STUK-YTO-TR 204 / APRIL 2004

USABILITY EVALUATION OF COMPLEX SYSTEMS

A literature review

Leena Norros, Paula Savioja
VTT Industrial Systems

In STUK this study was supervised by Jukka Kupila

STUK • SÄTEILYTURVAKESKUS
STRÅLSÄKERHETSCENTRALEN
RADIATION AND NUCLEAR SAFETY AUTHORITY

Osoite/Address • Laippatie 4, 00880 Helsinki
Postiosoite / Postal address • PL / P.O.Box 14, FIN-00881 Helsinki, FINLAND
Puh./Tel. (09) 759 881, +358 9 759 881 • Fax (09) 759 88 500, +358 9 759 88 500 • www.stuk.fi

The conclusions presented in the STUK report series are those of the authors and do not necessarily represent the official position of STUK

ISBN 951-712-845-2 (print)
ISBN 951-712-846-0 (pdf)
ISSN 0785-9325

Dark Oy, Vantaa/Finland 2004

NORROS Leena, SAVIOJA Paula (VTT Industrial Systems). Usability evaluation of complex systems. A literature review. STUK-YTO-TR 204. Helsinki 2004. 44 pp.

Keywords: complex sociotechnical systems, validation and verification, usability

Abstract

In this literature study the task of usability evaluation is approached through recent publications and known practises. The scope of the reviewed literature and the overall problematic of evaluating complex systems are first defined. Then existing evaluation methods and tools and a set of design guidelines and standards is introduced. In the review the distinction between the evaluation of the design process and evaluation of the outcome of the design process is made. Also the existing criteria for usability is provided. These criteria can be used as measures in the evaluation procedures described in the previous chapter. Then, the problem of validity in the evaluation and validation procedures of complex cognitive systems is shortly discusses. The final chapter covers the conclusions of the literature review.

There is a need for deepening the relationship between usage and design. Usage must be the driver also in the design of process technology, plant's safety, and instrumentation and control as well as that of the design of man-machine interface. The collaborative activity is called activity-oriented design. In order to enable activity-oriented design we need to create design methods that drive the design activity into this direction. Criteria for evaluating what is good design are not available in present standards. Criteria for good tools and practices must consider human in a favourable role in the various tasks of operating an industrial plant as opposed to the view of human as a source of error. This is the way to create an optimal system, which takes advantage of the unique capabilities of both the human operators and the automation systems.

Foreword

This research report covers the literature review conducted in research project IDEC (Interaction Approach to the Design of Control Rooms). IDEC project aims at formulating a scientifically founded method for the evaluation of human-system interfaces of complex industrial systems. In the project a method and a set of evaluation criteria will be created.

IDEC project is part of the SAFIR 2003–2006 Finnish public research programme on nuclear power plant safety. SAFIR programme is administrated by a steering group that has been nominated by the Ministry of Trade and Industry (KTM). This specific research has been funded by Ministry of Trade and Industry, Teollisuuden Voima Oy (TVO), Radiation and Nuclear Safety Authority of Finland (STUK) and Technical Research Centre of Finland (VTT). The authors would also like to thank the personnel of Fortum for good co-operation.

Contents

1 Introduction

This research report is a literature review concerning human factors evaluation methods used to assess and ensure the appropriateness of complex sociotechnical systems. By appropriateness we denote usefulness of the system as a tool in an intended activity. Later in this report appropriateness will be elaborated by introducing different notions of usability. In usability evaluation the characteristics of a specific system are examined. This is usually done for the purpose of enhancing the usability of the system but the evaluation can also serve the purpose of validating the design and implementation of the system. In this report we have divided the different evaluation methods to focus on either the design process with which a particular design outcome was produced, or the features of the outcome, the actual system.

1.1 Complex systems

This report concerns the human–technology interaction (HTI) of complex sociotechnical systems. With complex sociotechnical systems we refer to a set of interrelated system characteristics, which Kim Vicente presents in the beginning of his book "Cognitive Work Analysis" (Vicente 1999). The list consists of several system characteristics, which according to Vicente increase the demands laid on both, the system users and the system designers.

According to Vicente (1999) the system characteristics, which make a certain system complex and sociotechnical are for example:

- The number of relevant factors that the designer of the system must consider is enormous.
- Part of the system use is social interaction among the users and the system has many heterogeneous user roles.
- The system is distributed geographically and involves users in different locations.
- The system is dynamic in nature and there is a high degree of potential hazard in operating it.
- The system consists of many coupled and highly automated subsystems.
- There is a degree of uncertainty to the data that the system provides its users.
- The users of the system cannot directly observe whether the goals of the system have been achieved.
- The users have to deal with unanticipated events and disturbances of the system.

In our specific research project called IDEC (Interaction approach to development of control rooms) the case system is the control system of a nuclear power plant. IDEC research project aims at formulating a scientifically founded approach to the development of human–technology interaction in control rooms and operating centres of nuclear power plants. In the project a set of criteria that may be utilised in the validation of the design process and its result will be created. An interaction perspective, the so-called ecological approach is adopted in the construction of the criteria. We argue that artefacts should not be evaluated independently of the practices of their use. Thus, in connection to defining the criteria for a good control room and control system interface we also must consider the criteria for good practices of process control. Methods and tools for the comprehension of the user demands and actions, on the one hand, and means of translating and representing the demands in the evaluation metrics of the HTI of complex systems, on the other, are the central research problems in IDEC research project.

1.2 Structure of the report

In this literature study the task of usability evaluation is approached through recent publications and known practises. Chapter 2 defines first the scope of the reviewed literature and the overall problematic of evaluating complex systems.

Chapter 3 concerns the existing evaluation methods and tools. Also a set of design guidelines and standards is introduced. The chapter is divided into two main sections: Evaluation of the design process and evaluation of the outcome of the design process. In chapter 4 the existing criteria for usability is provided. These criteria can be used as measures in the evaluation procedures described in the previous chapter. Chapter 5 shortly discusses the problem of validity in the evaluation and validation procedures of complex cognitive systems. And chapter 6 covers the conclusions of the literature review.

2 Problem scope: human factors in design

Human factors is a human-centred approach to technology. It originates in the need to exploit scientific–technical innovations for the human tool-using activities. It aims at improving the appropriateness of the technological solutions (Hancock & Chignell 1995). Innovations in technology enhance the shared human intelligence and the possibilities for human to interact with the environment. Human performance may basically be considered from two points of view in design. Traditionally human behaviour has been perceived from the perspective of causing risk to the proper and safe functioning of the system. The other perspective emphasises the positive contribution of human performance for productivity and safety. In normal flow of actions, the favourable role of human performance is not evident but it becomes overt in unexpected situations. The fact that we still have pilots on board of commercial aircrafts even though technological prerequisites for unmanned flying are considered being available, is an example of exploiting the positive effect of human behaviour. In concert with Bernard Papin we see that if design only focuses on the minimising of the human contribution to risk this would sooner or later lead to putting the human operator aside the responsibility for control (Papin 2002). Therefore, in the design of technology it is necessary to balance between the favourable and negative effects of human intervention in the design of complex systems.

Lisanne Bainbridge drew attention to the tendency that the need for exploiting scientific knowledge of human behaviour emerges from the inherent paradoxes of the technological development (Bainbridge 1987). She used the notion “ironies of automation” that denotes the fact that automation does not free human operators from the control of natural processes but instead strengthens the role of human operators. The “ironies of automation” create design problems such as ensuring situation awareness and monitoring of relevant features of a process by the human operator, who was replaced by a machine that was considered a more accurate controller, or to problems of maintaining operational skills without possibilities to have operational experience. Bainbridge also argued that complexity of the process is not necessarily a negative feature from the human actor’s point of view. She referred to empirical results that made evident that high complexity, possibilities to have an effect on the process and coherent representation of process information reduce the level of mental strain and enhance well being (Bainbridge 1987).

Zuboff also dealt with the internal contradictions of development of new technologies (Zuboff 1988). She made an important distinction between the role of technology to automate and to informate. The development of technology is usually comprehended from the point of view of its capability of replacing human labour by automation. Yet, by improving measurements, and transimission, copying and representation of information new technology also increases the role of information as the mediator between the human actor and the object of activity. This complicates the traditionally intuitive and indirect interpretation of the state of the environment and calls for improved intellectual skills. Hence, it is the new technology’s ability of to informate that paradoxically emphasises the role of the human when the level of automation increases, as Bainbridge wrote.

2.1 Human factors research traditions

As we have indicated elsewhere (Norros et al 2003) the human factors research that is relevant for human-centred design of complex systems emerged from two lines of research. Around the 1940s the early engineering psychology developed into the research domain called ergonomics. Ergonomics and later cognitive ergonomics is an interdisciplinary research field that focuses on optimising the functioning of the human–technology interaction with regard to safety, efficiency and human well being. This is accomplished by taking into account the strengths and weaknesses of human performance. The development of industrial automation and the extensive use of IT in transport systems (especially air transport) and process industries reshaped work demands and types of workload on human operators (Leppänen & Norros 2002). This created the need for understanding human cognitive processes and mental load. These issues are addressed in the field of cognitive ergonomics that aims at improving human performance in complex production or transport systems, which often have high safety relevance.

The other line of human factors research was initiated in the late 1970s. It was labelled human–computer interaction research, or computer psychology, because it dealt in particular with the use of computers and software systems. The characteristic context of use was the office but office work was hardly considered in the analyses of the use of the computer. The approach focused first mainly on individual person's choices, which were affected by the usability of the tools and the user's preferences. The significant innovation in computer domain was the graphical user interface (GUI). The emergence of GUI was essentially based on taking into account the human factor as it was recognised that the computers were performing better if they are adapted to the way people think (Kirwan 2003, Norman 1986).

The computer supported co-operative work (CSCW) is a tradition in the computer psychology that has had a major impact in redefining the conception of human performance. The CSCW tradition has a divergent theoretical background but is characterised through a strong ethnomethodological orientation (Bannon 2002). Methodologically, CSCW draws from phenomenology but also from the cultural historical theory of activity (Bannon & Kuutti 2002, Leont'ev 1978, Vygotsky 1978). Lucy Suchman convincingly demonstrated the drawbacks of interpreting human conduct as being a sequential course action that results from following a predetermined action plan, as the prevailing information processing approach in psychology maintains (Suchman 1987). The author suggested the conception of situated action that acknowledges the constructive situation specific structure of action and claimed that plans are but weak resources of situatively constructed action. The emphasis on the wider context of use in the CSCW tradition contributed in bridging the gap between the computer psychology tradition and cognitive ergonomics research (Bannon 2001). The latter line of research has traditionally been more interested in the context in which human performance takes place and it has interpreted computers as information and control tools that mediate human–environment interactions.

Today, the two traditions are merging, mainly due to advances in information and communications technology (ICT). ICT provides a global infrastructure for all human activities in the knowledge society. The new human factors tradition may be considered as a science of design (Carrol 1997) because it considers human factors both from the point of view of the design and of the product in use. We have recently adopted a new term "human–technology interaction" (HTI) to acknowledge the significance of both earlier traditions for the further development of technologies for both professional and everyday activities of the knowledge society (Norros et al 2003).

2.2 From interface design to deeper levels of system design

One of the important design challenges in improving the control of complex environments is the development of new ways of information presentation. This challenge was seen already by Bainbridge and Zuboff who predicted that in the facilitation of human performance in the control of complex processes the traditional automation concepts should be aided by further solutions. Today, the implementation of digital automation and visual display units in the control centres has caused the predicted changes in the layout and forms of representation of process phenomena. However, more fundamental changes in the paradigms of representation are still under study, design, and prototyping. Task-oriented, functional or ecological information presentation approaches are currently under intensive study and design (Bye 2003, Pirus 2002, Saarni et al 2002a, Saarni et al 2002b, Saarni et al 2002c).

Relevant research questions in the design of information presentation were recently listed by Kirwan (Kirwan 2003). These topics are presented in the following, adapted list in which we also have indicated some recent studies in each problem area:

- Design for error recovery by improving error tolerance, error detection and correction (Kaarstad & Ludvigsen 2002, Wioland & Amalberti 1996)
- Support team work to improve reliability by e.g. promoting monitoring actions (McNeese et al 2001).
- Improve possibilities for situation assessment by e.g. enabling direct perception and comprehension of connections and integrating information (Papin et al 2003, Rasmussen & Pejtersen 1995, Woods 1995, Yamaguchi & Tanabe 2002)
- Improve prediction and anticipation by simulations and predictor displays and alarm systems (Saarni et al 2002b).
- Faciliate focusing on process and problem-driven performance by developing function and task-oriented representation of information (Pirus 2003a, Vicente 2002).
- Create transparency of automation and trust in automation by facilitating navigation in information systems, and by providing adequate feedback and improving reliability (Hollan et al 2000, Lee & Moray 1994, Strand 2002).

When more innovative solutions become necessary in the improvement of human–technology interaction, human-centred design principles must be applied even in the earlier and deeper levels of design of the system. Hence, for example designers involved in the design of the future nuclear power plant concepts see that the new plants should comply with stronger safety and productivity constraints than what are expected from the present plants or their modernised versions. This challenge cannot be reached by only technical improvements but, instead, the contribution of knowledge of human performance must be integrated in design (Papin 2002). Papin maintains, further, that both risk and favourable effects of human behaviour need to be included when utilising human factors in design. However, he does not consider realistic that engineers would take the human factor problems into account if these are detected very late in the design. Therefore it is most important to anticipate these problems in connection with very early design choices. When referring to design Papin does not only denote the design of man–machine-interface, which is the currently dominant level of human–technology interaction with regard to which human factors problems are solved. For Papin the human-centred design is a principle according to which human factors knowledge is taken into account also with regard to decisions concerning much "harder" issues. Figure 1 demonstrates his vision to human-centred design of nuclear power plants.

