THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Measures and method characteristics for early evaluation of safe operation in nuclear power plant control room systems

Eva Simonsen

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Product and Production Development
Division Design & Human Factors
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2016

Measures and method characteristics for early evaluation of safe operation in nuclear power plant control room systems
Eva Simonsen

Report no 99
ISSN 1652-9243

Published and distributed by
Department of Product and Production Development
Division Design & Human Factors
Chalmers University of Technology
SE-412 96, Gothenburg, Sweden
Telephone +46(0)31-772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2016

ABSTRACT

Safe operation is a central objective for high-risk industries such as nuclear power plants. Operation of the plant is managed from a central control room, which is a complex socio-technical system of physical and organisational structures such as operators, procedures, routines, and operator interfaces. When control room systems are built or modified it is of great importance that the new design supports safe operation, something that must be evaluated during the development process. Summative evaluations at the end of the development process are common in the nuclear power domain, whereas formative evaluations early in the process are not as customary. The purpose of this licentiate thesis was to identify demands on evaluation methods for them to be suitable for early assessment of the control room system's ability to support safe operation. The research consisted of two parts: to explore evaluation measures relevant for nuclear power plant control room systems, and to identify requirements on evaluation methods for them to be useful in early stages of the development process.

To explore the issue of evaluation measures two interview studies were performed with various professionals within the nuclear power domain. The purpose of the first study was to investigate aspects contributing to safe operation, while the second study sought to identify design trends in future control room systems and their potential usability problems. To complement these empirical studies, other researchers' choices of measures for control room system evaluations were analysed. The results showed that a combination of measures from six categories is necessary to fully access the control room system: system performance, task performance, teamwork, use of resources, user experience, and identification of design discrepancies. In addition, the resilience engineering perspective should be considered in control room system evaluations in order to assess the ability to handle unanticipated events.

Requirements on evaluation methods were investigated through analysis of characteristics of early product development phases. The result was that system representations in these phases are more conceptual, and that using these representations to perform tasks differs in some aspects from use of the final system. Empirical methods that directly study user interaction with the control room system are therefore less suitable for early evaluations. Analytical methods that study use indirectly are a better choice. An additional identified requirement is that if methods are to be utilised in industry, practitioners must find them useful in practice.

To conclude, further work is needed to identify useful analytical evaluation methods that can assess measures from the six categories. Suitable methods for early assessment of the capacity for resilient behaviour is another topic that needs further exploration.

Keywords: Control room, nuclear power, evaluation methods, human factors engineering, safe operation, early development, resilience engineering

ACKNOWLEDGEMENTS

This research project is funded by the Swedish Radiation Safety Authority.

First, I would like to thank my main supervisor and examiner, Professor Anna-Lisa Osvalder, for her guidance, patience and valuable input. I am also grateful for the help from my former co-supervisor, Professor Håkan Alm (Luleå University of Technology), and my current co-supervisor, Dr Andreas Kjellin (Swedish Radiation Safety Authority). You have helped me view my work from different perspectives and your comments have improved its quality.

Dr Yvonne Johansson at the Swedish Radiation Safety Authority saw the need for investigating evaluation methods within the nuclear power domain and has been my contact person at the Authority. I extend my gratitude to her for encouraging me and trusting in my ideas.

This research project has a reference group that was invited to comment on the work. The members of the group are: Dr Jonas Andersson (Viktoria Swedish ICT), Agneta Bengtsson (Oskarshamns Kraftgrupp), Per Øivind Braarud (Institute for Energy Technology), Johan Holgersson (Ringhals), Jari Laarni (VTT Technical Research Centre of Finland), Professor Lena Mårtensson (Royal Institute of Technology), Maren Rø Eitrheim (Institute for Energy Technology), Stefan Sördal (Swedish Radiation Safety Authority), and Associate Professor Clemens Weikert (Lund University). It has been a privilege to have access to your knowledge and experience.

Without the participants in my studies, there would be no thesis. I thank you for taking the time to share your expertise with me.

The importance of context is emphasised within the field of human factors. My colleagues, both at the division of Design and Human Factors at Chalmers as well as at the Human Factors division at Vattenfall, have provided a fun, inspiring, and supporting context in which to pursue research. I thank you for the positive trend in my learning curve over the years. I am especially grateful to my colleague and friend, Dr Lars-Ola Bligård, who unselfishly provided his support whenever I needed it.

