meticulously the ways of acting of the users.

On an instrumental level the way of acting means that the tool helps the users to focus on relevant phenomena in

Fig. 1 Framework of analysis of tools in activity
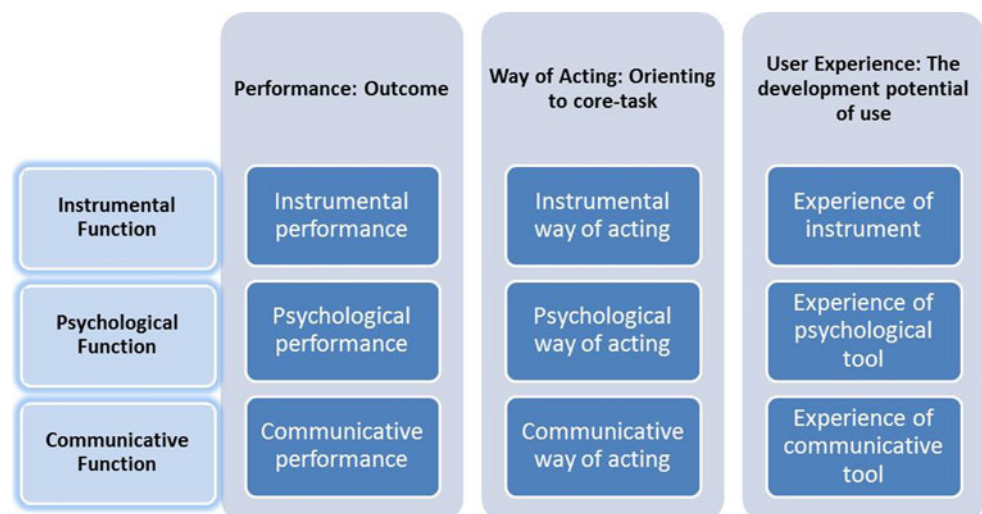
the process, it aids in focusing on the tasks that are most crucial and relevant in the given situation. This is manifested in a way of acting which can be characterized as focusing on core issues and effective prioritizing.

On a psychological level the way of acting means that the practices of use are such that they strive for understanding of the process situation thus being able to anticipate and being in control. On a psychological level this way of acting requires profound conceptual knowledge, and operative schemes about the process, its conformance with the natural laws and the controlling automation system. The tool should be such that it enables and supports the development of these psychological capabilities.

On a communicative level the way of acting refers to the features of teamwork and shared understanding of the prerequisites for good performance. A good way of acting in the communicative form is such that it enables the crew to share the object. This means that each individual is able to see the own influence on the joint object. In order to support communicative way of acting, the tool should support good teamwork shared awareness of the process status.

3.2.3 Measures of user experience

Measures of user experience emphasize the role of professional users in the design and acceptance of tools for the work. In addition to the tool providing the users the opportunity to carry out their tasks and do it in an appropriate manner, the tool should be such, that the users feel that it is a good tool for the particular work. The measures of experience are aimed at finding out about this quality. The over bearing quality of user experience is that the user feels that the technology has potential to develop into a meaningful tool for the activity and improves the interaction with the environment.

Instrumental experience means that users feel that the tool works well: there are no unnecessary complicacies in using the tool. Feelings of achievement belong to this category also.

In the psychological function, the user experience means that the user is confident in using the tool and the tool is embodied to the extent that the usage feels effortless and natural.

In the communicative function, the user feels that s/he can trust the tool the same way one can trust another operator within an operating crew. The tool is trusted to communicate all the needed process information in a way that is comprehensible to the users. The tool improves transparency and anticipation within a team and also improves shared understanding of the situation. In a wider perspective, it also mediates the values of good practice.

3.2.4 Using the framework for finding measures of systems usability

In the following chapter, we describe two case studies in which evaluations have been carried out in NPP control rooms with the aid of previously described evaluation framework.

4 The studies in hybrid control rooms

The initiative to look into evaluation methods for control rooms originated in a situation in which a thorough analysis of existing main control rooms of NPPs needed to be made prior to the start of colossal modernization projects. In a safety–critical technology domain, the approach to any modernization must be conservative: "If it works do not change it." This means that before anything is transformed, the current state of affairs must be understood as thoroughly as possible in order not to—by any chance—accidentally deteriorate the situation. For this purpose, extensive reference tests were conducted at two NPPs before effective modernizations of main control rooms were started. In the first phase, the focus was on the evaluation methodology (Norros and Savioja 2007). Conducting the reference tests served as a test bed for evaluation methodology development. It was possible to explore different kinds of task analysis, data collection, and data analysis methods to find out which would best support construction of understanding about the mechanisms that make the current tools usable for the activity. Later, in parallel with the progress of the modernization process, the method was implemented in more and more mature form to deliver required analyses of the control room designs. Results of these analyses are demonstrated in the following.

4.1 The plants and the control rooms, respectively.

Extensive studies were conducted at two different NPPs which were at the time planning and searching possibilities for the modernization of the main control room. Both plants originate from the same era, the turn of 1970s and 1980. In this paper, the plants are referred as plant A and plant B. Plant A consists of two pressurized water reactors producing jointly close to 1,000 MW electrical power, and plant B of two boiling water reactors producing together about 1,700 MW electrical power. Both plants have two main control rooms respective to the units in operation. Both plants have a similar concept of operation in which a crew consisting of three main operators (turbine operator, reactor operator, and a shift supervisor) operates the plant from the main control room in normal situations.

In plant A, the main control room is the end result of an extensive in-house design effort. As a result of earlier improvements the original analogue hardwired controls are used for operation but modern information system has been implemented for monitoring purposes. The on-going more thorough modernization of the I&C systems has created a hybrid control room that includes, in addition, a touch screen display for the operation of control rods. Plant A has also gone through modernization of the emergency operating procedures (EOP). EOPs are still in paper format but presentation is flow chart based and philosophy is partly state based (see e.g., Park 2009) diagnosing. Thus, the modernization efforts in plant A have started from the most safety–critical parts of the main control room: the EOPs and control rod maneuvering.

In plant B, the main control room constitutes also a hybrid concept as modern monitoring systems have been implemented. More importantly, the plant has gone through modernization of non-safety class turbine operation UIs. This means that half of the process is operated through displays via soft control methods and the other half (the reactor processes) through analogue hardwired controls. Thus, the modernization has started from the opposite direction in comparison with plant A.

The hybrid nature of the main control rooms in both plants was a central issue in the studies conducted. In both of the plants, the main control rooms were partly modernized, but more extensive renewals were in planning. Modernization of the control room systems opens up new possibilities for information presentation and user interaction with the process whereas at the same time the conventions developed over long periods of usage will change drastically. In a traditional control room, the view to the process is parallel: all the information concerning the process is spread on the wall panels and bench boards. As display based interface allows more extensive and elaborate information presentation, even in process situation based manner. But, as a downside, windowing typically hides some information and the operators must actively search and find relevant information. Also, requirements for operating crew's collaboration and communication are expected to change with the transformation from analogue UI to the digital one. In a traditional control room, it is relatively easy to know what sub system another person is working on judging by his or her physical location. Thus, awareness of other crew members' activities is relatively high in a traditional control room. This is expected to change when operators start using individual display units as the other person cannot know which particular window the other person has open on his or her screen.