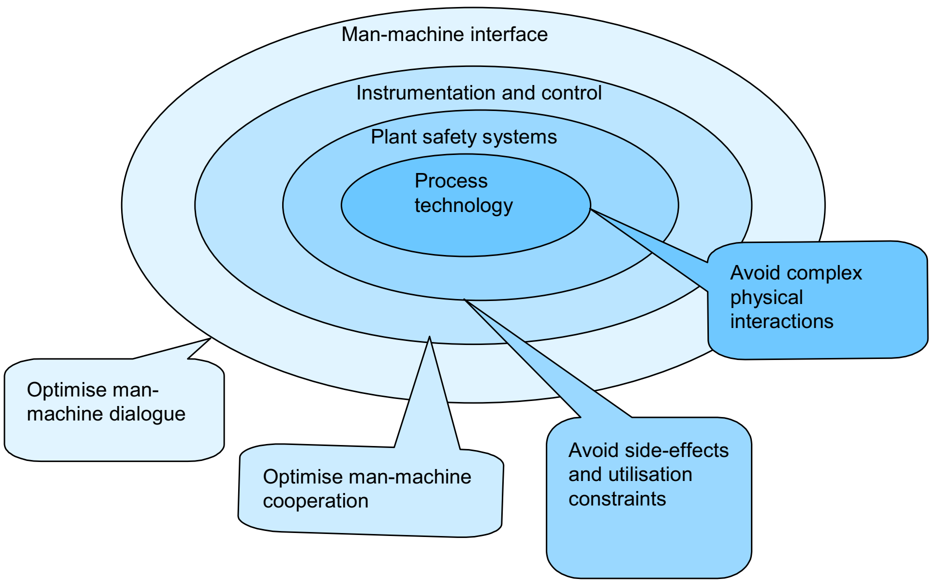

Figure 1. The multi-layered design process ad associated human factors objectives in nuclear power plant design (Papin 2002).

Involving human factors early in the design process and with regard to deeper levels of the system puts the challenge on the human factors community to define *synthetic* requirements and criteria for design. These criteria should be apt to be put in balance with techno-economical factors that usually orient the designers' decisions (Papin 2002). The author provides some ideas of the sought synthetic criteria. He maintains that they should be technologically determined but performance shaping features of the particular domain. Papin does not refer to the ideas of Vicente or Rasmussen in this connection but the line of thinking corresponds very well with the design approach expressed by these authors (see below) (Vicente 1999). According to Papin, complexity and time constraints are two important constraints for human behaviour that, at least, should be included as synthetic criteria for optimising sociotechnical systems.

2.3 Design challenges in NPP control rooms

2.3.1 Control room generations

Currently, extensive work is accomplished world wide in the modernisation and modification of the existing nuclear power plants. These plants represent design concepts that are usually labelled as the I & II generation designs (O´Hara 2003, Pirus 2003b). Early prototypes and the majority of commercially operating plants belong to this category. They have been taken into operation between 1950–1995.

Some III generation plants are already in operation, others are still under design or construction. Examples of generation III plants are EdF N4 in France, the Tepco ABWR in Japan, the European EPR600. The Westinghouse System 80 refers to a generation III control room concept. A further plant generation III+ denotes plant concepts that are designed to include some evolutionary changes in the process. These changes aim at improving safety and efficiency of power production. The generation III+ plants are expected to be in production around 2010. Advanced pressurised Water Reactor 1000 (AP1000), European Simplified Boiling Water Reactor (ESBWR) or Pebble Bed Modular Reactor (PBMR) are examples of the near future developments.

The IV generation has a perspective to be operational in 2030. Novel process solutions are designed for this plant concept. These are targeted to better sustainability, minimal waste, high economical efficiency, enhanced safety and proliferation resistance (O´Hara 2003).

2.3.2 Digitalisation of control rooms

The changes in control room technologies are considered necessary due to technical, availability and economical reasons that endanger the maintaining of old technologies. The improvements in the informative features of new automation are attractive because they may improve operators' process management and facilitate integration of operational control and other personnel, especially with maintenance (Östlund 2003). The III and III+ control room designs manifest a transition from the traditional analogue instrumentation and control technology to digital technology. Most important consequences of this transition with regard to the HTI are (O´Hara 2003, Pirus 2003b):

- The interaction with the process will take place via compact workstations and overview displays via which the new control principle called "soft control" is made available
- Large amounts of data may be integrated and made available for the operators
- The representation of process information is hierarchical and access to information is sequential
- Operator aids increase including improved alarm management, computerised procedures, computerised operator support systems (COSS)
- Restricted implementation of Virtual Reality solutions, but intelligent technology in further sense is limited.

2.3.3 Demands on operator actions

It is realistic to consider that new technical tools induce changes to the joint human–technology system and to the distribution of cognitive and operational activities of the cooperative system (Hollan et al 2000, Norros et al 2003). In the following we shall list some of the major challenges that the new control room designs put on the work and competencies of the operators, including cooperation within the team and in the organsation.

Challenges on operational control:

- Large amounts of process information and an

abundance of alarm information in disturbance situations challenge the operators' comprehension of the process state and its expected course in operational situations. The term "situation awareness" is often used for this complex of activities (Endsley 1995, Endsley & Garland 2000, Östlund 2003)
- The space for information presentation is small which necessitates a sequential use of information (the key whole effect) (Woods 1995)
- The parallel presentation of information changes the relationship between what is within the focal awareness of the operators and what information is within the subsidiary awareness. This changes the conditions for the formation and exploitation of tacit knowledge that is connected to understanding the meaning of subsidiary non-focal knowledge (Polanyi 1974)
- "Soft control" is expected to increase secondary tasks which may increase mental effort and work load (Pirus 2003b, Woods 1995)
- Feedback from the system must be optimised. Too fast and two small feedback may increase mental load (O´Hara 2003)
- Systems may fail, understanding the functions of automated systems must be ensured (O´Hara 2003)
- Identification, correcting and recovery from errors may become more difficult which endangers the reliability of the sociotechnical system (O´Hara 2003).

Challenges on cooperation:
- Transition to "soft control" diminishes the horizon of observation, the openness of the tools and the openness of interaction (Hutchins 1990).
- The new information tools change the propagation of information and decision making in the distributed cognitive system (Hollan et al 2000).
- Soft control requires conscious development of cooperation and communication practices (Hollan et al 2000).

The above-mentioned changes take place in the context of the nuclear power plant organisation. The values of the organisation and the shared expectations and preferences of the users have an effect on the acceptance of the changes and the actual implementation of the tools in usage. At the same time the new tools change the usage and organisational cultures. Therefore, when analysing and evaluating changes in tasks and work contexts we need methods that are capable of analysing human performance from a cultural and societal perspective. Suggestions for promising methodological approaches have recently been provided by various authors who all share the idea of understanding human activity as a culturally formed practice that is shaped in an adaptive tool-mediated interaction with the environment (Bannon 2001, Gauthereau 2003, Hutchins 1995, Norros submitted 2003, Woods 1995).

2.3.4 Human factors evaluation

One of the largest modernisation projects was recently accomplished in the Oskarshamn 1 Nuclear power plant in Sweden (Östlund 2002, Östlund 2003). One part of this modernisation was the upgrading of the control room. To minimise risk for human errors in this, and other modernised control rooms in the near future, the Swedish nuclear regulatory authority (SKI) developed a process-oriented comprehensive inspection method. The investigators of Brookhaven National Laboratory, J. O´Hara and J. Higgins, aided SKI in this task (O´Hara & Higgins 2002). The method used by the investigators is documented in the NUREG reports 0700 (O´Hara et al 2002a) and 0711 (O´Hara et al 2002b).

Brookhaven National Laboratory (BNL) is one of the leading research institutes in the area of human factors and cognitive engineering of complex environments. The U.S. Nuclear Regulatory Commission has exploited the shared competencies of BNL in providing guidelines for their human factors regulatory work. The above mentioned NUREG-reports present guidelines for human-system interface design and for the review of human engineering programmes in the nuclear industry but these documents also benefit other complex high-reliability organisations.

The present human factors evaluation activities focus mainly on the first outer layer of the multi-layered design model suggested by Papin (Figure 1). This applies also the NUREG reports. The VTT IDEC validation methodology – to the construction of which this literature review promotes – is also targeted to evaluations concerning the first and second design layers, i.e. man–machine interface, and instrumentation and control. This focusing is due to the immediate need to accomplish evaluations for control room upgradings, and for the evaluation of new control rooms that represent generation III or III+ plants. As we indicated above, these plant concepts are not foreseen to have major process changes and therefore design decisions do not concern the deeper layers indicated in Figure 1.

One of the benefits of the NUREG-guidelines is the construction of comprehensive system models that summarise the design relevant human factors issues and connect them to the course of the design process. The model depicted in Figure 2 summarises the human factors evaluation tasks in a design process (O´Hara 2003). This model is also included in a slightly different version for structuring the NUREG-0711 guidelines.

The human factors evaluation (HFE) activities depicted in Figure 2 are decomposed to five major areas, i.e. planning, human-system interaction design, training, verification and validation, and, performance monitoring. As Figure 2 indicates, it is assumed that human factors competence should be used in all these activities and that there are two types of human factors knowledge and methods that should be made available. The first type is knowledge that has been integrated into design guidelines and standards. This type of knowledge has a great value in ensuring that available scientific knowledge and experience is synthesised in a practically usable form. However, it is evident that, from the design point of view guidelines, and especially standards, tend to be outdated because they refer to existing technology and it takes time to reach a shared opinion about complex issues (O´Hara 2003, Pirus 2003b). Therefore, also other further human factors expertise is needed. Human factors evaluation analyses indicated in Figure 2 refer to to various kinds of studies, experimenting, analysis etc. that focus on understanding human performance in complex environments. The activities usually require considerable effort from human factors specialists, and they have close connection to research on human factors issues. One of the central methodical issues t is the development of proper task analyses. We shall return to this point in chapter 3.2.2.1 of this report.

The advantage of the above-referred NUREG guidelines is not only the providing a holistic

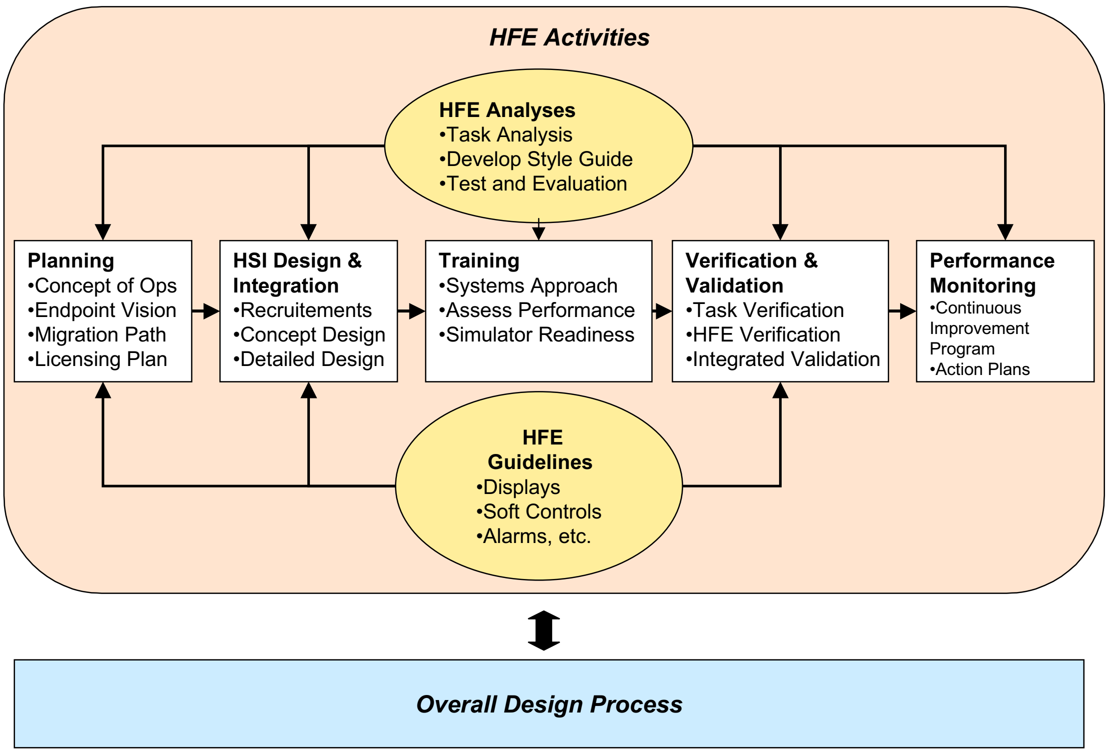

Figure 2. Human factors evaluation and the relation to design process according to John O'Hara (2003).

picture of the human factors evaluation issues. The other strength of the reports is that they introduce a great amount of issues or objects on which the evaluation should focus on. The principle used in the reports is to define what should be looked at. The evaluation consists of checks that a particular human factors issue, for example *concept of operation* (NUREG-0711) has been considered, or in the case of specific design elements such as dimensions of certain displays fulfil an acceptable level. The latter types of criteria relate to the physical features of the control room design. The very short (3 p.) appendix of the NUREG 0700 provides the design review principles and refer to the operator tasks (O´Hara et al 2002a). Underlying generic conceptions of human behaviour and possible ways of evaluating what may be considered “good” with regard to the operators’ practice or culture are not included. As a consequence, final criteria for evaluation of neither the interface nor training or any other human factors evaluation issue cannot be expressed. This problem was identified by O´Hara (O´Hara 2003), and it is one of the most difficult questions of the discipline. We see further, that the usability research tradition that originated in the human-computer interaction research, and developed further in the evaluations of the user interfaces of everyday consumer appliances has addressed this problem by introducing the concept of usability. Our target in the IDEC-project is to make use of the usability research tradition. Our aim is to integrate it to our own core-task oriented usability approach. We also see that the work by Papin and his colleagues (Papin 2002) has great relevance to defining core-task oriented usability criteria. These take position to what is good operator action and appropriate tools for this action, and the definition of good practice has connections to the synthetic constraints of the domain.