To maintain work-life, there must be a life outside of work. I am truly grateful to my friends, Cecilia in particular, who patiently listened and made fun things happen. The love and support of my family, Barbro, Lars-Erik, and Gustav, has always provided the foundation on which I stand. I thank you all for being there for me.

Finally I would like to thank my husband, Lars, for keeping me calm, well-fed, and laughing. Your support means everything.

APPENDED PUBLICATIONS

PAPER I

Simonsen, E., Osvalder, A. L. (2015) Aspects of the Nuclear Power Plant Control Room System Contributing to Safe Operation. In *6th International Conference on Applied Human Factors and Ergonomics*; 26-30 July 2015, Las Vegas, NV.

PAPER II

Salomonsson, E., Osvalder, A.-L., Alm, H. (2013) Requirements on Human Factors Methods for Evaluating Safety in Future Nuclear Power Plant Control Rooms. In *Enlarged Halden Programme Group Meeting: Man-Technology-Organisation (MTO) Sessions*; 10-15 March 2013, Storefjell.

PAPER III

Simonsen, E., Osvalder, A.-L. (2015) Exploring an approach to connect the design of nuclear power plant control room systems to the cornerstones of resilience. In *6th REA Symposium*; 22-25 June 2015, Lisbon.

TABLE OF CONTENTS

INTRODUCTION

This chapter describes the background of the doctoral project, its purpose, aim, and research questions. It also details the aim and research questions for the present work. The chapter ends with reading instructions for the licentiate thesis.

1.1 BACKGROUND

Swedish nuclear power plants were built in a period from the mid-seventies to the mid-eighties of the 20th century. Maintenance and modernisation demands have led to the initiation of a number of plant development projects. Either directly or indirectly, this led to changes in the plants' control rooms as well. The modification of control rooms creates a need to evaluate whether the changed design continues to support safety, productivity and the working environment. The same applies to newly built nuclear power plants too.

Against this background, the Swedish Radiation Safety Authority initiated a study (Osvalder and Alm, 2012). The aim was to study and critically review methods and procedures used today to evaluate changes in control rooms and their possible impact on safety, productivity and the working environment, and also to discuss the need for modified or new methods.

The study by Osvalder and Alm (2012) showed that Swedish nuclear power plants do not have a common view about, or established methods for, how control rooms should be evaluated with regard to safety. Other problems noted were the lack of baseline measurements, limited use of usability testing, and that methods for risk assessment were used in a simplified manner or not at all.

The report also pointed out that existing risk analysis methods are component-based and only study the interaction between an operator and single components. The need for a more systemic approach to analysing control rooms was emphasised. It was also questioned whether the methods used today are generally adapted to the technology found in older control rooms, or if methods are able to analyse more modern control room designs as well. Osvalder and Alm (2012) stated that practitioners only use a few of the methods available, and that they need methods that are flexible and simple to use.

The report became the foundation for a research project of which this licentiate thesis is a part. The purpose of the project is to improve and further develop knowledge of methods for evaluation of modified and newly designed control rooms for process control, with a focus on safe operation.

Within the scope of this purpose, the main goals of the research project are to:

1. Provide knowledge, methods and guidelines to support the Swedish Radiation Safety Authority in its role as a supervisory and licencing authority.
2. Modify existing and develop new methods, guidelines, and principles.
3. Support and improve national competence in the domain (for example owners, consultants and manufacturers) as well as academia.

The present work concerns the human factors contribution to nuclear safety and safe operation. Many definitions of the term 'human factors' exist, but the definition utilised in this licentiate thesis is the one from the International Ergonomics Association (2016): "*the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance*".

In their report, Osvalder and Alm (2012) referred to a report from a workshop held by a Nuclear Energy Agency committee regarding modifications of nuclear power plants (OECD/NEA Committee on Safety of Nuclear Installations, 2005). This report states that human factors efforts must start early in the development project in order for them to be effective. Introducing them later usually increases costs and limits the opportunities for improvements to the system. Human factors efforts include evaluations, so human factors evaluation methods must be suited for use in early stages. In addition to their interview study, Osvalder and Alm (2012) reviewed the procedures for human factors work within plant modifications from all Swedish nuclear power plants. Reading this review, the support for and emphasis on human factors verification and validation is evident, whereas earlier evaluations are not as clearly stipulated. This, in addition to the author's own experience of working within the Swedish nuclear power domain, points to a need for research into methods for early evaluation.