The motivation for the studies in both plants was to gain more profound understanding about how the current hybrid control room works as a tool in the collaborative process control activity. In the beginning, there was worldwide little experience of the significance of the control room changes to operator work. It was common to emphasize that, as the modernization does not induce changes in the power production process, the so called primary tasks of process control remain the same. Only the secondary tasks, involving handling of interfaces, search for information, navigating in the control system, etc., would be reshaped. The question emerged concerning the significance of changes in secondary tasks on the operator work. This issue was not thoroughly considered in the international guidance (O'Hara et al. 2003). In addition to this rather general research question, both plants had more specific issues that they were interested in. Plant A was additionally interested in the role of EOPs in the construction of operator activity. Plant B was additionally interested in the possible discrepancies induced by the differences in the UIs of reactor and turbine process control.

Below (Table 1) the research setting within each plant is summarized.

4.2 Research setting at the training simulator

The data collection was conducted as part of normal operator training in both plants. NPP operators go through extensive simulator training each year. The training simulator in both plants is a high fidelity dynamic process simulator which is nearly an exact copy of the main control room of the plant. Also the process response of the simulator is close to a perfect match with that of the real plant.

4.2.1 Participants

In plant A, all 12 operating crews participated in the study which means that altogether 44 operators acted as users in the usage experiment. In general, an operating crew consists of three persons, but in many of the crews, there were trainees who also took part in the simulator exercises. The crews acted in the study, just as they would act normally in the control room which means that the trainees had, depending their experience, either an assisting or a master role. The operating experience of the operators in plant A varied from 1 to 32 years of experience. There were 18 participants in the experience group 1–9 years, 13 participants in the experience group 10–19 years, and 13 participants in the experience group over 19 years.

In plant B, six operating crews participated in the study. Altogether, there were 24 participants and the operating experience varied from 0 to 31 years. 12 operators had operating experience of 0–9 years, 3 operators had operating experience of 10–19 years, and 9 operators had operating experience exceeding 19 years.

Table 1 Research setting concerning plant A and plant B

	Plant A	Plant B
Plant type	Pressurized water reactor	Boiling water reactor
Main control room	UI for control rod manipulation digital, all other UIs analogue. Digital process information system.	Turbine side UI digital, reactor side analogue, digital process information system
Modernization phase	Flow chart EOPs (paper based), modernized control rod manipulation	Turbine side modernized
Plans to modernize	Plan to modernize the whole main control room	No plan for full modernization
Research questions	Formulate a reference for validation	How does the modernized turbine UI work?
	Find out how the new EOPs are used	Formulate reference for possible future needs

4.2.2 Scenarios

In a training simulator, practically any process situation can be simulated. It is up to the imagination and creativity of the simulator trainer to create scenarios that serve a good learning possibility for the operators who participate in the training. Similarly, in the reference tests, the proficiency of the simulator trainers was relied upon to create meaningful scenarios, in which the process control activity could be studied in the light of the afore-mentioned research questions.

As plant A was interested in use of emergency operating procedures, two accident scenarios and one complex failure (incident) were created. The scenarios at plant A were: 1. Loss of coolant accident, 2. Primary—secondary leak, 3. Electrical bus failure. Scenarios 1 and 2 are so called design basis accidents, which means that specific emergency operating procedures cover the necessary process interventions which operators are required to carry out. In scenario 3, it is not evident which operating procedure would best suit the situation.

Also, in plant B, three scenarios were created for tests. As the plant's own interest lay in normal operations and the operators' ability to take advantage of the modernized turbine side control system UIs, the scenarios were decided to be smaller scale failures, so called incidents. Scenario 1 was a failure in decay heat removal system, scenario 2 was an ejector failure and scenario 3 an automation failure in a pre-heater line.

All of the scenarios were represented as functional situation models[2] (FSM) to enable the analysis of operator activity. FSM is a task model in which the critical process events are described both chronologically, and from the point of view of critical safety functions of the NPP

process. This means that a process event, for example, a valve failure is connected to the critical safety function which the failure endangers. For example, in a loss of coolant accident the critical safety function of core cooling is endangered. In a functional situation model, the required operator actions (represented, e.g., in a hierarchical task model) are connected to the critical safety functions. For example, in a loss of coolant accident, it is extremely important to start auxiliary feed water pumps to maintain the critical safety function of core cooling. Thus, in FSM the action, "start auxiliary feed water pump" is connected to the safety function "maintain core cooling." The detailed modeling is needed because these functional connections between actions and purposes are not always evident. The FSM technique was developed particularly for analysis of the three different tool functions.

By modeling the scenarios with the FSMs it is possible to connect the operator action steps to the higher level goal of that specific step. This type of analysis of the simulated process situations allows the use of activity theoretical evaluation of the operating practice of the crew because in AT the appropriateness of an action is determined by the goal it is meant to achieve. The analysis is not limited to judging whether the crew is able to start the auxiliary feed water pumps (task completeness) timely enough (completion time) but we can also say whether they did it for the right purpose. This can be analyzed by following which process signs the crew follows after starting the pump. If they look at signs related to the goal, they must have made the connection (conscious or from the gut) of the action and the goal. Thus, in order to be able to evaluate whether the operator practices are appropriate it is essential to be able to connect observed behavior to the higher level goal it strives to achieve.

Another goal of the FSM was to guide the further analysis of the collected data. The functional situation models cover the main events in the scenario and thus guide in choosing scenario relevant episodes in the video data for careful analysis. The modeling of scenarios was conducted jointly with the simulator trainers at both of the plants.

---

[2] Developing the methodology to describe situations in a way that makes explicit the connections of the events in the situation to higher level safety and efficiency functions, has been on our research agenda for many years. An overview of the earlier joint work in this issue is provided by Norros (2004). The use of FSM is explained also by Salo et al. (2009).

4.3 Collecting and analyzing the data

In order to have a rich and profound understanding of the collaborative process control work conducted by a NPP operating crew, a wide variety of data collection and analysis methods (Table 2) was utilized in the studies in both plants. Below, the data collection methods are described according to analysis of which of the three tool functions the data would be used.

4.3.1 Data collection and analysis for instrumental function

Instrumental function means that the crew is able to perform the required process interventions with the tool.

A process expert, the simulator trainer in both plants, acted as an expert evaluator of crew's process control performance in the scenario. The process expert had together with the researchers prepared beforehand a form, in which four general process control demands were contextualized according to the scenario in question. Thus, the forms were scenario specific but the categories were the same for every scenario. The general categories were: Diagnosis, process information search and retrieval, use of procedures, and collaboration. The expert evaluator rated the crews' performance in each category and the respective scenario-related sub category on a five point scale. In judging the performance of the crew, the process expert considered task completeness (errors) and time in giving the score. The quantitative data were treated with statistical measures. This data were utilized in analysis of performance.

In the end of the simulator session, every participant filled in a systems usability questionnaire. The questionnaire consists of altogether 51 positive usability statements

concerning the control room. Of those 18 concerned the instrumental function of the tool, that is, exploring how well the operators feel that the system works as their tool. The operators rated each statement on a four point scale varying form "totally agree" to "totally disagree." For each statement, it was also possible to make free form comments or specifications. The quantitative data collected with the questionnaire was treated with statistical measures. The free form comments were integrated with the usability remarks made in the process tracing interview. This data were mainly utilized in the analysis of experience (feeling of a well-functioning tool) but also some way of acting related statements (focus + prioritizing) were included in the questionnaire.

It was expected that if the process expert would claim that crew's process control was adequate, the process within safe margins, and that if the operators' consensus were that the control room system constitutes a well-functioning tool; we could say that the tool functions as an instrument.

4.3.2 Data collection and analysis for psychological function

The psychological function of a tool means that the operators know how to use the tool and understand how the different operations affect the process. On a performance level this means that cognitive load is not excessive. The psychological function also means that operators receive information from the process which allows them to understand and make sense of the current process state.