3 Evaluation of HTI and the methods and tools used in it

Riihiaho (2000) states that traditionally, there are two kinds of methods to evaluate the usability of a given product or system: usability inspections, and empirical user testing. The major difference between these two is that inspection methods can be applied without direct user involvement whereas user testing always involves actual users as participants. Also CNSC (CNSC 2003) and most of the evaluation standards (ISO , IEC, NUREG) divide evaluation methods into two categories according to their nature: verification and validation methods (V&V). This categorisation is similar to that of the traditional usability evaluation because verification means a process of demonstrating that equipment and system have been designed as specified, which can in most cases be done without direct user involvement. Validation, on the other hand, demands users as participants. Validation means the process of determining the degree to which the human–machine system design and supporting mechanisms facilitate the achievement of operational goals of the system. (CNSC 2003.)

In addition to evaluating the usability of the actual system, also the design process, in which the system being evaluated was produced, can be evaluated. The methods and tools used are of a wholly different nature when the evaluation is directed at the design process.

In the forthcoming sections we have divided the evaluation methods into two categories: ones that concern evaluation of the design process and ones that concern the outcome of the design process. Within these two categories there are two separate viewpoints: compliance to the standards and guidelines and overall performance. Typically compliance to the standards is evaluated with different inspection methods and performance is evaluated with functional tests and integrated system validations. This classification is done based on reviewed literature (Nicic 2003, O´Hara 2003, O´Hara 1999).

3.1 The Design Process

With the design process we refer to the way the system under evaluation is produced. The design process incorporates the life cycle-model of the system, the tools used to build the system and the individuals building the system.

The evaluation of the design process is one way to approach the quality of the system being examined. By evaluating the design process it is possible to state whether the various usability issues were considered at the appropriate level during the process of system design. Several standards and guidelines exist for design process evaluation. The standards reviewed in this report are either nuclear industry specific, or concern human-centred design on a more general level.

Common to all the standards and guidelines is that they lay out the separate phases of the process and all the activities relevant in each phase. This information can be used as a checklist to make sure, that all the necessary phases of the process have been carried out. But information about *how* each phase should be carried out and *what* are its consequences for the next phase is not specified in the standards.

3.1.1 Compliance with the standards and guidelines

In the following sections we introduce standards that can be used in the evaluation of the design process.

3.1.1.1 Human factors engineering program

The report NUREG-0711 (2002) lays out the human factors engineering (HFE) principles that need to be considered in a modernisation project of a nuclear power plant. The report gives twelve different phases of the design process in which HFE issues need to be considered. These are:

1. HFE Program Management
2. Operating Experience Review
3. Functional Requirements analysis and Function Allocation
4. Task Analysis
5. Staffing & Qualification
6. Human Reliability Analysis
7. Human-System Interface Design
8. Procedure Development
9. Training Program Development
10. Human Factors Verification and Validation
11. Design Implementation
12. Human performance monitoring.

The report explicates the aspects of these different design phases that need to be included in the design process but it does not give any acceptance criteria, which would verify the quality with which the phase was carried out. NUREG-0711 (2002) can be used as a checklist to make sure that all the relevant HFE issues are dealt with in the design process.

3.1.1.2 Design Process for Control Rooms (CR) of Nuclear Power Plants

The standard IEC 9064 (IEC-9064 1989) sets fundamental requirements for the process of designing the main control-room of a nuclear power plant. The standard does not give guidance on how to carry out the design process in practice but in the appendix of the standard there is some supplementary information in the form of a design guide. This guide includes points of philosophy and methodology, explanations and detailed recommendations and specifications for the control room design process. (IEC 1989)

The standard begins addressing the issue of NPP control room design by stating the functional objective of the main control room (CR) which is to provide the operator with accurate, complete, and timely information regarding the functional status of plant equipment and systems. On a more abstract level the overall objective of the main control room is to provide a space and the equipment for safe and efficient operation of the plant in all the operational states and accident conditions. The main control room serves the purposes of the plant's operational goals. In addition, it provides an environment under which the CR staff is able to work without discomfort, excessive stress, or physical hazard. (IEC 1989)

The human factors engineering principles that should be taken account of in the design process are: anthropometric, perceptual, cognitive, psychological, and motor response. (IEC 1989)

Functional design of the CR should be based on a system-based approach. This is achieved by applying a top–down approach in the design. According to IEC 9064 (1989) the following four steps should be included in the process:

1. **Function analysis** is an analysis of the objectives of the NPP, and the hierarchical structure of the goals
   - identification of functions
   - information flow and processing requirements.

2. **Functional assignment** is done to determine which functions should be performed by man, and which by machine. It should consider:
   - identification of operator capabilities
   - I & C system processing capabilities.

3. **Verification and validation of function assignment** is performed to show evidence that a proposed function assignment takes the maximum advantage of the capabilities of man and machine without imposing unfavourable requirements on either of them. Modifications of the functional assignment and the verification of it shall be made iteratively until all criteria are met. The validation of functional assignment is performed to demonstrate that the functional goals can be achieved.

4. **Job analysis** of the verified function assignment is performed to clarify the content and requirements of the operators' jobs.

The steps 1–3 are covered more thoroughly in the standards IEC 61839 and IEC 1771 (IEC-1771 1995, IEC-61839 2000).

Standard 9064 (IEC 1989) also pays special importance to the process of verification and validation of the man/machine interface. The verification of the control room system design is a normal design review – a part of quality assurance procedures – which assures the correct incorporation of functional specifications and other technical information into the proposed control room system concept.

- The process of verification & validation shall include preparation, evaluation and resolution phases. The evaluation shall include the operating procedures and training programme
- The control room system integration shall incorporate all the functional specifications and all other technical requirements correctly
- A dynamic simulator is necessary for the validation

3.1.1.3 Functional Analysis and Assignment in the Design of NPP CRs

The standard IEC 61839 (2000) specifies functional analysis and assignment procedures for the design of the control-room system for nuclear power plants. The standard also gives rules for developing criteria for the assignment of functions. The standard is a complement to the standard IEC 9064 (IEC 1989).

The standard (IEC 2000) specifies how to carry out the functional analysis and assignment of functions referred to in the standard IEC 60964. The process of functional analysis is the first step in the design of a control-room. It aims initially to identify all of the functions required to operate the plant, then to assign the functions to humans or to machines.

The basic technical team for the FA and A should include the following areas of expertise (IEC 2000):

- Nuclear and non-nuclear systems engineering
- Systems analysis
- I & C systems design
- Information and computer systems design
- Human factor engineering
- Plant operation
- Development of normal operation and emergency procedures.

The functional analysis process includes the identification of functions and the identification of information flow and processing requirements. The process begins with identification of functions. The functions are derived from the fundamental objectives of the plant operation. Then the top level functions are broken down into a hierarchy of functions where the lowest set of functions are the control functions which are then assigned to humans or to machines. (IEC 2000)

In the functional analysis the identification of control functions needed is based on general nuclear power plant decomposition. The decomposition is obtained by presenting the results of the overall plant design in a hierarchical manner, with the plant operational goals. These goals are developed further as subgoals producing a hierarchical goal structure. The identification of functions follows immediately from the identification of goals. (IEC 2000)

The designer shall subdivide each of these functions successively and develop a set of rules to identify when the hierarchical analysis is completed to a sufficient level of detail. The bottom level functions shall form a complete set, which is itemised and stated in functional terms. Some iteration could be necessary to reach the final level of functions. The resultant hierarchy will have the functional goals at the top, system-level functions in the middle and detailed control functions at the bottom level. The bottom level control functions are later in the process assigned either to humans or machines. (IEC 2000)

The next step in the process is the identification of basic information and processing requirements needed for the accomplishment of each control function. This is done through:

- Individual function analysis.
- Identification of time requirements and representative events of all operational sequences, all design basis events given in the safety analysis report and beyond design events like core fusion, steam explosions etc.

Functional assignment is the distribution of functions between the human and the automated constituents of the system. Assignment of functions to humans means achieving them by manual control, monitoring, high-level mental processing, or their combinations. Assignment of functions to machines means achieving them by automation. In the process of functional assignment the control functions must first be methodologically grouped. The suggested groups are: must be automated, are better automated, should be given to humans, and should be shared. After the categorisation all the actions needed for the accomplishment of the functions are defined in detail. The last step is the identification of the typical measurements of the functions. (IEC 2000)

3.1.1.4 Verification and validation of the design of the main control room of a NPP

The standard IEC 1771 (IEC-1771 1995) specifies verification and validation (V&V) procedures for the design of the control-room system of nuclear power plants and gives verification and validation criteria for the assignment of functions and for the integrated control-room system. The standard can be applied to a renewal project of an existing control-room or a construction project of a totally new control-room. The standard is a complement to the standard IEC 9064 (IEC 1989).

Development oh new control rooms

In the development of a new control-room design, analysis should be performed in an effort to optimise the assignment of functions (to man or machine) and the tasks that implement the functions. The result of each of the two phases is checked using V&V activities. The purpose of the V&V activities is to assess the adequacy of the interfaces between operator and plant processes as found in the control-room. (IEC-1771 1995)

Each step of the V&V activities should include the following activities (IEC-1771 1995):

- **Preparation** which includes development of evaluation criteria, definition of V&V methodology, identification of source documents, organisation of the evaluation team, definition of workspace and equipment for the evaluation team, and schedule definition for the review.
- **Evaluation**.
- **Resolution**.

In the **verification** of the **function assignment** evidence shall be presented that a proposed function assignment takes the maximum advantage of the capabilities of man and machine without imposing undesirable requirements on either of them. (IEC-1771 1995)

In the **validation** of the **function assignment** the control-room system is demonstrated to be capable of achieving all the functional goals. The validation shall also demonstrate that, if required, in case of loss of automated functions, manual back-ups are available. (IEC-1771 1995)

Verification of the **integrated control-room system** shall be performed by evaluating the proposed design specifications against the applicable functional requirements, design requirements and criteria. (IEC-1771 1995)

Validation of the **integrated control-room system** shall be performed to prove that the system can achieve the performance intended. Special attention shall be given to the time dependant dynamic characteristics of the integrated system. The validation should asses the suitability of the design to support the interactions between: control-room and the operator; control-room and the operating procedures; control-room and the training program; operator and the other staff inside and outside the control-room. (IEC-1771 1995)

Development of existing control rooms

Evolutionary designs are those, which introduce modest changes when, compared to predecessor control-rooms. The design aspects that should be taken into consideration when deciding whether the change is evolutionary are: display and control man–machine interface, alarm presentation, operational methods and general response of the control-room man–machine interface to plant upsets. The degree of innovation in the new design sets the criteria whether the change is an evolutionary one or not. (IEC-1771 1995)

Pre-existing materials can serve as acceptable substitute for portions of the V&V activities in the case of an evolutionary design. Such material can be i.e. information on design principles, operating procedures, training material and operating experience and control-room design review information. The applicability of such material depends on their quality and the extent of differences between the new design and it's predecessor. (IEC-1771 1995)

- **The verification of function assignment** may not need to be an extensive top-down analysis for evolutionary designs. Any changes in human–machine function assignment and their integration with other functions shall be verified though. (IEC-1771 1995)
- In **the validation of the function assignment** for the areas of change of function assignment, it shall be shown that the time-

dependant characteristics of functions and their assignments are in accordance with human physical and cognitive capabilities. Potential overload conditions should be identified and tested. (IEC-1771 1995)

- **The verification of the integrated control-room system** shall concentrate on the areas of change and their integration in the control-room. In the process all the relevant design requirements shall be shown to have been satisfied. (IEC-1771 1995)
- **The validation of the integrated control-room system** shall concentrate on the new design areas. Also the interaction between the new and the old design is adequate. The validation is done through execution of all plant procedures and it should take place in a control-room training simulator. (IEC-1771 1995)

3.1.1.5 Application of visual display units (VDUs) in the main control room of a NPP

The standard IEC 1772 (IEC-1771 1995) presents design requirements for the application of VDUs in main control-rooms of nuclear power plants. The standard is a complement to the standard IEC 9064 (IEC 1989).

In the beginning of the design process the goals of the display system shall be identified. These can be i.e. safety, availability, and operability. (IEC-1772 1995)

In the design process special attention shall be paid to the relationship between the information to be presented and any associated controls. The design of the VDU system shall develop and document a clear definition of the intended purpose of the displays, their safety role and their basic performance requirements. (IEC-1772 1995)

The principal users shall be identified for each group of VDUs. These may be the reactor or other plant operators, the operation supervisor, maintenance staff or management. Normally the most important users to consider are the operators in the main control-room. The design targets should be to enhance the operators' role towards that of a safety and performance optimiser. (IEC-1772 1995)

The information to be displayed shall be defined in principle and then in detail by analysis of the operators' and other users' need for information in different operating conditions. (IEC-1772 1995)

The requirements for a display or suite of displays shall be determined with a thorough and systematic analysis of the proposed use of the data being displayed. For each proposed item of information the designer shall take into account the following principles (IEC-1772 1995):

- number of users
- purpose of use => required reliability
- are comparisons with other data required
- when and how often and how quickly the data is required
- the accuracy needed
- are errors in interpretation acceptable
- the degree of detail or abstraction which is required
- the time of an event which causes an important transient.

The verification of a VDU-based information system should be carried out for a well-specified set of operational state data including abnormal states and fault conditions. In the verification special attention should be given to ensure consistency in situations where variables are displayed at several locations at the same time. (IEC-1772 1995)

The validation of a VDU-based information system should be carried out defining representative scenarios of operation, disturbed situations or accident conditions and information goals for different users of the system. (IEC-1772 1995)

3.1.1.6 Ergonomic Design of Control Centres

The standard ISO 11064 parts 1,2,3 handle the ergonomic design of control centres and in particular the arrangement of control suites and control-room layout (ISO-11064 2000). The standard covers all types of control centres typically employed for process industries, transportation and logistic control systems, and people deployment services.

The first part of the ISO 11064 series specifies ergonomic principles, recommendations and requirements to be applied in the design of control centres, as well as in the expansion, refurbishment and technological upgrades of control centres.

The first part of the series is organised in nine separate principles (ISO 2000):

1. **Application of human-centred design approach;** in which the combination of humans and machines, in its organisational and environmental context, is considered as an overall system to be optimised.
2. **Integrate ergonomics in engineering practice**; which is carried out by organising the design project in such a way that an integration of technical and ergonomic expertise is encouraged.
3. **Improve design through iteration;** in the design process evaluation shall be repeated until the interactions between the operators and designed objects achieve their functional requirements and objectives.
4. **Conduct situational analysis;** Existing or similar situations shall be analysed so that the functions of the future system can be thoroughly understood and anticipated beforehand. Situational analysis should include task analysis, operator interviews and incident analysis.
5. **Conduct task analysis;** which considers all modes of system operation including start-up, normal operation, shut-down, anticipated emergency scenarios, periods of partial shutdown for maintenance, the results used in the design process and the development of staffing plans. The task analysis methods may vary according to the scope and content of each individual project.
6. **Design error tolerant systems;** although human error cannot be totally eliminated it is still important to strive for error-tolerant design.
7. **Ensure user participation;** user participation throughout the design process is essential to optimise long-term human–machine interaction by instilling a sense of ownership in the design. Experienced users can offer valuable empirical contributions to the control centre design.
8. **Form an interdisciplinary design team;** to oversee and influence all phases of the design project. Team may include system and process engineers, ergonomists, architects, and industrial designers.
9. **Document ergonomic design principles;** to reflect the ergonomic design basis of the project. This can include for example reasoning or significant task analysis findings. The document should be updated whenever there is a change. An appropriate procedure should be developed for this process.

The standard (ISO 2000) introduces a framework for the design process. In the framework the control-centre design process is divided into five dependent phases. Typically all phases should be executed with the overall effort distributed in accordance with the scope of the design project.

The recommended framework involves following phases:

1. **Clarification** is carried out to clarify the purpose, context, resources and constraints of the project when starting the design process. In the clarification the existing situations, which could be used as references are taken into account.
2. **Analysis and definition** produce a document of control centre's functional and performance requirements culminating in a preliminary functions allocation and job design.
3. **Conceptual design** is carried out to develop initial room layout, furnishing designs, displays and controls, and communications interfaces necessary to satisfy the needs identified in the analysis.
4. **Detailed design** develops the detailed design specifications necessary for the construction and/or procurement of the control centre, its content, operational interfaces and environmental facilities.
5. **Operational feedback** should be collected in a post commissioning review to identify successes and shortcomings in the design in order to positively influence subsequent designs.

3.1.1.7 Human-centred design processes for interactive systems

The standard ISO 13407 provides guidance on human-centred design activities throughout the life cycle of computer-based interactive systems (ISO-13407 1999). It mainly addresses the planning and management of human-centred design activities.

The standard ISO 13407 (1999) states that the incorporation of a human-centred approach is characterised by the following principles:

- The active involvement of users and a clear understanding of user and task requirements.
- An appropriate allocation of function between

users and technology.
- The iteration of design solutions.
- Multi-disciplinary design.

There are four human-centred design activities that should take place during a system development project:
1. To understand and specify the context of use.
2. To specify the user and organisational requirements.
3. To produce design solutions.
4. To evaluate designs against requirements.

The human-centered design process should start at the earliest stage of the project (e.g. when the initial concept for the product or system is being formulated), and should be repeated iteratively until the system meets the requirements, as illustrated in Figure 3.

3.1.2 Performance

With performance of the design process we refer to the quality with which the design process is carried out. Performance evaluation differs from the compliance viewpoint in the sense that the former is only concerned with whether certain phases of the design process were carried out and certain documents produced. When the performance of the design process is evaluated it is actually evaluated with what quality the phases were carried out. We want to say that it is not enough just to say that a requirement phase has been conducted. It is also necessary to evaluate whether the actions taken during the phase were appropriate and whether the results of the phase were of good quality.

Jokela has written about the evaluation of the performance of user-centred design (UCD) process (Jokela 2001). He suggests three different types of criteria for the performance: quantity, quality, and integration. Quantity refers to the extent to which UCD processes are carried out where as quality refers to the quality with which they are carried out. Finally, integration refers to the extent with which the results of the UCD processes are subsequently integrated into the product design.

SPICE (ISO/IEC-15504-5 1998) also deals with the process performance. The viewpoint of SPICE is different though. SPICE defines process performance as:

"The extent to which the process achieves the process outcomes by transforming identifiable input work products to produce identifiable output work products. As a result of full achievement of this attribute
- *the scope of work to be performed and work products to be produced are understood;*
- *work products will be produced that support the achievement of the process outcomes"*

The difference between Jokela's (2001) process performance indicators and those of SPICE (ISO 15504) is that the latter consider the level of process performance good, if all the phases of the design process are just carried out and the expected outputs are produced. Jokela, on the other hand, says that it is equally important to evaluate the

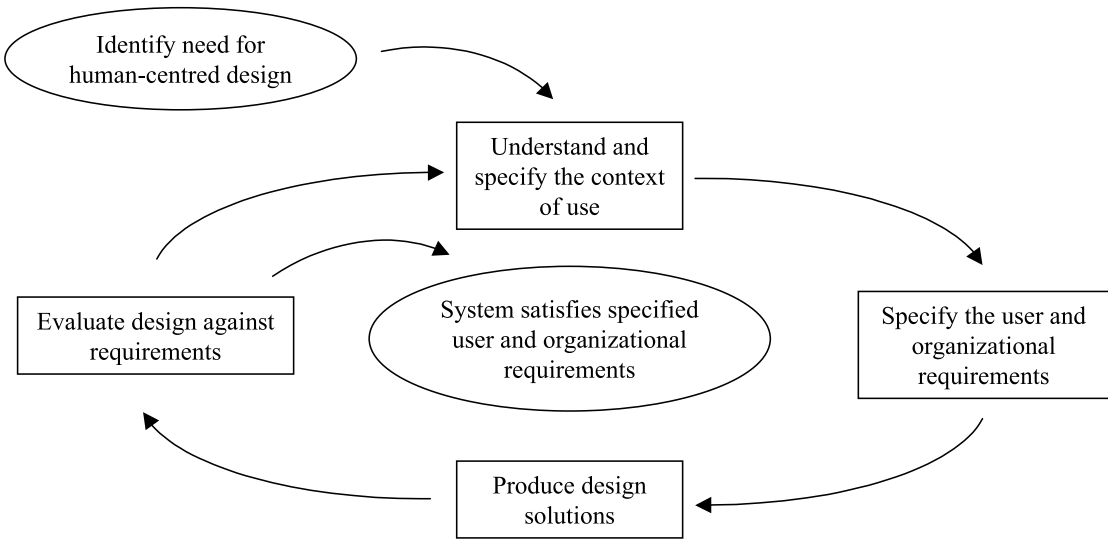

Figure 3. ISO 13407 (1999) human-centred design activities and their interrelations.

quality of the outputs and the way they are used in the subsequent phases of the design process.

As Jokela (2001) suggests that the quality of the design decisions should also be under evaluation in the process of evaluating performance of the design process, he gives some criteria for good quality. These are mainly related to the human-centredness of the decisions made. We have interpreted that Papin (2002) discusses same types of quality attributes of the design process when he denotes that the designers should use synthetic requirements in the design process. Papin suggests that these kind of synthetic requirements should at least be to always try to *reduce the complexity* of design and always try to reduce the effects of *time constraints* of the system on the operators' activities. We would like to add one more synthetic requirement. This would be to always try to reduce the *uncertainty* in the system.

3.2 Outcome of the design process – the control room

3.2.1 Compliance with the standards and guidelines

The outcome's compliance with the standards and guidelines is evaluated in a usability inspection. The usability evaluation of the actual system (versus the design process) can be divided in two: user testing and usability inspection. The difference between these two is that in usability inspection no direct user involvement is needed because the inspection is carried out by user interface experts who are trained to notice typical usability problems in various kinds of systems.

Of the two methods user testing is considered somewhat more important, yet users are sometimes hard to find and it takes time to organise the tests. Therefore in order to find the most obvious usability problems some usability inspections are typically performed prior to the actual user testing. Inspection methods do not require extensive preparations and can be fairly easily applied and integrated into a development process even before any prototypes are prepared. (Riihiaho 2000)

In usability inspections different evaluators find different usability problems. Therefore adding the number of evaluators enhances the amount of problems detected. However the first five evaluators can typically detect the majority of the existing usability problems that can be found in an inspection. (Riihiaho 2000)

3.2.1.1 Nuclear industry specific design standards

There are also some nuclear industry specific standards and guidelines that can be used as usability inspection tools. The problem with the use of these standards is that they don't explicitly specify what needs to be designed to accomplish the requirements. For example IEC 9064 (1989) states that "The (data acquisition and processing) system should be designed such that the system faults do not cause any unsafe state or unacceptable economic losses in the plant operation." This is an essential requirement, but it leaves the designer with very many questions. Also in validating the design of the system the evaluator has to make the decision whether some specific requirement can be accomplished the way the designer suggests.

The standards and design guidelines work well as checklists for the designers on what aspects to take into account in the design. The designer can read the checklist and see which ones of the requirements are relevant. An interpretation of the relevant requirements needs to be made after the requirements are identified. It is up to the evaluator to decide whether the interpretation that the designer has made is adequate.

The most important nuclear industry specific design guidelines and standards are:

- IEC 9064 Design for control rooms of nuclear power plants (1989) specifies the functional design requirements for the control room system and equipment that perform the assigned monitoring and control functions. Also specifies the interface between the man and the control room equipment.
- IEC 1772 Nuclear power plants – Main control room – Application of visual display units (IEC-1772 1995) specifies requirements the visual display units used in the control room of nuclear power plants. The standard is quite specific in for example stating the desired sizes of specific display figures etc. A failure of an information system means that the information is degraded and not sufficient and precise enough to understand or perform a safety task

properly. Failure situations are taken into account in the design by using redundancy.

- In the screens and touch panels for information and control which may have safety relevance (such as those necessary to perform actions according to safety related procedures inside and beyond design) some redundancy should be provided. This is to ensure that a single component failure in the system does not prevent operation of its general function.
- In the screens and touch panels necessary for safety, such as those of a dedicated safety, panel redundancy and diversity of information and control means should be used.
- All the displays should be as simple, clear and comprehensible as possible. Where complex or highly detailed displays are necessary, good organisation and structure are required. Where safety criteria require raw, unprocessed or accident resistant data to be presented in addition to the processed information, the display organisation and identification shall differentiate between these types of information.
- Where colour is used with safety significance, other kinds of coding (position, symbol shape or text) shall be used to ensure that safety significance can be clearly noticed by the operators without reliance to colour. In general the use of pictorial display is recommended. Displays shall be designed so as to benefit from humans' ability to make comparisons and detecting contradictions.

- NUREG 0700 is a human-system interface review guideline developed by the U.S. Nuclear Regulatory Commission (NRC). It covers a wide range of issues relevant in the human factors engineering sense. The guidelines are organised into four basic parts. Part I contains guidelines for the basic human-system interface elements (information display, user-interface interaction and management and controls). Parts II to IV contain guidelines for the review process of the human-system interfaces in nuclear power plants.

3.2.1.2 User interface standards and guidelines review

In guideline reviews the usability inspectors check whether the system being evaluated conforms to the given guidelines. Typically the designers have already used the same guidelines in the user interface design, but the guidelines tend to be very general, so the designers might interpret them in a non-optional way and therefore reviews are needed. (Riihiaho 2000)

3.2.1.3 Heuristic evaluation

One of the most common usability guidelines is the Nielsen's set of ten usability heuristics (Nielsen 1993), Table I**.**

Riihiaho (2000) suggests that the output of a heuristic evaluation be a list of usability problems. The list should state the heuristic rule that the problem violates and the severity of the problem.

3.2.1.4 Cognitive walkthrough

Cognitive walkthrough is a task oriented usability inspection method. Its focus is on ease of learning. Cognitive walkthrough is based on a theory of learning by exploration according to which users try to infer what to do next using cues that the system provides. Users do not read manual or want any formal instructions before they start to use new systems. Instead users learn by doing and exploring. (Riihiaho 2000 citing Lewis *et al.*1990, Polson *et al.* 1992, Wharton *et al.* 1994 and Lewis & Wharton 1997)

The method guides the analysts to consider users' mental processes in detail instead of evaluating the characteristics of the actual interface. The method can be used very early in the design process to evaluate designers' preliminary design ideas and hence no running version of the system is required. On the other hand the context of the tasks and the users' characteristics must be well specified so that the analysts are able to consider the users' mental processes. (Riihiaho 2000)

In the walkthrough, the analysts comment on the sequence of actions that the users should execute to accomplish their tasks. The walkthrough should always follow the right sequence of actions, that is the sequence that the designer has planned the user to follow. If problems arise in this sequence, they are recorded but the analysis is continued as if the problem didn't exist. (Riihiaho 2000)

Table I. Nielsen's (1993) usability heuristics.