After each scenario the subjective workload was measured by using the NASA-TLX self-evaluation questionnaire (Hart and Staveland 1988). The questionnaire consists of 6 scales: mental demand, physical demand,

Table 2 The data collection methods during the simulator sessions

Method	The tool function analyzed with the method	Perspective to activity	Way of using the method
Orientation interview	Communicative	Way of acting	Once per person before all simulator runs
		Experience	
Observation of process control activity	Psychological	Way of acting	During each simulator run
	Communicative	Performance	
Expert judgment of performance	Instrumental	Performance	During each simulator run
Task load measurement	Psychological	Performance	After each simulator run
Process tracing interview	Psychological	Way of acting	After each simulator run
		Experience	
Systems usability questionnaire	Instrumental	Experience	One time after all simulator runs
	Psychological		
	Communicative		

temporal demand, performance, effort, and frustration. On each scale the subject marks a point between very low and very high (e.g., perfect and failure, for performance) which best describes his/her work load in the previous scenario. The quantitative data collected with the form was treated with statistical measures. This data were used in the analyses of performance (task load).

During the simulator runs the operator activities were observed by researchers and video recorded using both overview cameras and head-mounted cameras. The video data were analyzed in multiple ways. In an exploratory phase, a few of the video runs were transcribed into a chronological form in a spread sheet in which all the process events, operations, communications, some directions of gaze and operator movements could be examined in parallel.

This data were utilized in the analysis of way of acting in the following way. Based on the functional situation model some parts of the simulator runs were analyzed qualitatively concerning the operating practices and sense making. For this purpose, meaningful episodes were first selected. Then, the operator activity was meticulously followed from the videos in the selected episode. The semiotic concept of habit (Norros 2004) was utilized in the analysis. In the semiotic analysis of habits, the analysis does not consider only what people do; Attention is also paid to how they do, that is, the habit. The habit is manifested in the ways people make use of information (signs) available in the environment (Salo et al. 2009). In our analysis, we first looked at what operators did (I in Fig. 2). The deeds were things like process operations, communications, movements, etc. In the next phase, we looked at the signs in the environment (S in Fig. 2) which were observed or perceived by the operators either prior or in doing the deeds. With this information (the deed and the information based on which it was done) we were able to

Fig. 2 The semiotic model of habit (Peirce 1991; see also Norros 2004 p. 74; Norros and Salo 2009)

infer the objective which they were striving for (O in Fig. 2). This object was then compared with the goals of process control in the particular scenario which had previously been explicated in an FSM. By comparing the needs of the process and the goals in operator activity we assessed the operators' overall understanding of the process state.

After each scenario the operators were given a chance to reflect on the process control work by conducting a process tracing interview. In plant A, the process tracing interview was conducted as a group interview mediated by a trainer, and in plant B, the process tracing was conducted as an individual interview by the researchers. In both plants, the structure of the interview was similar. The operators were asked to re-live the scenario by describing what had happened in the process. After bringing up an individual process event (or state change) detailing questions about the significance of the event, the UI where it was detected and the required follow up actions were asked. During the process tracing interview the operators were also asked to comment on information presentation and interaction issues of the used control room systems. Process tracing interview data were transcribed and analyzed qualitatively. In the analysis statements concerning the appropriateness of UIs were collected. This data were utilized in the analyses of way of acting (conceptualization of the process situation) experience concerning UI.

In the systems usability questionnaire, there were 17 statements concerning the psychological function of the tool. These statements contained various operationalizations of the tools psychological function ranging from experience of suitability for personal style (embodiment and self-confidence with the tool) to encountering interruptions in work due to interface issues. This data were used mainly in the analyses of experience but also way of acting and performance were considered when appropriate.

It was expected that if the operators do not have excessive task load scores, have good understanding of the situational state of the process and if they have positive experiences concerning the tool in a psychological sense, thus the tool would function well in its psychological role.

4.3.3 Data collection and analysis for communicational function

The communicational function builds upon the previous two functions. That is, if the tool functions as an instrument and in the psychological role, it has the possibility to function in a communicational role also. The communicational role means that the tool enables and supports communication and collaboration in work. This means, for example, that personnel within the organization share understanding of the importance of process state, and have an appropriate trust in the functioning of the tool.
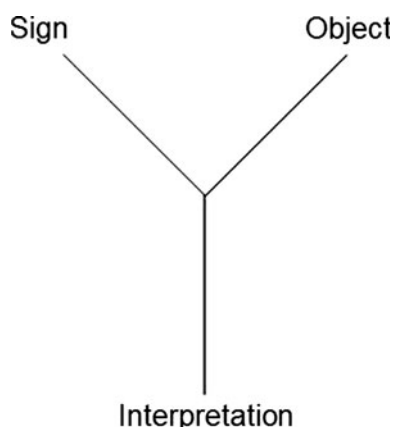
Before the simulator run, the first data collection method in chronological order within the test day was an orientation interview. The results of the orientation interview have been reported by Norros et al. (2011b). The orientation interview is an individual interview concerning the operators' personal stance toward the object in the work, the process. Orientation or professional orientation in this case refers to a person's personal attitude and understanding about the objectives of the work and about the object and its intrinsic characteristics. In the interview, six defining questions concerning the NPP process and the operators work were asked from each operator individually. For example, operators were asked what they view as the core task in their work, to reflect about the role of procedures in process control and to elaborate on their conceptions concerning role of alarms in initiating operator activity. All the interviews were audio recorded and transcribed. This data were utilized in the analysis of way of acting and experience (shared awareness, sense of control).

The answers were classified utilizing a three class system. The classification dimension is derived from pragmatist philosophy (Peirce 1998b) which makes a distinction between reactive and reflective relationship with the environment. In our previous studies on orientation (Klemola and Norros 2001; Norros and Klemola 1999), we have ended up to use the dimension reactive—confirmative—interpretive to express the different relationships (Norros 2004). A grounded approach was utilized, first, to determine how this evaluation dimension should be operationalized with regard to five themes: Role of procedures, Coverage of procedures, Focus of procedure control, Trigger for action and Concept of a good operator. Second, we determined which answers belong to which category in the operationalized scale. Summary of the characterization of each class and an example of how the categories were adapted with regard to "Role of procedures" is presented below (Table 3). All the answers were read several times by three researchers and in the end each person made an initial judgment about the class in which the specific answers would belong to. These initial judgments were then discussed jointly and a consensus was reached and a final decision made.

A reactive orientation reflects an attitude in which the actor sees him/herself as passive follower of the process. For example, concerning procedures, a reactive answer would be total reliance on procedures without reflection about the appropriateness of the particular procedure or the professional skills required in using it. Concerning the characteristics of the controlled process the actor concentrates on the own sub processes without further reflection about the totality of the overall process status. On the other end of the scale an interpretive attitude means that the actor has an active relationship with the environment. The actor is sees the whole activity as a meaning making process in which personal contributions play a vital role. An interpretive answer is such that it reflects the person's ideology that by doing and following the consequences I can learn more about the process. The middle class, a confirmative answer, regards work as shaped by rules and external demands. In the confirmative category, the role of procedures and acting by the rules is overly emphasized.

Detailed quantitative analysis of the amounts of communication and movements were conducted for all simulator runs. This analysis was done based on the observation, that is, video data. All the communications and movements of the operators were calculated, and the data were treated with statistical methods. This data were utilized in the analyses of performance and way of acting (verbal interactions, spatial interactions, and teamwork practices).