1. Simple and natural dialogue	Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility. All information should appear in a natural and logical order.
2. Speak the users' language	The dialogue should be expressed clearly in words, phrases, and concepts familiar to the user, rather than in system-oriented terms.
3. Minimise the users' memory loads	The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
4. Consistency	Users should not have to wonder whether different words, situations, or actions mean the same thing.
5. Feedback	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
6. Clearly marked exits	Users often choose system functions by mistake and will need a clearly marked "emergency exits" to leave the unwanted state without having to go through an extended dialogue.
7. Shortcuts	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users.
8. Good error messages	They should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
9. Prevent errors	Even better than good error messages is a careful design that prevents a problem from occurring in the first place.
10. Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, be focused on the user's task, list concrete steps to be carried out, and not be too large.

The cognitive walkthrough method was developed to evaluate the ease of learning of the system. Therefore it can also be used in the case of complex sociotechnical systems to evaluate the parts of the interface that are rarely used and are used in situations where it is not possible to use manuals or written instructions. The use of these parts of the system has similarities with walk–up–and–use systems with which the method has been found the most effective.

3.2.2 Performance

The human-centred approach to design maintains that technology should be comprehended from the point of view of providing tools for human activity (Flach et al 1995). From this perspective, it is evident that the final way of evaluating technology is to assess its appropriateness in the aimed use. In our case the outcome of the design process is a complex information and control system for a safety critical production process that constitutes an entire working environment for a team of operators. In such cases it is clear that the evaluation of performance is a difficult task that requires a lot of effort. The development of methods for the analyses of performance requires research work, the results of which also improve generic understanding of human behaviour. Methods developed in human factors engineering research over the past 50 years provide a reservoir of methods for human performance analysis. These methods are usually designed for experimental research and quasi-experimental studies in simulated environments. Another source of information concerning human performance are the studies that focus on human reliability and on incidents and accidents (Hollnagel 1993, Hollnagel 2002). The recently increased interest in accomplishing studies of operator actions in real working contexts (Cannon-Bowers & Salas 1998, Klein et al 1993, McNeese et al 2001, Norros submitted 2003, Woods 1993, Zambok 1997) is expected to improve the usefulness of human factors studies for practical purposes.

It is not within the scope of this report to review the vast amount of literature of performance analysis in complex environments. We should, however draw attention to the work accomplished within the OECD Halden Reactor Project in the area of performance assessment in simulated laboratory environments, in HAMMLAB (Andresen & Dröivoldsmo 2000). In Halden such analyses are conducted for integrated system validation (O´Hara et al 2002b) purposes, and also for human reliability analyses. These both activities belong to the human factors evaluation activities depicted in Figure 2 (human reliability analysis is not indicated in Figure 2 but is one of the subtasks within the planning phase). Andresen and Dröivoldsmo provide a summary of the HAMMLAB data collection methods, which include verbal protocols, automated registrations from the experimental facility, questionnaires and interviews. They also explain how the acquired data is refined. Very informative is the chapter that explains what measures are used to evaluate human behaviour. The Halden tool kit includes assessments of diagnostic behaviour, erroneus actions, task complexity, performance, situation awareness, and workload.

3.2.2.1 Task analyses as tools for integrated system validation

Task analysis is one of the central methods for performance evaluation. In fact, the term task analysis refers to a large set of methods used for analysis of human behaviour from different perspectives. A most comprehensive account of task analyses methods is provided by Kirwan and Ainsworth who listed and described more than 40 well documented task analysis methods (Kirwan & Ainsworth 1992).

Drawing on a frequent paper by Kirwan (Kirwan 2003) we may state that the most central task analyses methods are the following three types methods:

- Hierarchical task analyses
  Hierarchical tasks analysis is characterised by a top–down analysis of the work process composed of a larger task complex. It consists of decomposing the objectives, tasks, operations and plans. It also defines the tools used in the task. The analysis should describe the task in considerable detail. Data is collected by conducting observations, interviews, walk through/ talk through techniques, work participation and collection of operating experience.

- Link analyses
  Link analysis focuses on interactions in performance. These may consists of human–human interactions and communications among different actors, but may also include interactions with the process or with the tools. The aspects analysed from the data may include the character, density, content of interactions. The results are usually provided in the form of graphical representations. Link analyses have been found useful in studies of complex user interfaces.

- Timeline analyses
  Timeline analysis techniques are divided into vertical a and horizontal analyses. The former focuses on the analyses of personnel issues such as operator tasks, task allocation, or communication. It is focuses on analysis of critical phases of the task flow and some versions are used to evaluate task load. The horizontal analysis is more detailed and focuses on the sequential temporal structure of actions and is used for analysing bottlenecks in task performances.

Kim Vicente accomplished a theoretically oriented analysis of existing task analysis methods (Vicente 1999). He pointed out generic problems in the use of task analysis methods in design. One problem is that the methods focus on the analysis of existing work. The concepts used in the analysis assume the explication of the tools used in the work. When the new tools are finally implemented the task analyses are no more valid because the task is not independent of the tools and has changed with the change of tools (Carroll et al 1991). Vicente notes, however, that the iterative character of design, rapid prototyping methods and use of simulations have, of course, improved the situation, but the fundamental problem of dependence between tools and tasks remains. Therefore, drawing on Jens Rasmussen's work he proposes that the analysis should not only exploit sequential description of actual tasks but, instead, orient to the intrinsic performance shaping constraints of the domain (Vicente 1999). Vicente uses the term "formative" analysis to denote the idea of defining the generic boundaries for action without specifying the actual course of actions, the task.

Norros (Norros submitted 2003) utilised Vicente's analysis and expressed the different types of task analyses in a table. The table was constructed by identifying the focus or object of the analysis and the type of modeling that underlines the analyses (Table II). In the table the formative, domain-centred analysis of Vicente is completed by introducing a further object of analysis, work

Table II. Categorisation of different task analysis methods (developed based on Vicente 1999 (Norros submitted 2003).

	Object of analysis			
	Task			
Type of modelling	**Instruction-based**	**Costraint-based**	**Work domain**	**Practices**
Normative	Defines what *should be done*; utilises an input-output or *sequential* approach instrumental vocabulary	Defines tasks as what *should be avoided* in reaching the result; utilises an *event-oriented* sequential approach and instrumental vocabulary		
Descriptive		Describes *actual behaviours*; utilises event-oriented sequential approach and instrumental vocabulary; *constraints* emerge		
Formative			Defines result-critical *boundaries of action*; utilises vocabulary that orients toward *behaviour-shaping intrinsic work constraint*	Defines habits with regard to actor's accounts of the intrinsic constraints; the *internal good* of the practice may be analysed and *standards of excellence defined*
=> Type of explanation	*Causal explanations* of *actual* realisations of performance		*Functional* explanations based on explicating the *potentials* of performance in a domain	*Understanding* explanations of actions based on the clarification of the reasons for action

practices. Analysis of practices utilises the formative analysis of the domain that defines the affordances of the environment for action. The human actor may take the affordances into account constrained by the availability of appropriate habits. The concept of habit originates in the pragmatist tradition. It allows to represent the task on the generic disposition level. Thus, the actions are not only described as they are realised in a particular situation, but they are also comprehended from the point of view what kind of meaning they signify, and through what kind of potential they provide for action. With this completion the formative task analysis provides a new type of ecological approach to analysis of action.

The formative-habitual approach to modeling task is one of the basic tools in the core task analysis methodology (CTA) developed and applied at the VTT human factors group. At present, the CTA includes tools for the evaluation of practices and organisational culture. In the IDEC-project a method for the evaluation of the usability of complex information tools is developed.

3.2.2.2 Usability testing

Usability testing means gathering of information about the use of systems from particular users who are not involved in the design process. The goal of usability testing is to improve the usability of the system being tested. (Riihiaho 2000)

In general usability testing works so that there is the system or some level prototype of it and potential users of the system. The users are given test scenarios in which some tasks are included. The users perform the tasks and the observers of the test record the users' actions and comments. Typically the users also think aloud in order to help the observers understand why they take the actions that they take. In the analysis phase of the test the observers analyse the data they have gathered and make improvement suggestions based on the data.

The most typical type of usability test is one where the participants are the user, the administrator and a group of observers. The administrator has the responsibility of conducting the test and (s)he for example introduces the system to the user. The administrator also introduces the scenarios and the tasks to the user. If the system being tested is at a prototype level the administrator can also act as the engine of the prototype and with a paper prototype switch the layout sketches according to the states of the system. The observers of the test follow the test preferably from some distance, for example from another room via video transmission or through a glass wall so that their discussion will not disturb the test. These kinds of usability tests are usually carried out in usability laboratories, which have good facilities for audio–visual recording of the test. Users are typically asked to think aloud during the test in order for the observers more clearly to understand the users' actions.

3.2.2.3 Paired-user testing

Paired-user testing is a test setting in which two users try to solve the tasks together. One reason for testing in pairs is to make thinking aloud more natural. (Riihiaho 2000) In paired user testing the users might have their tasks on paper so an administrator doesn't have to be present.

3.2.2.4 Usability walk through

Pluralistic usability walk through is a method, which combines elements from a usability test and usability inspection. A pluralistic usability walk-through session involves participants from three groups: potential users, system developers and usability professionals. Together the participants gather information about the systems usability by inspecting hardcopy panels of the system. All users who try to accomplish given tasks participate in the usability analysis. In the end of the session the whole group discusses the findings they have made.

3.2.2.5 Contextual inquiry

Contextual inquiry (Beyer & Holtzblatt 1998) can also be used as a usability evaluation method. It differs from the above mentioned methods in that it is always carried out in the real context of the system for example users are observed while they perform their normal work tasks. In contextual inquiry the observer must be able to interrupt the users work to ask specifying questions about users tasks that (s)he does not understand. Contextual inquiry is typically used as a data gathering method in a user centred design process and it is a part of a larger framework contextual design. (Beyer & Holtzblatt 1998)

4 Criteria for evaluating complex HTI

In the following the relevant usability and appropriateness criteria found in the literature is presented. We emphasise the criteria that can be used in the evaluation of complex industrial systems.

4.1 Standardised usability criteria

The ISO Standard 9241-11 defines usability as follows: "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (Figure 4). Thus, according to ISO usability criteria are the effectiveness, efficiency and satisfaction, which in an evaluation situation are operationalised into concrete quantitative measures. In this sense the effectiveness is measured as the accuracy with which the users are able to achieve specified goals. Efficiency on the other hand means the resources with which the goals were achieved. Satisfaction refers to freedom from discomfort and the positive attitudes to the use of the product.

The usability of a product or a system is not an attribute of the product alone. A product can have very different levels of usability when used in different context, so the context of use should be clearly defined for the evaluation. The context includes the users, their tasks, their goals, and equipment as well as the physical and social environment in which the system is being used.

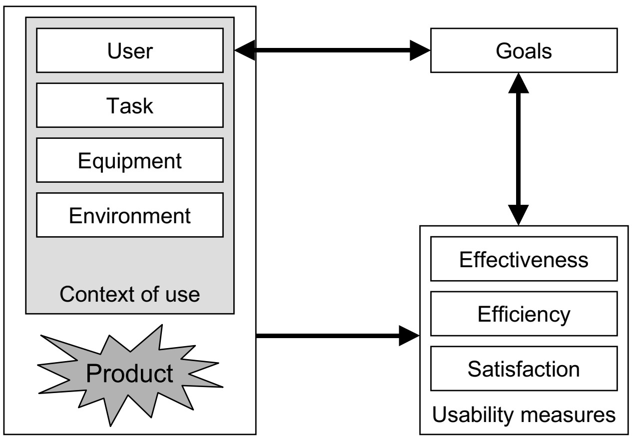

Figure 4. The ISO usability framework (ISO 1998).

4.2 Nielsen's criteria for usability

Jakob Nielsen wrote his famous book "Usability Engineering" in 1993 many years before the ISO 9241 (1998) standard. Nielsen saw usability as a narrower product characteristic than did the experts who wrote the standard. Nielsen (1993) makes a distinction between utility and usability. In his usability definition utility denotes the correct functionality, whereas the ISO 9241 (1998) views the correct functionality as one of the usability measures in the form of effectiveness.

Nielsen (1993) defines usability as a five dimensional quality attribute of the system. These dimensions are: learnability, efficiency, memorability, error prevention, and satisfaction (Figure 5).

Learnability means that a system should be easy to learn so that the user can rapidly start getting some work done with the system. Nielsen (1993, p. 28) states that highly usable products have a steep learning curve for novice users. (Nielsen 1993) In the case of complex industrial systems learnability is not that important of a usability goal though, since the industrial organi-

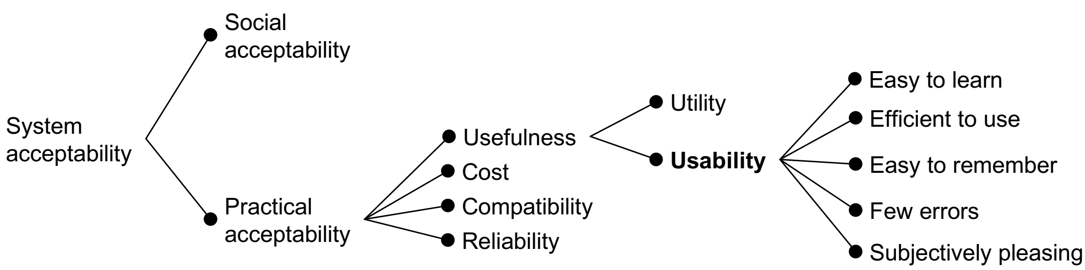

Figure 5. Nielsen's (1993) model of system's acceptability and usability as an attribute of it.

sations typically have the resources to train the users thoroughly.

Efficiency on the other hand is very relevant when it comes to industrial systems. It refers to the expert users' steady-state level of performance at the time when the learning curve flattens (Nielsen 1993).

Memorability refers to the casual users' ability to use the system. Casual are ones that use the system intermittently. The amount of system use is significantly less than would required for real expertise to develop. However in contrast to novices the casual users have used the system previously. (Nielsen 1993.) Memorability is important when considering industrial system because there are many systems that are used only under special conditions from time to time.

It is obvious what error prevention means: There should be a minimum amount of possibilities for the users to take incorrect actions. But if the users anyhow take some incorrect actions this should not lead to any catastrophic situations. This attribute of usability is very relevant with industrial systems.

Subjective satisfaction refers to how pleasant it is to the users to use the system. Subjective satisfaction is more important in a nonwork environment, such as home computing, games etc. (Nielsen 1993)

4.3 Keinonen's criteria for usability

Keinonen (2000) has developed a model of usability (Figure 6) for consumer products but many of the attributes he defines are also relevant with complex industrial systems.

The usability attributes that Keinonen (2000) presents are:

- Functionality (FNC)
- Logicality (LOG)
- Information presentation (PRE)
- User manual, documentation (DOC)
- Usefulness (USE)
- Ease of use (EoU)
- Affect on emotions (AFF).

The attributes described above are connected to either personal preferences of the consumer, or directly to the product, or to its functionality. The first level of usability has to do with how the user perceives the concrete characteristics of the user interface of the product (LOG, PRE, DOC, and FNC). The next level of usability evaluation contains the consumer's feeling about the quality of the interaction between her and the product (USE, EoU). The third level of usability, which is also the most common in use when usability is evaluated, is the interactions general effectiveness on users' emotions. This level is emphasised by consumers' personal aesthetic and value based criteria (AFF). (Keinonen 2000)

The usability attributes presented above are originally created to cover the usability aspects of consumer products. Still some of them are more than relevant also in the case of work-related industrial systems. To our view especially the criteria functionality, logicality, information presentation, and documentation should always be considered when evaluating the interface of industrial systems.

We believe though, that in the future also the attributes PRE and AFF, which refer to the individual users' opinions and attitudes, will also become important when industrial systems are evaluated. They both have a connection to *user acceptance* of the system the importance of which is increasing due to increased level abstraction of in the industrial systems. Abstraction of information systems means that the functionality of the system is no longer apparent or obvious.

4.4 The Halden classification of validation criteria

The OECD Halden Reactor Project (HRP) has over twenty years of experience in the study of human factor issues in the NPP domain. In 1983 an exper-

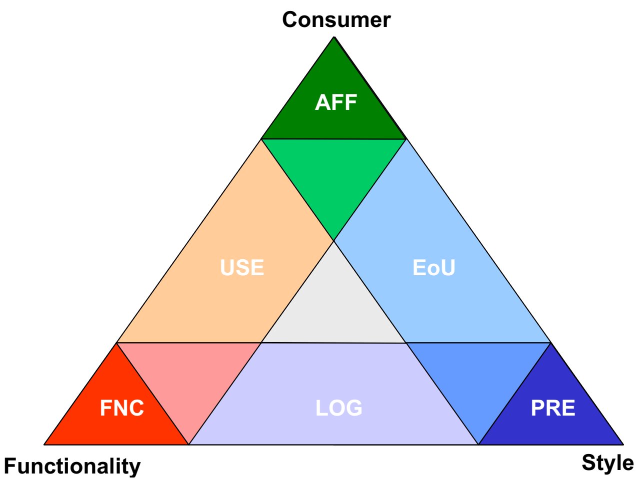

Figure 6. Keinonen's model of the attributes of usability (Keinonen 2000).

imental research facility named the Halden Man–machine Laboratory (HAMMLAB) was established for the study of operator performance in simulated NPP environments. After having started the work with smaller scale NPP process simulations HAMMLAB acquired a full-scope simulation of the Loviisa PWR. The Loviisa NOkia Research Simulator, NORS was used for studies that concerned human performance and human reliability. Different types of operator aids were also tested in HAMMLAB using the NORS simulator. Experts from Loviisa power plant operations supported Halden research staff in the usage of the simulator, and many operator crews from the plant visited Halden and operated the simulated process in experimental studies. A few years ago Halden acquired a more generic simulator facility that is build on the APPROS simulation environment. It is capable of simulating different types of power processes. Three new simulators, i.e. FRESH; HAMBO and PETSHI were established and an Integration Laboratory constructed for improving the testing and implementation of new features to the simulators (Bye 2003, Jokstad et al 2002).

The scientists of the HAMMLAB have developed an experimental research approach that consists of a large arsenal of methods and measures to study human performance in HAMMLAB (HRP 2002). Comprehensive facilities have been constructed for the collection, registration and processing of experimental data (Jokstad et al 2002). The North-American Human Factors tradition has had a major influence on the Halden research approach but more recently it has also gained from the francophone tradition.

Recently, HRP provided a position paper that summarised the current line of thinking in terms of the usability validation requirements. The proposal reflects the work of two sub-programmes of the HRP research programme the "Development of design support bases on HAMLAB data sets" and the "Integrated system validation" (Miberg-Skjerve & Skraaning 2003). The conception outlined in this paper has great relevance to our aims in the IDEC project.

Drawing on the experience in human performance studies in HRP the paper outlines a framework for usability criteria. The criteria are designed for the verification and validation of innovative operational design concepts. The validation and verification process should ensure that the designed system fulfils the design specifications, and that that the system functions as intended. Usability criteria of the existing designs are insufficient, partly invalid and may cause mode errors. This is due to the fact that the implementation of new tools induces changes in the working conditions and in the structuring of operator behaviour. Thus, the aim is to define generic *types of usability criteria* that could guide the design process. In defining the types of criteria the authors refer to many of the above mentioned standards. They draw especially on the human factors work of the Brookhaven National Laboratory (O´Hara et al 2002b).

The development of HRP usability criteria bases on the premises that in the nuclear power domain the validity of a design should, first ensure that the new system is *operational* , i.e. that it contains the desired functionality and a human machine interface that allows the operator to perform the defined tasks satisfactorily. Second, passing a validation test should in this reliability-critical domain also include that a new system should *support the performance of the joint human–machine system*. Moreover, the HRP framework assumes that a validation test should provide evidence of longer-term effects of the new equipment. It should be necessary to ensure that operators not only can but also *will* make use of the tools in their practice. By introducing the last premise the HRP investigators draw attention to motivational and emotional aspects of human behaviour. In other contexts scientists have underlined the significance of this premise by stating that the new equipment should be *meaningful* for the users (Bannon 2002, Mallain 2003).

The above-mentioned premises are reflected in the conception of usability that was adopted by the HRP. This conception consists of three *types of criteria*: *efficiency, effectiveness and satisfaction (interpreted as acceptance)* (p. 3). These types are borrowed from the definition of usability by the ISO standard on Ergonomic requirements for office work with visual display terminals (VDTs) (ISO-DIS-9241-11 1998). *Efficiency* refers to the extent to which the operators are able to operate the system in a safe and reliable manner. Efficiency is measured with regard to the *work process* (situated action) with the aid of various kinds of performance measures. *Effectiveness* denotes efficiency with regard to promoting the productivity targets. This criterion is, thus, considered to refer to the *joint human–machine system and its outcome.*

The HRP validation framework puts much weight on the *acceptance* criterion. The authors draw on the ideas of William B. Rouse who stated that one of the key objectives in the human-centred design is to foster user acceptance by ensuring that the operators are provided with the roles they consider to be satisfactory (Rouse 1991). It is, of course, true that the operators of complex systems seldom have the option not to use a system that is an integrated part of the control room. It is usually expected that the operators adapt to the changes in the tools and in their working environment. The authors state that the risk, however, remains that operators develop informal routines, in which they avoid the non-acceptable features of the system. Consequently, the system is used differently than intended and less comprehensively. Miberg-Skjerve and Skraaning thus suggest that the operator's perception of the designed system could serve as a latent condition that continually will influence the motivation and practice of usage. The user acceptance implies that the operators require the support of the equipment. If operators, who are sufficiently skilled to use the system, do not appear to exploit the system in validation tests it should be inferred that they do not consider the system to be of practical use.

Problems in acceptance may be caused by deficient functionality or interface design. They may, furthermore, be due to phenomena such as failing trust in automation or anxiety for degrading of competencies connected to the introduction of new technologies. It is furthermore underlined that issues of job satisfaction, self-esteem, professional standing among colleagues and individual merit may often be overlooked. All these issues may have an effect on efficiency and safety of practices and should therefore be taken into account in the validation of new equipment.

The validation framework of that the HRP research group proposes consists of four levels (p. 6–7):

- Baseline: This evaluation level refers to the designer's own validation that is intertwined with his/her design decisions during the design.
- Level 1 – User acceptance: This evaluation level refers to the operators' evaluation of the designed system's characteristics and adequacy in relation to the intended functions of the system. The fulfilling of these functions in different conditions are reflected and the operators' concerns of the long-term effects on work. Issues of job satisfaction, competencies etc. may be addressed. Evaluations are based on operators' judgements about the envisioned design solution.
- Level 2 – Benefits to work process: This evaluation level refers to systematic testing of work processes on individual and team work level with realistic mock-ups, simulations or prototype versions of the design. The evaluations focus on the effects on safety. Evaluations consider both performance and user acceptance. Performance measures include physical positioning and anthropometry, cognitive and behavioural demands, use of procedures, alarms handling, operators' comprehension of the situation, out of the loop effects, discrepancy from good operating practices, group processes. User acceptance measures include issues such as long-term motivation, development of competencies, degradation of teamwork skills, trust in automation.
- Level 3 – Benefits to system performance: This evaluation level type refers to the design solutions capability to enhance the overall system performance with regard to productivity, i.e. safety and effectiveness. It should also be clarified whether the operators actually use the system as intended. The validation tests should provide answers to questions whether the physical system state are tolerable, whether productivity standards are adhered to, or whether the operators use the system effectively.

Validation may focus only on single levels or on combinations of these levels. The authors, however, state that to argue that a system is valid in terms of usability a systematic and comprehensive coverage of all levels should be required.

Miberg-Skjerve and Skraaning state that the previous validation studies of the HRP usually concentrate on the levels 2 and 3 with the focus on efficiency criterion. The authors maintain that in other usability studies the focus is on usability on levels 1 an 2 but the level 3 measures concerning the effect on productivity are often missing. These observations lead the authors to consider the connections between the levels of their validation framework.

The question is whether the evaluations by the users (level 1) and the work performance evaluations (level 2) should be connected to the system output in terms of safety and effectiveness (p. 9–10). The authors take the position that in the analysis of standard and well analysed systems, such a link can probably be established. They, however maintain that in the case of innovative systems this is an unreasonable demand. Following arguments were provided to support this conclusion. First, the authors note that system performance measures are insensitive and, therefore, sufficient performance variation cannot be observed. Second, it is impossible to have a representative sample of possible events for the validation tests. Third, optimal human–machine system performance is vague during the design phase and effected by the design solutions itself. Finally, the authors state that system performance effects are immediate whereas the validation criteria used in levels 1 and 2 refer to long term effects.

Corresponding problems of design and validation were brought up by Vicente when he elaborated difficulties to anticipate the context of use and the demands of the user during design (Vicente 1999). The solution for connecting the judgement- and performance-based evaluations of usability to the system performance that he suggested was to use a formative modeling approach to describing the system performance. This approach was initially proposed by Rasmussen (Rasmussen 1986, Rasmussen & Svedung 2000) and elaborated in an instructive way by Vicente in the above referred book. Because it is impossible to predict or define the situation specific sates and performances of the system comprehensively during design, this modeling technique withdraws from using a prescriptive or descriptive vocabulary in defining the system performance (see Table II). Instead, it approaches the system in functional terms and identifies result-critical intrinsic constraints of the domain. They constitute result-critical boundaries that ought to be maintained under all conditions and, thus, they provide generic criteria for the proper functioning of the system. The boundaries for result-critical action should orient the designer in his/her evaluation of the design solutions.

4.5 Core-task orientation as a criterion for system performance

Drawing on Rasmussen, the VTT human factors research group also developed a formative modeling approach for the description of the work domains. Within the methodology named the core-task analysis (Norros submitted 2003, Norros & Nuutinen 2002) the VTT group utilised formative modeling as a basis for understanding human performance and decision-making in complex socio-technical systems. Dynamism, complexity and uncertainty are three generic features of the environment that create challenges for the human actor to cope with it. These features correspond to those that Woods defined as dimensions of the environment that effect coping with the domain (Woods 1988). Woods also included risk to his concept of the generic environmental features. While we agree that risk is significant we would suggest that it should be considered as a further but not equal category as the three others. The distinction of these features provides a generic background for our analysis of the characteristic constraints of the system that constitutes the VTT core-task modeling method.

In core-task modeling we distinguish two levels of modeling of the constraints. First, these may be defined as pertaining to the work domain (NPP operations, anaesthesia, maintenance work etc). The constraints are defined on a generic level with the help of the critical function concept. This concept denotes a decomposition of a complex system into its main result-critical operating functions, which may be fulfilled by different means and physical systems and components (Corcoran et al 1981, Pirus 2003a, Pirus 2003b, Rasmussen 1986, Rasmussen & Svedung 2000).

In his Abstraction–Decomposition Space (ADS) model Rasmussen distinguished two dimensions of decomposing systems, the part whole, and the abstraction level. A causal analysis of system behaviour may be decomposed to the level of detail that is known to the anlyst. The abstraction level dimension refers to the further kind of decomposition. It is based on functional abstraction.

From this perspective relational structures may be distinguished that represent " 'practically isolated relationships' which are valid for a variety of systems, and they have long been considered the only acceptable scientific representation of phenomena. … This type of model does not necessary represent the actual, in–the–world behaviour of the phenomena of interest, but is effective for understanding basic mechanisms and to define limits of performance and conditions for optimal function" (p. 31) (Rasmussen & Svedung 2000). Pirus adopted a compatible method but defines function in a way that is more oriented towards the design practice. Functional analysis denotes decomposing the system to operating function as " a set of additional or redundant means needed to fulfil a precise operation mission in response to the constraints imposed by the environment" (Pirus 2002, Pirus 2003a).