Table 3 Orientations reflecting a person's epistemic attitude in work

	Characterization of orientation	Orientation with regard to "role of procedures in process control"
Reactive	Reflects passivity and lack of expectations concerning the situation	Procedures support if the operator is insecure or when he has problems with memory; procedures prevent memory-dependent or other mistakes; procedures define the correct procedures and their order; put-down of procedures. For example, "Procedures support my work and provide security, you can have a look at them whenever in doubt"
Confirmative	Taking the situation for granted, acting in a pre-defined way, over emphasizing rules and procedures	Procedures support actions, procedures are needed in rare or demanding situations; procedures unify the ways of working and guarantee safety. For example, "They are central. About everything is done according to the procedures. Of course, routine tasks are such that procedures are not needed. But procedures are important. And we must work according to them"
Interpretive	Questioning the observed phenomena and building expectations of future events, emphasizes interpretation of situational demands	Procedures provide framework and method according to which the operations are performed. For example, "We operate according to them. They provide such a framework that they support [our work]"

Altogether 16 statements concerning the communicative function were included in the systems usability questionnaire. These statements contained various operationalizations of the tool's communicative aspects from the relevance of applied nomenclature to trustworthiness experienced by the operators. The quantitative data were treated in statistical manner. This data were utilized in the analysis of experience (e.g., trust in technology).

It was expected that the orientations, communications, and movements, in addition to reactions to the statements concerning communicative aspects of the tool, reveal the communicative function of the tool. Orientation is something that is constructed during a person's career. In earlier studies, it has been found that within a group of novices within a practice differences in orientations exist and that the differences that can be found in the interview answers are portrayed in actual practice (Klemola and Norros 2002). That is to say: Persons with interpretive orientation maintained also a reflective work practice. And this reflective work practice is something that in an unexpected (not proceduralized) situation can save the game. If operators are able to, in a tough situation, to reflect on their activity and the process state, it is possible that they can come up with creative solutions which have not been thought by the designers of the system. If a communicative function of the tool is achieved the tools is such that it enables and supports these reflective processes.

4.3.4 Summary of data collection and analysis

Below (Fig. 3) the specific indicators of systems usability utilized in these studies are presented in the framework introduced earlier in this paper.

4.4 Results on systems usability in the two hybrid control rooms

In this section, the summarized results of the two studies are presented concerning systems usability, that is, how the three general functions of a tool in an activity seem to be fulfilled in the hybrid control rooms. Because of the variety of data collection and analysis methods used, the results of the studies are also multitude. The statistical results of the studies cannot be described in full detail in this paper, but the focus is on the qualitative and descriptive parts of the results which demonstrate the benefits of the used evaluation and analysis approach.

4.4.1 Instrumental function

In the instrumental function, both main control rooms worked quite well. In plant A, the crews' performance (expert rating) was overall good (mean values between 3 and 4 out of 5), and the performance differences between crews were not statistically significant. Based on the results
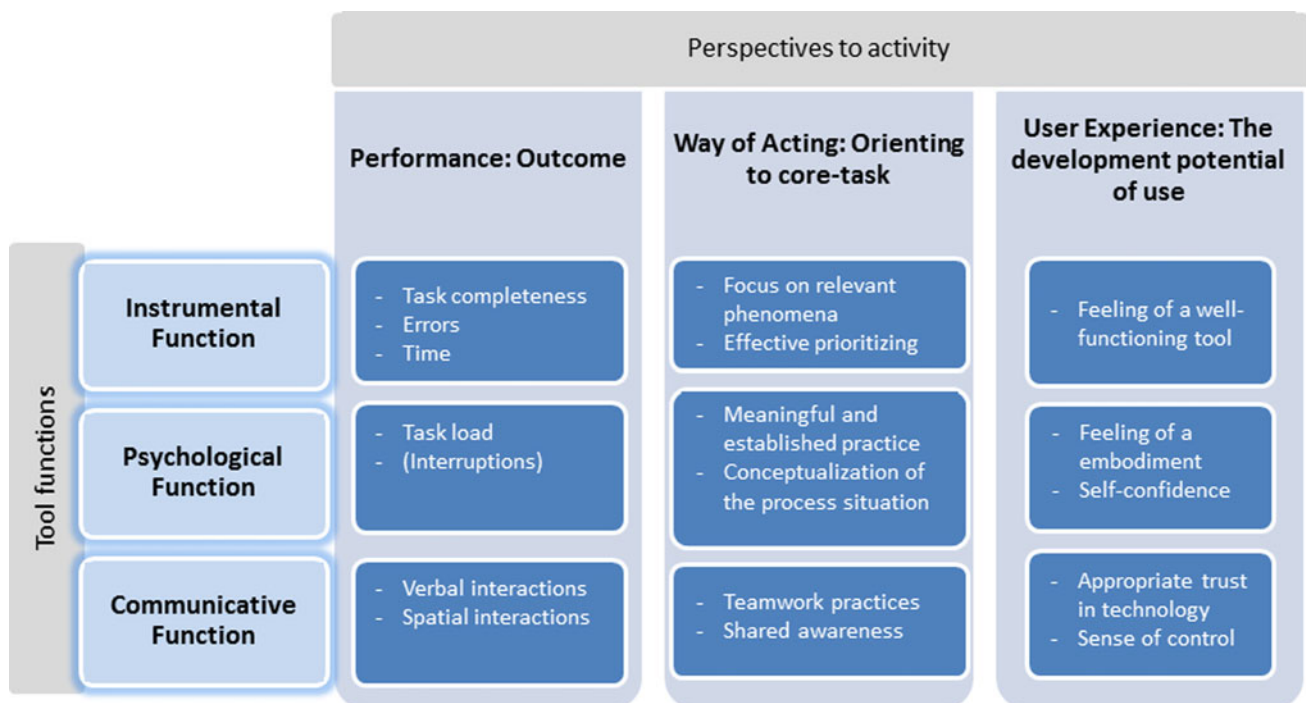

Fig. 3 Indicators of systems usability which were utilized in the studies. The indicators are embedded within the systems usability framework. Interruptions (*in parentheses*) were not analyzed

of the expert performance rating it can be said that the performance outcome in each scenario was on a satisfactory level. In this conclusion, the measure of satisfactory level performance is that of not endangering safety, that is, completing the requirements expressed in the emergency operating procedures. In plant B, the overall performance ratings of the expert varied between 2 and 5, but the average ratings were quite good varying between 3.5 and 3.9 (out of five). The lowest ratings were given for detection and use of procedures.

In plant B, the modernized turbine side UI had some problems also in the instrumental form. It does not always help operators in concentrating on relevant information as there is no alarm filtering in the displays. Every alarming component is blinking in a disturbance situation. Statement concerning the alarms was made by one operator in the process tracing interview:

> The colors do illustrate somehow, but as in a disturbance situation, there are many blinkings in many different positions, it becomes disturbing… And, the bigger the disturbance, the more there are alarming components.

The generally positive feedback (Figs. 4, 5) acquired with the usability questionnaire concerning instrumental function speaks for a well-functioning instrument. The most negative experiences concerned complexity caused in operators' work by the UI, error possibilities and especially recovery from usage errors with the UI. Also in the open comments some critique on the UI was presented. The main problem was that the needed information is spread all around the control room. Also, the information monitor might be far from the operating interface. Alarm system presents itself as problematic to the operators. Alarms are too many, and thus, it is difficult to know where to focus, and the information value of alarms is not high enough especially in the situations when it would be needed.