In the VTT core task modeling approach it is further assumed that, as response to the possibly contradicting constraints of the environment human operators must take into account and balance the critical functions in operational actions for achieving the intended outcomes. The critical functions also constraint design as they define the boundaries of the result-critical functioning of the future system. Balancing between the functions constitute the core-task demands of the particular work (Klemola & Norros 1997, Norros & Klemola in press, Nuutinen & Norros 2001, Oedewald & Reiman in press, Reiman & Norros 2002).

The core-task modeling method also includes a second level modeling that focuses on identifying the manifestation of the critical function and the core-task demands in specific situations. Functional situation models include analyses of the available resources (information, operative methods, procedures etc.) from the point of view of their role for the maintenance of the core-task demands of the work in the particular situation (Norros submitted 2003). Functional situation models have been utilised as reference in analyses of operator performance in various contexts and for various purposes. These include analyses of competencies and development of training (Hukki & Holmberg 2002, Hukki & Norros 1998, Nuutinen 2003), human-reliability analysis (Holmberg et al 1999), and the analysis of accidents (Norros & Nuutinen 1999b, Nuutinen & Norros 2001). Watanabe and colleagues have proposed an interesting method for developing event scenarios for nuclear power plant operators' training. In correspondence to the VTT method, also this approach draws on Woods's complexity dimensions (Watanabe et al 2002) but these dimensions are not contextualised with regard to the critical functions as the VTT core task modeling approach does.

As was indicated above the core-task modeling technique was used in connection to the analysis of operator performance. By modeling the core-task and situated constraints it is possible to create the connection for the analysis of behaviour with reference to the human–machine system output. Thus, modeling forms a reference for making inferences of the situated adaptiveness and appropriateness of the operator behaviour. The core-task analysis method includes a further part in which task-performance-related evaluation items are defined and concrete criteria are developed for the evaluation of habits of action. These criteria are anchored with the situational constraints but the deeper evaluation dimensions are drawn from theoretical considerations regarding adaptive behaviour in an uncertain environment (Hukki & Norros 1998, Norros submitted 2003). Drawing first Ewald Ilyenkov's ideas (Ilyenkov 1977, Ilyenkov 1984) and later on pragmatist conceptions of human conduct (Peirce 1998b) we have stated that situationally appropriate behaviour in an uncertain world is characterised by an interplay of doubt and and belief regarding the state of the environment. Continuity in behaviour is accomplished by developing habits that are repeated. By virtue of repetition of habits, actions become situationally adapted and they may develop. The stronger the personal ability and interest to attend a situation as a specific instance and object of knowledge the stronger tendency there is to act in an interpretative way. The weaker this tendency is, the more action tends towards reactiveness and mechanical routines.

Naikar and colleagues have utilised the Cognitive Work Analysis method of Vicente (Vicente 1999) for designing team tasks (Naikar et al 2002). Corresponding to the use of the Core Task Modeling by the VTT research group Naikar and colleagues developed first situation models on the basis of generic functional description of the domain. Thereafter, instead of defining design criteria for information presentation they proceed to infer generic decision-making and operational task requirements for co-operative team action. The derived criteria were used to compare optional team constellations.

In our first applications of the core-task analysis approach we focused the analysis of operators' performance with aim of improving learning and training. The core-task modeling has also been exploited in the context of validating control room designs. The group accomplished a comprehensive validation study in the nuclear power plant domain. In this study a new operator aid for the management of disturbances was validated in full-scope simulator experiments (Norros et al 1997, Norros & Nuutinen 1999a). A detailed description of its use as a validation tool is under preparation.

The (formaive) core-task modeling approach appears to provide one solution to the problem addressed by Miberg-Skjerve and Skraaning with regard to connecting human–machine system output criteria to the other validation criteria. The advantage of this modeling approach is that because the result-critical intrinsic demands of the domain, i.e. the core-task demands, constitute practice internal generic success criteria for work, they are meaningful features of work (Norros submitted 2003). The act of taking the core task demands into account in behaviour – which may be evidenced by the analysis of either actual operations or conceptions of work – demonstrates that a particular person, or organisation, values the significance of these features for achieving the intended outcomes of the human–technology system. Consequently, by utilising the core task concept, there is a possibility to establish the connections between system performance and efficiency of work performance in reference to particular situations, and with longer-term user acceptance, taken that acceptance may be interpreted to be linked to the overall significance and meaningfulness of work.

5 Quality of evaluation

As we are studying suitable methods for the HTI evaluation of complex cognitive systems we must also consider the criteria that makes the evaluation itself a valid procedure. According to O´Hara (1999) the validity of an evaluation method exists in four different forms: system representation, performance representation, test design, and statistical conclusion validity.

Typically the evaluation and validation of complex systems is performed in quasi-experimental settings. Quasi-experiments is the term used for research conducted, often infield settings, where performance comparisons are made between groups that are not formed by random generalisation. In this kind of test setting it cannot be directly assumed that the groups are representative of the population and that the observed effects can be unambiguously generalised. The reason why quasi-experiments are used is that they relate more to the practice and real usage of the system than pure laboratory experiments would. (Cook & Campbell 1979, O´Hara 1999)

In quasi-experiments the validity of the obtained results is ensured through four forms of experiment validity (O´Hara 1999):

1. *System representation validity* is logically supported such that it may be concluded that the integrated system is representative of the actual system in all aspects that are important to the performance

2. *Performance representation validity* is logically supported such that the measures of integrated-system performance and their associated criteria reflect good measurement practices and are concluded to be representative of important aspects of performance

3. *Test design validity* is logically supported such that a comprehensive testing programme was conducted by an independent, multidisciplinary team and there are no plausible biasing, confounding or masking effects to make the predictions of system performance ambiguous.

4. *Statistical conclusion validity* is logically supported such and based upon a convergence of the multiple measures such that it can be concluded that the performance of the actual system will be acceptable.

6 Concluding remarks

In this report, we have reviewed recent and also some earlier seminal literature concerning the evaluation of the user interfaces of complex industrial systems. In addition to scientific work we have reviewed the standards and guidelines that are relevant in the design of industrial human technology interaction. Based on this comprehensive study we want to express the following conclusions.

Within the evaluation and validation of complex industrial systems there are two objects of evaluation that need to be considered. They are: the design process in which the system is produced, and the system itself. The system (or part of it i.e. the control room) is, of course, the one that has the significance for usage, but in order to design an appropriate system the designers need to make design solutions that affect the issues of usage. These solutions are made in all the phases of the design process. The various kinds of design process models, covered by standards and guidelines, aim at defining what these solutions are, and in which phases of the design process they need to be made.

There are several methods that can be used in the evaluation of HTI solutions of complex industrial systems. In this literature review we have divided the methods into two categories. The categories are: inspections, and performance evaluations. Both of them can be used in evaluation of either the design process or its outcome. The categorisation is based on the fact that the appropriateness of a given system expresses itself either as compliance with relevant standards or the quality of performance. Compliance with the standards is evaluated with an inspection and the quality of use in a performance evaluation.

An inspection is a procedure in which the design solutions are compared to the existing standards and guidelines. If the solutions are in compliance with the relevant standards and guidelines the inspection is passed. The inspections should always be performed by multidisciplinary teams, which consist of experts in design, human factors, domain knowledge etc.

In performance evaluation the overall performance of the system is assessed. In performance evaluation it is also necessary to have a multidisciplinary team performing the evaluation. At the moment performance evaluation is something that is carried out towards the end of a given design process and the inspections are the ones that are performed during the design process. In our opinion also the performance evaluation should be started sooner in the design. A specific goal of our IDEC research project is to enable the evaluation of system performance as early as possible.

All of the above mentioned types of evaluation procedures need always to be based on relevant and appropriate evaluation criteria. Based on the reviewed literature we conclude that there seems to be a problem with defining suitable criteria. It is very difficult to say when the performance of the human–machine system is at an acceptable level. What makes performance acceptable anyhow? The existing standards do not help us in solving the question nor is there any scientific literature denoting what is acceptable. It is trivial to say what kind of performance is *not* acceptable, but it is another problem to define how close we can go to the absolutely not acceptable, and still remain acceptable.

Within the tradition of usability research several criteria exist for the appropriateness of the system being evaluated. We maintain that these criteria need to be integrated into the evaluation and validation tradition of HTI solutions of complex industrial systems. We also want to emphasise that the definition of appropriate and relevant acceptance criteria requires for multi-disciplinary work conducted jointly by the relevant stakeholders.

When defining the acceptance criteria for the outcome of a design process, e.g. a control room, we must also consider how the designers can affect the realisation of these criteria already in the design phase. These are the criteria for an appropriate design process. There are two dimensions to the appropriateness of the design process: temporal and scope. This is to say that first we need to consider that *when* in the design process the usage of the system is approached. Secondly, we need to consider, *which* design *solutions* actually have an *affect on the usage.* Based on the reviewed literature we maintain that even the design solutions concerning the process technology have an affect on the usage, thus these solutions should be made considering the appropriate use of the system.

One important criterion for the performance type evaluation is that the overall system interface needs to be such that it imposes good work practices on the users. For this purpose we need to create and define criteria for good work practice. We maintain that the criteria for good work practice is context related but we also believe that there are some generalisations that can be made about good practices. We also denote that the evaluation of work practices in complex industrial settings should not be based on the notion of human as the weakest link in the system. We want to emphasise the favourable actions of human operators, which are based on humans' ability to adapt and adjust to novel situations.

The validation method and criteria under construction in IDEC-project are based on one core idea. This idea is that the user, and the usage of the system being created need to be taken fundamentally account of in all the phases of the design process from the design of process technology to the design of user interfaces. We call this design approach an *activity oriented design*. The IDEC framework aims at creating methods and criteria to evaluate whether activity oriented design have been used. The first year of IDEC-project has mostly been concerned with the evaluation of design solutions. In the future we would like to expand our evaluation framework to cover also design-related issues.

Finally we want conclude our literature review in the following three points.

- *There is a fundamental need for deepening the relationship between usage and design. Usage must be the driver also in the design of process technology, plant's safety, and instrumentation and control as well as that of the design of man–machine interface.*
- *In order to enable such profound design we need to create design methods that drive the design activity into this direction.*
- *Criteria for evaluating what is good design are not available in present standards. Criteria for good tools and practices must consider human in a favourable role in the various tasks of operating an industrial plant as opposed to the view of human as a source of error. This is the way to create an optimal system, which takes advantage of the unique capabilities of both the human operators and the automation systems.*

References

Andresen G., Dröivoldsmo A. 2000. Human performane assessment: methods and measures Report HWR-353, OECD Halden Reactor Project, Halden.

Bainbridge L. 1987. Ironies of Automation. In: J. Rasmussen, K. Duncan, J. Leplat,(ed.). New Technology and Human Error. Chichester: John Wiley & Sons.

Bannon L. 2001. Toward a social and societal ergonomics: A perspective from computer-supported to cooperative work. In: M. McNeese, E. Salas, M.R. Endsley,(ed.). New trends in cooperative activities: Understanding system dynamics in complex environments. Santa Monica, CA: The Human Factors and Ergonomics Society. 9–21.

Bannon L. 2002. Taking "Human-Centred Computing" Seriously. In COCONET: Context-Aware Collaborative Environments for Next Generation Business Networks, ed. Helsinki.

Bannon L., Kuutti K. 2002. Shifting perspectives on organizational memory: From storage to active remembering. In: S. Little, P. Quintas, T. Ray,(ed.). Managing knowledge: An essential reader. London: Open University and Sage Publications.190–210.

Bye A. 2003. The OECD Halden Reactor Project programme 2003–2005. Presented at OECD NEA, SEGHOF meeting, Paris,

Cannon-Bowers J.A., Salas E., eds. 1998. making decisions under stress. Implications for individual and team training. Washington: American Psychological Association.

Carrol J.M. 1997. Human-computer interaction: psychology as a science of design. International Journal of Human-Computer Studies:46, 5105–22.

Carroll J.M., Kellogg W.A., Rosson M.B. 1991. The task-artifact cycle. In: J.M. Carrol,(ed.). Designing interaction: Psychology at the human–computer interface. Cambridge, England: Cambridge University Press.74–102.

CNSC. 2003. Human factors verification and validation plans Report G-278, Canadian Nuclear Safety Commission, Ottawa.

Cook T., Campbell D. 1979. Quasi-experimentation: design and analysis issues for field settings. Boston: Houghton Mifflin.

Corcoran W.R., Porter N.J., Church J.F., Cross M.T., Guinn W.M. 1981. The critical safety functions and plant operation. Nuclear Technology 55690–712.

Endsley M.R. 1995. Toward a theory of situation awareness in dynamic systems. Human Factors 3732–64.

Endsley M.R., Garland D.J., eds. 2000. Situation awareness. Analysis and Measurement. Mahwah, New Jersey: Lawrence Erlbaum.

Flach J., Hancock P., Caird J., Vicente K.J. 1995. Global perspectives on the ecology of human-machine systems. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gauthereau V. 2003. Work practice, safety and heedfulness. Studies of organisational reliability in hospitals and nuclear power plants Report 842, University of Linköping, Linköping

Hancock P., Chignell M. 1995. On human factors. In: J. Flach, P. Hancock, J. Caird, K.J. Vicente,(ed.). Global perspectives on the ecology of human–machine systems. Hillsdale, N.J.: Lawrence Erlbaum Associates.14–53.

Hollan J., Hutchins E., Kirsch D. 2000. Distributed cognition: Toward a new foundation for human-computer interaction research. AMC Transactions on Computer-Human Interaction 7:2, 174–96.

Hollnagel E. 1993. Human Reliability Analysis, Context and Control. London: Academic Press.