The conclusion concerning the instrumental function is that especially the analogue UI works well in the instrumental form. The operators typically carry out the tasks almost perfectly and the feedback is concrete and timely and the user experience is almost hands on the process. The new emergency operating procedures in plant A worked quite well in performance sense. The operators did not have any major difficulties in following the procedure, and thus, the intended effects were reached. The shortcomings that were detected concern recovery from errors and some cumbersome positions caused by the interface. The digital touch screen-based interface in plant A was considered troublesome, as sometimes it is unclear whether the command of the user has been implemented or not.
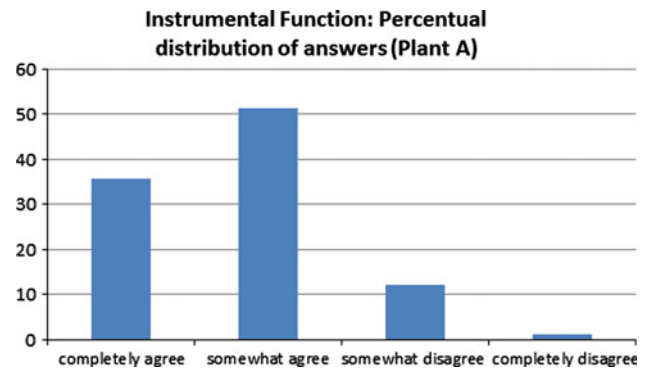

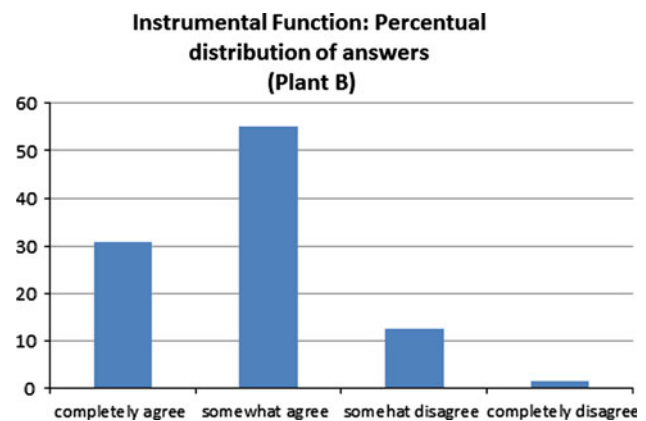**Fig. 4** Instrumental function in plant A according to questionnaire data

Fig. 5 Instrumental function in plant B according to questionnaire data

4.5 Psychological function

In the two accident situations in plant A, the crews' task loads were overall lower than in the electric bus system failure (Fig. 6). This is evidence for the psychological function of the EOPS: For the accidents, an emergency operating procedure exists, and for the electric bus failure, there is no specific procedure. The electric bus failure scenario as such was not as severe from safety point of view as the accident situations but the task loads were still higher. The scenario's effect on each task load factor except frustration was statistically significant ($p < 0.05$) or very significant ($p < 0.001$).

Despite the excellent results concerning psychological function of the EOPs the psychological function of the present day hybrid control rooms is not as unproblematic as the instrumental. This result was represented in the questionnaire data concerning psychological function (Figs. 7, 8). The most negative experiences concerned difficulty to learn to use the systems, support for finding right operative solution in an unclear situation, help of the procedure in
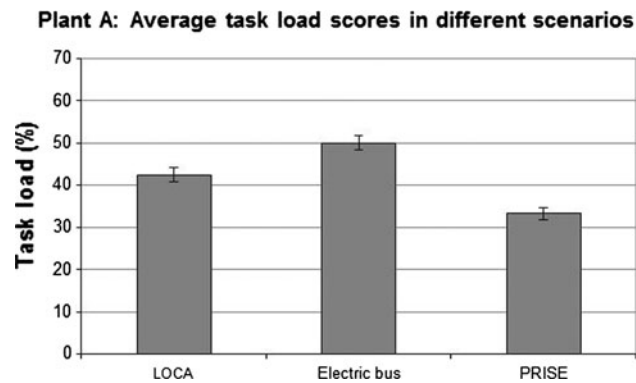
Fig. 6 Task load scores in different scenarios ($p < 0.05$). LOCA refers to loss of coolant accident and PRISE to primary–secondary leak, both of which are severe accidents. Electrical bus failure scenario resulted in higher task load values than the other two
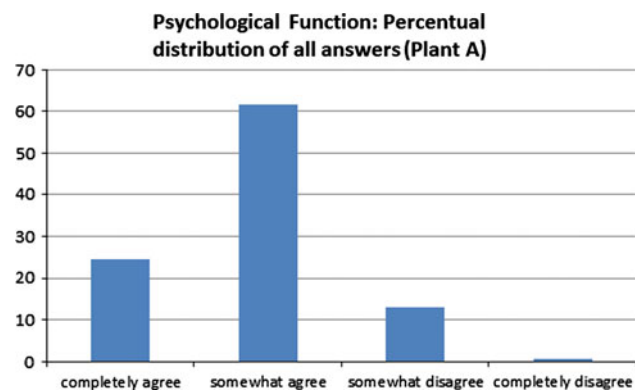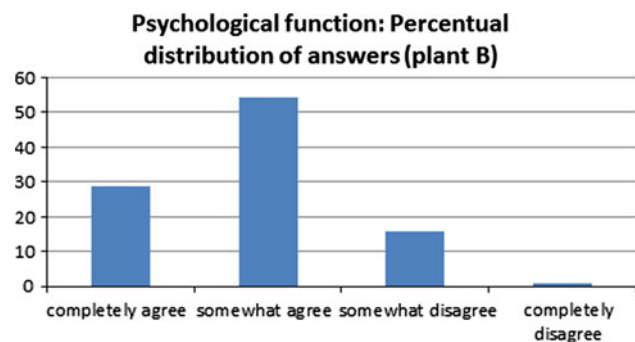

Fig. 7 Psychological function in plant A

Fig. 8 Psychological function in plant B

understanding process situation, support for personal styles, and support for adaptive activity.

One issue that came up in the different data forms is that the operators consider the UI very difficult from learning point of view. This holds for both plants. In the systems usability questionnaire, the statement concerning the effort required to learn to use the system was one of the lowest scoring statements in both plants. In addition, in process tracing interviews, remarks were made concerning the learnability of the digital UI.

> I have used it so little and then there are these logic displays…And it takes so long that I don't care to look there first… it is like digging.

In the above statement, an operator is describing why s/he did not use the logic displays even though on some level s/he claimed to know that the solution to the problem could have been found there. S/he just does not know exactly where the information resides and feels that s/he does not know how to use the interface well enough.

Although the long training period addresses the issue of learning and competencies in use, the experience of the operators cannot be neglected. The poor learnability is partly due to the hybrid nature of the control rooms. As there are UIs of different generations all installed in the same control room, the logics of operating these systems also varies. Thus, the operators feel that they have to learn each and every system separately and there is no consistency within the design on the level of the control room.

In the detailed analysis of operating practices, one quite remarkable difference between the crews in operating practice was found. In plant A, in scenario 1, (loss of coolant, LOCA) there was a small additional failure. One of the plant protection signals (containment isolation) did not function correctly. The checking of the protection signals is not in the beginning of the LOCA procedure, but nevertheless, the operators, if they are aware of the endangered safety functions, are aware of the status of the containment isolation also. Only one crew managed to notice this failure and take the respective measure to correct the situation before it was mentioned in procedure. We considered this a problem in the psychological function of the control room because only one crew was able to orient to the core demand of ensuring containment isolation.