Hollnagel E. 2002. Understanding accidents – From root cause to performance variability. Presented at IEEE 7th Conference on Human Factors of and Power Plants, Scottsdale, Arizona,

Holmberg J., Hukki K., Norros L., Pulkkinen U., Pyy P. 1999. An integrated approach to human reliability analysis. Decision analytic dynamic reliability model. Reliability Engineering and System Safety 65239–50.

HRP. 2002. Proceedings. Presented at Enlarged Halden Programme Group Meeting, Storefjell.

Hukki K., Holmberg J.E. 2002. Enhancing risk-informed ways of management of nuclear power plant fire situations. In: R. Kyrki-Rajamäki, E. K. Puska,(ed.). FINNUS. The Finnish Research Programme on Nuclear Power Plant Safety 1999–2002. Final Report. Espoo: Technical Research Centre of Finland.261–7.

Hukki K., Norros L. 1998. Subject-centred and systemic conceptualisation as a tool of simulator training. Le Travail Humain 313–31.

Hutchins E. 1990. The technology of team navigation. In: J. galegher, R.E. Kraut, C. Egido,(ed.). Intellectual team work. Hillsdale, NJ.: Lawrence Erlbaum Associates, Publishers.191–220.

Hutchins E. 1995. Cognition in the Wild. Cambridge: MIT Press.

IEC-1771. 1995. Nuclear power plants – main control room – verification and validation of design. IEC.

IEC-1772. 1995. Nuclear power plants – main control room – Application of visual display units (VDU). IEC.

IEC-9064. 1989. Design for control rooms of nuclear power plants. IEC.

IEC-61839. 2000. Nuclear power plants – design of control rooms – functional analysis and assignment. IEC.

Ilyenkov E. 1977. Dialectical logic: essays on its history and theory. Moscow: Progress.

Ilyenkov E. 1984. Learning to think (in Finnish). Helsinki: Tutkijaliitto.

ISO-11064. 2000. Ergonomic design of control centres.

ISO-13407. 1999. Human centred design process for interactive systems. Geneve: ISO, 26.

ISO/IEC-15504-5. 1998. Software Process Assessment – Part 5. An assessment model and indicator guidance. ISO/IEC.

ISO-DIS-9241-11. 1998. Ergonomic requirements for office work with visual display terminals (VTDs) - Part 11. Guidance on usability.

Jokela T. 2001. Assessment of user-centred design processes as a basis for improvement action. An experimental study in industrial settings. Doctoral thesis. University of Oulu, Oulu.

Jokstad H., Fält C.O., Karlsson T., Kvalem J., Skraaning G.j. 2002. Halden Man-Machine Laboratory as of 2002. Presented at Halden Enlarged Programme Group Meeting, Storefjell,C3/4/1–4.

Kaarstad M., Ludvigsen J.T. 2002. Background study for further research in performance recovery. Presented at Enlarged Halden Programme Group Meeting, Storefjell,C2/5/1–16.

Keinonen T. 2000. One dimensional usability. In: T. Keinonen,(ed.). How to design usability? (in Finnish). Helsinki: Helsinki University of Art and Design.

Kirwan B. 2003. Design Process and Human-System Interfaces. Presented at International Summer School on Design and Evaluation of Human-System Interfaces, Halden , Norway, 1–26.

Kirwan B., Ainsworth. 1992. A guide to task analysis. In,(ed.). London: Taylor & Francis.

Klein G.A., Orasanu J., Calderwood R., Zsambok C.E., eds. 1993. Decision Making in Action: Models and Methods. Norwood, NJ: Ablex Publishing Corporation.

Klemola U.-M., Norros L. 1997. Analysis of the clinical behaviour of anaesthetists: recognition of uncertainty as basis for practice. Medical Education 31449–56.

Lee J., Moray N. 1994. Trust, self confidence, and operators' adaptation to automation. International Journal of Human–Computer Studies 40153. 84.

Leont'ev A.N. 1978. Activity, Consciousness, and Personality. Englewood Cliffs: Prentice Hall.

Leppänen A., Norros L. 2002. Teollisten prosessien inhimilliset käyttövarmuustekijät - katsaus alueen tutkimussuuntauksiin ja tutkimuksen kehitystarpeisiin. Työ ja Ihminen 16:2, 105–18.

Mallain P. 2003. Evaluation criteria for the successfulness of HTI-interaction. Grenoble: Smart Objects Conference,

McNeese M., Salas E., Endsley M.R., eds. 2001. New trends in cooperative activities: Understanding system dynamics in complex environments. Santa Monica, CA: The Human Factors and Ergonomics Society.

Miberg-Skjerve A.B., Skraaning G. 2003. A classification of validation criteria for new operational design concepts in nuclear process control. Presented at NEA/CSNI Workshop on Modifications at Nuclear Power Plants, Paris, France.

Naikar N., Pearce B., Drumm D., Sanderson P.M. 2002. Designing teams for first-of-a-kind, complex systems using the initial phases of cognitive work analysis: Case study. Human Factors 45:2, 2022–217.

Nicic A. 2003. Validation of the Human-Technology Interaction of NPP control rooms. The approach of the Canadian Nuclear Safety Commission. Leena Norros. Paris: OECD/NEA CNSI Workshop on Modifications at Nuclear Power Plants.

Nielsen J. 1993. Usability engineering. Boston: Academic Press.

Norman D.A. 1986. Cognitive Engineering. In: D.A. Norman, W. Draper,(ed.). User centered system design: New perspectives on human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum. 31–61.

Norros L. submitted 2003. Acting under Uncertainty. The Core-Task Analysis in Ecological Study of Work.

Norros L., Holmberg J., Hukki K., Nuutinen M. 1997. Slutrapport om validering av Säkerhets-ock ÖSI-panelen Report, VTT, Espoo.

Norros L., Kaasinen E., Plomp J., Rämä P. 2003. Human–Technology Interaction. Research and Design. VTT Roadmap Report 2220, Technical Research Centre of Finland, Espoo.

Norros L., Klemola U.-M. in press. Naturalistic analysis of anaesthetists' clinical practice. In: R. Lipshitz, B. Brehmer, H. Montgomery,(ed.). Methodologies of Naturalistic Decision Making (tentative title). Hillsdale, NJ.: Lawrence Erlbaum Associates.

Norros L., Nuutinen M. 1999a. Contextual analysis of the nuclear power plant operators' utilization of a disturbance management system. Presented at 7th European Conference on Cognitive Science Approaches to Process Control, Villeneuve d'Ascq, France, 87–92.

Norros L., Nuutinen M. 1999b. Development of an approach to analysis of Air Traffic Controllers' working practices. Presented at Human Error, Safety and System Development, Liege, Belgium.

Norros L., Nuutinen M. 2002. The Core-Task Concept as a Tool to Analyse Working Practices. In: N. Boreham, M. Fischer, R. Samurcay,(ed.). WHOLE.

Nuutinen M. 2003. Sukupolvenvaihdos turvallisuuskriittisen työn haasteena: ohjaajaharjoittelija oppipoikana vai tutkivana oppijana. Työ ja Ihminen 17:2, 173–89.

Nuutinen M., Norros L. 2001. Co-operation on bridge in piloting situations. Analysis of 13 accidents on Finnish Fairways. In: R. Onken,(ed.). The cognitive work process: Automation and interaction. Proceedings of the 8th Conference on Cognitive science approaches to process control. Munich: European Association of Cognitive Ergonomics. 3–14.

Oedewald P., Reiman T. in press. Core task modeling in cultural assessment – A case study in nuclear power plant maintenance. Cognition, Technology & Work.

O´Hara J. 2003. Overview of different types of control rooms and their human system interface solutions. Presented at International Summer School on Design and Evaluation of Human System Interfaces, Halden , Norway,3/1–28.

O´Hara J., Brown W.S., Lewis P.M., Persensky J. 2002a. Human-system interface design review guidelines Report NUREG-0700, U.S. Nuclear Regulatory Commission, Washington, DC

O´Hara J., Higgins J. 2002. Human factors engineering review of the Oskarshamn Unit 1 modernisation program Report, Statens Kärnkraftinspektion, Stockholm

O´Hara J., Higgins J., Persensky J., Lewis P., Bongarra J. 2002b. Human factors engineering program review model Report NUREG-0711, United States Regulatory Comission, Washington, DC

O´Hara J.M. 1999. A quasi-experimental model of complex human–machine system valdiation. Cognition, Technology & Work 1:1, 37–46.

Papin B. 2002. Integration of human factors requirements in the design of future plants. Presented at Enlarged Halden Programme Group Meeting, Storefjell,C3/1/–10.

Papin B., Quellien V., Paulus V. 2003. Design and first evaluation of a prototype of function-based NPP display dedicated to the shift supervisors. Presented at Paper presented at Workshop on "Innovative human-system interfaces and their evaluation, Halden, 1–7.

Peirce C.S., ed. 1998b. The Harvard Lectures on Pragmatism, Vols. 2. Bloomington and Indianapolis: Indiana University Press. 133–241 pp.

Pirus D. 2002. Computerized operation using formal plant functional breakdown. Presented at Enlarged Halden Programme Group Meeting, Storefjell,C3/8/1–8.

Pirus D. 2003a. Functional analysis – Operating functions Report, Electricité de France, Septen, France.

Pirus D. 2003b. Human–system interfaces. Types and principles. Presented at International summer school on Design and Evaluation of Human–System Interfaces, Halden,9/1–69.

Polanyi M. 1974. Personal knowledge. Chicago: The University of Chicago Press.

Rasmussen J. 1986. Information Processing and Human-Machine Interaction. Amsterdam: North-Holland.

Rasmussen J., Pejtersen A.M. 1995. Virtual ecology of work. In: J. Flach, P. Hancock, J. Caird, Vicente K.J.,(ed.). Global perspectives on the ecology of human–machine systems. New York: Wiley & Sons. 121–56.

Rasmussen J., Svedung I. 2000. Proactive risk management in a dynamic society. Karlstad: Swedish Rescue Services.

Reiman T., Norros L. 2002. Regulatory Culture: Balancing the Different Demands on the Regulatory Practice in Nuclear Industry. In: A.R. Hale, A. Hopkins, B Kirwan,(ed.). Changing Regulation – Controlling Hazards in Society. hhh: Elsevier.

Riihiaho S. 2000. Experiences with usability inspection methods. Licenciate thesis. Helsinki University of Technology, Espoo.

Rouse W.B. 1991. Design for Success- A human-centered approach to designing successful products and systems. US: John Wiley and Sons.

Saarni R., Andresen G., Nystad E. 2002a. Integrated task-oriented display system: First user tests Report HWR-701, OECD Halden Reacator Project, Halden.

Saarni R., Andresen G., Nystad G. 2002b. Integrated Tas-oriented Display System: First User Tests. Presented at Enlarged Halden Programme Group Meeting, Storefjell,C2/8/1–2.

Saarni R., Jokstad H., Meyer B.D. 2002c. Display and Alarm System for the FRESH PWR Simulator. Presented at Halden Enlarged Programme Group Meeting, Storefjell,C3/5/1–4.

Strand S. 2002. Trust and Automation: The influence of Automation malfunctions and System Feedback on Operator Trust. Presented at Enlarged Halden Programme Group Meeting, Storefjell,C2/13/1–4.

Suchman L.A. 1987. Plans and Situated Actions: The Problem of Human-Machine Communication. New York: cambridge University Press.

Watanabe S., Takahashi M., Kitamura M., Suzuki K. 2002. Characterisation and evaluation of anomaly event scenarios for training of nuclear power plant operations. Presented at Enlarged Halden Programme Group Meeting, Sotrefjell,C2//1–-9.

Vicente K.J. 1999. Cognitive Work Analysis. Toward a Safe, Productive, and Healthy Computer-Based Work. Mahwah, NJ: Lawrence Erlbaum Publishers.

Vicente K.J. 2002. Ecological Interface Design: Progress and Challenges. Human Factors 44:1, 62–78.

Wioland L., Amalberti R. 1996. When errors serve safety: towards a model of ecological safety. Presented at Cognitive Science Engineering in Process Control (CSEPC), Kyoto,194–1.

Woods D. 1988. Coping with Complexity:The psychology of Human Behaviour in Complex Systems. In: L.P. Goodstein, H.B. Andersen, S.E. Olsen,(ed.). Tasks, Errors, and Mental Models: A festschrift to celebrate the 60th birthday of professor Jens Rasmussen. London: Taylor & Francis. 128–48.

Woods D. 1995. Toward a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition. In: J. Flach, P. Hancock, J. Caird, K.J. Vicente,(ed.). Global perspectives on the ecology of human-machine systems. Hillsdale, NJ.: Lawrence Erlbaum. 157–88.

Woods D.D. 1993. Process-tracing methods for the study of cognition outside the experimental psychology laboratory. In: G.A. Klein, J. Orasanu, R. Calderwood, C.E. Zsambok,(ed.). Decision Making in Action: Models and Methods. Norwood, NJ: Ablex Publishing Corporation. 228–51.

Vygotsky L.S. 1978. Mind in Society. The Development of Higher Psychological Processes. Cambridge, Mass.: Harvard University Press.

Yamaguchi Y., Tanabe F. 2002. Creation and Evaluation of an Ecological Interface System for Operation of Nuclear Reactor System. Presented at Enlarged Halden Programme Group Meeting, Storefjell,C3/10/1–6.

Zambok C. 1997. Naturalistic decision making: Where are we now? In: C. Zambok, G.A. Klein,(ed.). Naturalistic Decision Making. Hillsdale, NJ: Lawrence Erlbaum.

Zuboff S. 1988. In the Age of the Smart Machcine. New York: Basic Books.

Östlund A.M. 2002. Granskning av O1MOD Kontrollrummvalidering Report 8.15-001094, Statens Kärnkraftinspektion, Stockholm

Östlund A.M. 2003. Säkerheten i fokus i det nya kontrollrummet. In Nucleus, ed. 26–9.