In plant B, the scenario 3 was by far the most difficult for the operators to handle. It was analyzed that the problem lay mainly in the psychological function of the modernized turbine side UI. Scenario 3 involved a failure in the automation system which was very difficult for the operators to distinguish from a process failure. As the UI of the turbine side of the control room was already modernized the new digital interface would have given the operators an opportunity to detect that it was indeed an automation failure. But only one crew was able to make this detection and thus avoid an unnecessary reactor scram. This was considered a problem in the psychological function as only this one crew had managed to develop way of acting which makes good use of the new features of digital UI. Among the other crews there were even claims about knowing that the key to solving the problem could be found in the new interface, but deficiency in own skills prevented the person

from exploiting it. One operator explained that s/he knew that the conditions could be checked from FUP displays (logic display) but s/he had decided not to use them.

> Then I thought that I won't start using this FUP picture. But in principle you could see the conditions from here.

In the semiotic analysis of operator practices and sense making in plant A, we found differences between operating crews. Some crews utilized available process information more profoundly than the others. The exact differences were in the amount of information sources that the crews utilized in initial detections and in the amount of (out loud) reflection on the meaning of information. Five crews (out of) twelve utilized several process information sources and reflected out loud on their meaning on the overall status of the process. This can be interpreted as a problem in the psychological function of the tool. Not all operator crews were able to make use of the available possibilities of making sense of the process situation. The amount of information sources utilized is an important indicator as it reflects how profoundly the crew strives for understanding of the situation.

4.5.1 Communicative function

In plant A, there were some differences between the crews in communicative performance. Differences existed in the amounts of communication and movement around the control room, especially for the shift supervisor. Also there were differences in the extent of using and communicating process information for decision making. In plant B, the effect of the UI on communicative practice was evident in the amount of movements of the operators. The turbine operators who use the modernized UI moved less than other operators. Another notable result of the analysis of movements was that of total time of movement of the shift supervisor in plant B. The shift supervisor was away from his own station more than either of the two other operators.

In the further analysis of the operating practice of the successful crew in plant B (managing to solve the automation failure in scenario 3), it was detected that the amount of communication in this crew was higher than in the other crews. The same difference was reflected when the time that the crew spent physically together, communicating and interpreting information provided in the UIs was calculated.

Practices of collaboration varied quite much between the different crews in both plants. In a control room with strong communicative function, the overall quality of collaboration would be better and more consistent throughout the crews. This would be the case if the interface would provide support for collaboration and communication. This phenomenon was most evident in plant B where half of the control room consisted of digital UIs. In plant B, in scenario 3, effective and rich collaboration was the key to successful mastery of the process situation for the one crew which managed the situation well. From this evidence it was inferred that it is possible that hybridity induces polarization to the operating practices of different crews. The crews which manage to take advantage of the new features in the system develop different practices from the other crews. The control room's communicative function of conveying meanings within the organization cannot be claimed to be fulfilled if different crews end up working with very different operating practices. Hence, the hybridity of the control room might deteriorate the communicative function.

There were qualitative differences in the orientations of the crews in both plants. When all the answers of all the operators in plant A were coded 29 % reflected an interpretive orientation, 48 % a confirmative orientation, and 23 % a reactive orientation. The same holds for plant B: When all the answers were coded and the results calculated 23 % of answers reflected an interpretive orientation, 58 % a confirmative orientation, and 19 % reactive orientation. The confirmative orientation is by far the dominant in both plants which reflects prevailing attitude of understanding process control work as confirmatory activity in which it is enough to follow and obey the rules and procedures.

In the orientation interviews, the operators did not present very strong interpretive attitude concerning process control work and the object in the activity. Interpretive orientation would be characterized by a questioning attitude which enables building expectations of the future states of the process, and learning from operation experience. This is a central requirement in complex process control. Without questioning attitude it is not possible to handle unexpected situations that are not described in procedures or thought of when designing the UI. The somewhat poor results of the orientation interviews are connected to the communicative function of the control room. It seems that the meaning of the different process measures and automotive functions is not communicated to the operating personnel as profoundly as should be the case. When the interpretive orientation is low or recessive some of the adaptive power that human activity brings to the functioning of a complex process control system is lost.

According to the questionnaire data, the communicative function of plant A and plant B are different (Figs. 9, 10). The statements scored low in plant A concerned EOPs ability to communicate process information, control room systems' ability help finding solutions, and understandability of the sounds utilized in the control room.
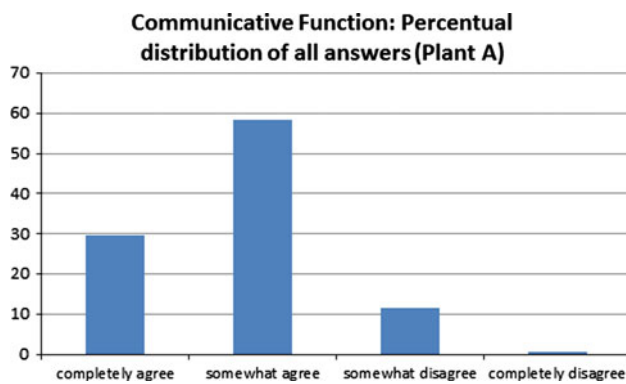
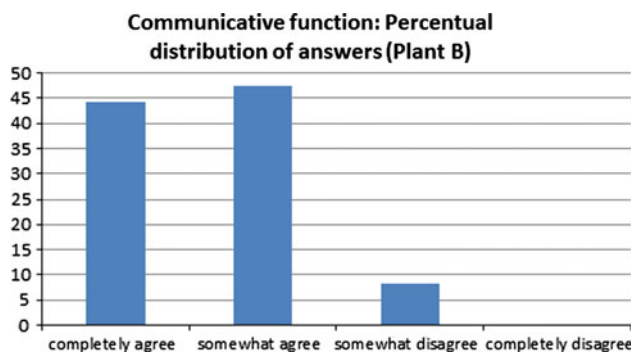**Fig. 9** Communicative function according to questionnaire data in plant A

Fig. 10 Communicative function according to questionnaire data in plant B

4.6 Feedback to the plants and summary of the results

The detailed results were fed back to the plants in various forms. Several workshops were held as part of class room training of the operators. All the operators who participated in the experiment took part in the workshops. In the feedback sessions, all the results and their meaning were discussed thoroughly. Also the trainers and other human factors responsible personnel participated in the discussions. Separate sessions were also held for the personnel responsible for the design of the control rooms.

The systems usability problems which were discovered in the experiment with the current control room systems were delivered back to the participants and the affiliated training and design organizations. In addition, improvements suggestions were made (Table 4) concerning the UI and related training.

5 Discussion

The reference evaluation of two hybrid control room solutions combined a micro level qualitative analysis of operator practices and quantitative methods originating from a more objectivistic evaluation tradition. In addition, operators' experiences of the system were exploited in the overall synthesis of the results. The combination of the different approaches to understand how the tool functions in an activity complies with the requirements stated by Harrison et al. (in press) in their quest for HCI as a successor science, which would put analytic, objectivistic and hermeneutic methods into dialogue in the evaluation (and development) of UIs. The evaluation approach has been labeled and characterized as contextual because it is based on understanding the activity, the part of which the tool is, as a context for usage and evaluation.

Hybrid nature of a control room is one phase in the long evolution of control room UIs which are often in use every day of the year, 24 h a day, in industrial plants. The activity theoretical approach to evaluation suits well this frame as it takes into account the developmental trends prevalent in the activity. To complement this aspect of activity, theoretical approach an analysis was conducted concerning the design philosophy utilized in the respective plants and the historical development of the control room into its current states. Also the future modernization plans were clarified prior to the evaluation (Laarni et al. 2006).

The hybrid nature of the current control rooms is a product of a process in which different technologies have been implemented to an existing control room. Implementing a new part to a functioning safety–critical system can pose a threat to the prevailing state and thus threat safety. Hybridity of the control room was in this paper considered as an example of a process of decrementalism (Dekker 2011), which may cause drift in the overall socio-technical system. We agree with Dekker's point and see that appropriate control room evaluations are necessary whenever even small changes are implemented. Evaluations should enable catching even the small changes which are induced in the activity. In the paper, we have shown that there are challenges concerning the adequacy of current evaluation methods in catching the small changes in activity. We proposed that it is important to complete presently available methods with new ones that consider the underlying mechanisms which produce the performance in the system. This means adopting a multi-method evaluation approach.

In the evaluation approach described and demonstrated in this paper, the evaluated tool is approached through three distinct functions of a tool: instrumental, psychological, and communicative. The functions complement each other and thus are not independent nor can they be analyzed totally separately. The aim in using them is to understand profoundly what the role of tools in constructing, shaping, and developing an activity is.

The instrumental function, especially in the performance sense resonates well with the evaluation approaches based

Table 4 Summary of the results concerning problems in systems usability and the related improvement suggestions to the plants

Function	Issues raised in evaluation/related perspective to activity	Improvement suggestion
Instrumental	Recovery from input errors and mis-operations/performance	In process control it is not possible to recover from an input error, the strategy should be to avoid them by, for example, visualizing the effects of the intended operation to the user prior to execution of the operation
	Cumbersome and awkward positions and movements required to operate/way of acting	Increase terminals: make information available in the physical location where it is needed
	Not clear whether commands have been implemented in the digital interface/performance	Improve user feedback in the system, for example, visualize implementation of commands
	Not easy to determine which phenomena is important/way of acting	Development of alarm handling and presentation
Psychological	Slow and difficult learning process/way of acting	Develop training methods, combine theory and practice of training
		Make sure there is consistency in the operating logic across all control room systems
	Skills in using features of digital interface are not sufficient/way of acting	Develop presentations to match the operating needs in the situation
		Develop training of the tool
	Differences in way of acting among the crews, especially in the amount of information sources exploited/way of acting	Develop process overview information presentation
		Increase training on process couplings
Communicative	Confirmative professional orientation/way of acting	Train operators in interpretive practice
		Develop UI solutions to support reflection in action
	Variation in practices of collaboration/way of acting	Add collaborative features/support for collaboration to the UIs and procedures

on the formal scientific method such as typical human factors evaluation. But by adding the measures of way of acting and user experience even the instrumental function is understood more broadly. The quality of way of acting is contextual. What might be good acting in one context is not necessarily that in an another. The analysis of the core task in the particular activity must be made in order to see whether the tool supports good way of acting.

The psychological function brings to the evaluation the fact that tools and technologies shape in a profound way the ways of working. When changes are made in the tools it is important to consider the implications on the activity. The new tools should develop the capabilities and be in harmony with the preferences of people concerning the development of their work. It is also necessary to include the training of users as early as possible in the design, so that feedback to design can be collected and there is enough time for learning new skills. The usability and HCI methods have represented this view point since 1980s and 1990s but the design of complex technical environments sets new challenges to implement the good principles in the design process. In analyzing way of acting, in considering the psychological function of the tool, it is possible to evaluate whether the tool supports users' developing understanding of the complexities of the controlled process. In this sense, using the tool makes the operator realize intricacies concerning the process

and thus develop the capabilities of controlling it. In addition, the tool should be such that operators have an experience of a tool that suits their practice.

The communicative function views the tool as a medium which is supposed to create, carry, and distribute meanings across the socio-technical system. This approach suites well with the so called the third wave HCI which has claimed to view design as meaning making (Ylirisku et al. 2009). The way in which the communicative function was operationalized in the above studies will be developed further by considering the extent to which the object is shared by the collaborative team. In our on-going work, we are applying the typology of collaborative actions proposed by Raeithel (1983) that draws on the ideas of AT.

By utilizing the three general functions that a tool has in an activity system, it was possible to gain profound understanding about which aspects of the tool benefit the activity and which still need development. The methodology is quite heavy to use as it relies on extensive modeling of the domain and meticulous analysis of usage activity. Nevertheless, it is possible to utilize the tool function approach also in studies that cannot utilize the same amount of resources as is appropriate in safety–critical domains. The three tool functions have been introduced to increase the understanding of how profoundly and from many perspectives technology affects human activities.

6 Conclusion

In this paper, a contextual evaluation approach for the evaluation of tools in safety–critical work has been outlined. The method has also been demonstrated by two case evaluations in NPP domain in the evaluation of the hybrid control rooms. In the evaluation of the appropriateness of the control room, a variety of data collection and analysis methods were utilized. The results were harmonized using the concept of systems usability which refers to the three general function of a tool in an activity system: instrument, psychological tool, and a communicative tool.

References

Bannon L (1991) From human factors to human actors. In: Greebaum J, Kyng M (eds) Design at work: cooperative design of computer systems. LEA, Mahwah, pp 27–44

Béguin P, Rabardel P (2000) Designing for instrument-mediated activity. Scand J Information Systems 12:173–190

Bødker S (2006) When second wave HCI meets third wave challenges. NordiCHI. doi:10.1145/1182475.1182476

Braarud P-O, Skraaning Jr G (2006) Insights from a benchmark integrated validation of a modernized NPP control room: performance measurement and the comparison to benchmark system. American Nuclear Society, NPIC&HMIT, Albuquerque

Burkhardt J, Détienne F, Hébert A, Perron L, Safin S, Leclercq P (2009) An approach to assess the quality of collaboration in technology-mediated design situations. ECCE 2009, Helsinki, pp 30:1–30:9. Retrieved from http://portal.acm.org/citation.cfm?id=1690508.1690551

Carroll JM (1997) Human–computer interaction: psychology as a science of design. Int J Human-Computer Studies 46(4):501–522. doi:10.1006/ijhc.1996.0101

Dahl Y, Alsos OA, Svanæs D (2010) Fidelity considerations for simulation-based usability assessments of mobile ICT for hospitals. Int J Human-Computer Interaction 26(5):445–476. doi:10.1080/10447311003719938

Daniellou F (2005) The french-speaking ergonomists' approach to work activity: Cross-influences of field intervention and conceptual models. Theoretical Issues in Ergonomics Science 6(5):409–427. doi:10.1080/14639220500078252

Dekker S (2011) Drift into failure—from hunting broken components to understanding complex systems. Ashgate, Burlington

Dewey J (1999) The quest for certainty. A study of the relation of knowledge and action (finnish translation). Gaudeamus, Helsinki

Dourish P (2001) Where the action is—the foundations of embodied interaction. MIT Press, Cambridge

Engeström Y (1987) Learning by expanding: an activity-theoretical approach to developmental research. Orienta-Konsultit, Helsinki

Gay G, Hembrooke H (2004) Activity-centered design: an ecological approach to designing smart tools and usable systems. MIT Press, Cambridge

Harrison S, Sengers P, Tatar D (in press) Making epistemological trouble: third-paradigm HCI as successor science. Interact Comput. doi:10.1016/j.intcom.2011.03.005

Hart SG, Staveland LE (1988) Development of a multi-dimensional workload rating scale: results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Human mental workload. Elsevier, Amsterdam, pp 139–183

Huang E, Mynatt E, Trimble J (2006) Displays in the wild: understanding the dynamics and evolution of a display ecology. In: Fishkin K, Schiele B, Nixon P, Quigley A (eds) Pervasive computing. Springer, Berlin, pp 321–336

Hyysalo S (2002) Transforming the object in product design. Outlines, Critical Social Studies 4(1):59–83

Kaptelinin V, Nardi B (2006) Acting with technology: activity theory and interaction design. MIT Press, Cambridge

Kestenbaum V (1977) The phenomenological sense of John Dewey: habit and meaning. Humanities Press, Atlantic Highlands

Klemola U-M, Norros L (2001) Practice-based criteria for assessing anaesthetists' habits of action: outline for a reflexive turn in practice. Med Educ 35(5):455–464. doi:10.1046/j.1365-2923.2001.00894.x

Klemola U-M, Norros L (2002). Activity-based analysis of information characteristics of monitors and their use in anaesthetic practice. ECCE 2002, Catania

Koski-Jännes A (1999) From addiction to self governance. In: Engeström Y, Miettinen R, Punamäki R-L (eds) Perspectives on activity theory. Cambridge University Press, Cambridge

Kuutti K (1991) The concept of activity as a basic unit of analysis for CSCW research. ECSCW 1991, Amsterdam, pp 249–264

Laarni J, Norros L, Rinttilä E (2006) Control room modernisation from an evolutionary point of view—experiences gained at the loviisa NPP. Workshop on Future Control Station Designs and Human Performance Issues in Nuclear Power Plants, Halden

Leontjev AN (1978) Activity, consciousness, and personality (finnish translation) Kansankulttuuri Oy, Helsinki

Lepreux S, Abed M, Kolski C (2003) A human-centred methodology applied to decision support system design and evaluation in a railway network context. Cogn Technol Work 5(4):248–271. doi:10.1007/s10111-003-0128-9

Lützhöft M (2004) "The technology is great when it works": maritime technology and human integration on the ship's bridge. Dissertation, Linköping University

Määttänen P (2009) Toiminta ja kokemus (action and experince). Gaudeamus, Helsinki

Mattsson S (2011) Measures of collaboration in CSCW: usability and collective measures in remote and co-located problem-solving. Master's thesis, Chalmers University of Technology

McCarthy J, Wright P (2004) Technology as experience. MIT Press, Cambridge

Merleau-Ponty M (1986) The phenomenology of perception. Routledge & Kegan Paul, London

Nachreiner F, Nickel P, Meyer I (2006) Human factors in process control systems: the design of human–machine interfaces. Saf Sci 44(1):5–26

Neale DC, Carroll JM, Rosson MB (2004) Evaluating computer-supported cooperative work: Models and frameworks. In: Proceedings of CSCW 2004. Chicago, IL, pp 112–121. doi:10.1145/1031607.1031626

Nickel P, Nachreiner F (2008) Evaluation of presentation of information for process control operations. Cogn Technol Work 10(1):23–30. doi:10.1007/s10111-007-0074-z

Norman DA (2006) Logic versus usage: the case for activity-centered design. Interactions 13(6):45-ff. doi:10.1145/1167948.1167978

Norros L (2004) Acting under uncertainty. The core-task analysis in ecological study of work. VTT, Espoo

Norros L, Klemola U (1999) Methodological considerations in analysing anaesthetists' habits of action in clinical situations. Ergonomics 42(11):1521–1530

Norros L, Nuutinen M (2005) Performance-based usability evaluation of a safety information and alarm system. Int J Hum-Comput Stud 63(3):328–361. doi:10.1016/j.ijhcs.2005.03.004

Norros L, Salo L (2009) Design of joint systems: a theoretical challenge for cognitive systems engineering. Cogn Technol Work 11(1):43–56

Norros L, Savioja P (2007) Towards a theory and method for usability evaluation of complex human-technology systems. Activités 4(2):143–150

Norros L, Liinasuo M, Hutton R (2011a) Evaluating the potential of new technological tools for safety critical work. Interact Comput 23(4):299–307. doi:10.1016/j.intcom.2011.05.003

Norros L, Liinasuo M, Savioja P (2011b) Operators' conceptions of procedure guidance in NPP process control. In: Proceedings of EHPG. Halden, Norway

O'Hara J (1999) A quasi-experimental model of complex human-machine system validation. Cogn Technol Work 1(1):37

O'Hara J, Pirus D, Beltratcchi L (2003) Information display: considerations for designing modern computer-based display systems (no. EPRI-1002830). EPRI, Palo Alto

Park J (2009) The complexity of proceduralized tasks. Springer, London

Peirce CS (1991) Peirce on signs. In: Hooper J (ed) Writings on semiotic by Charles Sanders Peirce. University of North Carolina Press, Chapel Hill

Peirce CS (1998a) The Harvard lectures on pragmatism. In: T.P.E. Project (ed) The essential Peirce. Selected philosophical writings. Indiana University Press, Bloomington, pp 133–241

Peirce CS (1998b) The Peirce edition project. Introduction. In: T.P.E. Project (ed) The essential Peirce. Selected philosophical writings. Indiana University Press, Bloomington, pp 133–241 (pp XXX)

Raeithel A (1983) Tätigkeit, arbeit und praxis. Campus, Frankfurt am Mein

Randell R, Wilson S, Fitzpatrick G (2010) Editorial—evaluating new interactions in health care: challenges and approaches. International Journal of Human-Computer Interaction 26(5):407–413. doi:10.1080/10447311003719847

Roto V, Law E, Vermereen A, Hoonhout J (2011) User experience white paper - bringing clarity to the concept of user experience. http://www.allaboutux.org/files/UX-WhitePaper.pdf. Accessed 1 December 2011

Rückriem G (2003) Tool or medium? The meaning of information and telecommunication technology to human practice. A quest for systemic understanding of activity theory. Paper Presented in FISCAR '03. Espoo

Rückriem G (2009) Digital technology and mediation: a challenge to activity theory. In: Sannino A, Daniels H, Gutiérrez KD (eds) Learning and expanding with activity theory. Cambridge University Press, New York, pp 88–111

Salo L, Norros L, Savioja P (2009) Using operating procedures in NPP process control. In: Proceedings of ECCE 2009, pp 258–266

Savioja P, Norros L (2008) Systems usability—promoting core-task oriented work practices. In: Law E, Hvannberg E, Cockton G (ed) Maturing usability. Quality in software, interaction and value. Springer, London, p 123

Savioja P, Norros L, Salo L (2008) Evaluation of systems usability. Proceedings of ECCE 2008, pp 110–117

Schraagen J, Militello L, Ormerod T, Lipshitz R (2008) Naturalistic decision making and macrocognition. Ashgate, Burlington

Stanton N, Salmon P, Jenkins D, Walker G (2010) Human factors in the design and evaluation of central control room operations. CRC Press, Boca Raton

Stevenson D, Hutchins M, Smith J (2010) Human-centered evaluation for broadband tertiary outpatient telehealth: a case study. Int J Hum Comput Interact 26(5):506–536. doi:10.1080/10447311003719979

Twidale M, Randall D, Bentley R (1994) Situated evaluation for cooperative systems. In: Proceedings of CSCW 1994. Chapel Hill, North Carolina, pp 441–452. doi:10.1145/192844.193066

Upton C, Doherty G, Gleeson F, Sheridan C (2010) Designing decision support in an evolving sociotechnical enterprise. Cogn Technol Work 12(1):13–30. doi:10.1007/s10111-008-0124-1

Vicente KJ (1999) Cognitive work analysis. Toward a safe, productive, and healthy computer-based work. Lawrence Erlbaum Publishers, Mahwah

Vygotsky LS (1978) Mind in society. The development of higher psychological processes. Harvard University Press, Cambridge

Yliriseku S, Halttunen V, Nuojua J, Juustila A (2009) Framing design in the third paradigm. In: Proceedings of the 27th international conference on human factors in computing systems, Boston, MA, USA, pp 1131–1140. doi:10.1145/1518701.1518874