Evaluation cannot be undertaken without knowing what to evaluate. The control room system's ability to support safe operation is a phenomenon that must be operationalised to make it possible to evaluate. If the control room is to be able to conduct safe operation, the road to assessing this goes through identifiyng the aspects that contribute to safe operation.

Given the above preconditions the following research questions were formulated for the research project:

RQ1: Which aspects must be evaluated to assess the control room system's ability to support safe operation of the plant?

RQ2: When evaluating the control room system's ability to support safe operation early in the development process:

a. What characteristics must evaluation methods have?

b. Are there suitable evaluation methods?

c. If there are no suitable methods, how must existing evaluation methods be modified in order to be suitable?

1.2 PURPOSE AND RESEARCH QUESTIONS

The purpose of this licentiate thesis was to identify demands on evaluation methods for them to be suitable for early assessment of the control room system's ability to support safe operation from a human factors perspective. The work has been focused on the nuclear power domain.

To fulfil this purpose, the present work answers research question 1 and 2a of the overall research project, namely:

RQ1: Which aspects must be evaluated to assess the control room system's ability to support safe operation of the plant?

RQ2: When evaluating the control room system's ability to support safe operation early in the development process:

a. What characteristics must evaluation methods have?

1.3 PERSONAL CONTEXT

My educational background is industrial design engineering. Studying engineering made problem-solving and the creation of artefacts central to me. The industrial design focus in my education highlighted the need to base this problem-solving and creation on a thorough analysis of use and user needs. I have worked as a practitioner within the field of human factors engineering for ten years, with the focus on control room system development. My work was mainly carried out within the Swedish nuclear power domain, but it also included work with control room systems in other domains, such as train dispatch and combined heat and power plants. My practical experience made me aware of the importance of methods that are usable in practice. Methods are tools, and like all tools they bring value only when they are used. Evaluations are a vital part of development work, and integrated systems validations in control room simulators are common practice in the Swedish nuclear power domain. I find, however, that the use of formative evaluations can be improved further, and that methods for early evaluation are needed. Another insight gained during my work-life experience was the importance of addressing the control room system. Operator performance relies on more than the design of operator interfaces; other parts of the system, such as procedures, training, and routines, affect human behaviour as well. These experiences have to a large degree shaped the scope and focus of this licentiate thesis.

1.4 READING INSTRUCTIONS

This chapter provides the background, aim, and research questions that form the foundation and direction of the work presented in this thesis. In addition, it gives a description of the author's educational and professional background and their implications for the present work.

Chapter two presents the research approach and the methodology of the two interview studies on which this thesis is based.

The third chapter explains concepts and terms important for understanding the results. To some extent, this applies to chapter four as well, but in addition this chapter explores existing evaluation methods and characteristics of early evaluation to answer research question 2a. Chapter 5 combines input from the interview studies and literature to seek the answer to research question 1.

Chapter 6 discusses the methods, results and implications of the present work. Chapter 7 presents the conclusions of this licentiate thesis.

RESEARCH APPROACH

The studied object, the nuclear power plant control room system, is a complex socio-technical system. Identification of the aspects that contribute to the control room system's ability to support safe operation (research question 1) solely using an empirical approach was possible in theory, but not in practice. In theory variables could be changed and the corresponding effect on safe operation could be monitored, but the complexity of the system and its environment made this approach impossible. Therefore, an empirically based rationalist approach to knowledge acquisition was chosen to answer research question 1. Two interview studies were performed to utilise the knowledge of professionals within the nuclear power domain. These provided qualitative empirical data which, combined with qualitative data from literature, constituted the base for rationalistic reasoning regarding the aspects that contribute to safe operation. The answer to research question 2a (required method characteristics) was explored solely through rationalistic reasoning based on qualitative data from literature.

2.1 STUDY I

The aim of Study I was to identify a foundation for evaluation measures by finding aspects of the control room system that contribute to safe operation from a human factors perspective.

The design of the control room system and the way it is operated will largely affect its performance, which makes personnel responsible for design and operation a valuable source of information. Thus Study I was an interview study to utilise the experience of professionals within the Swedish nuclear power domain. The professional roles chosen were those influencing human factors-related aspects rather than technical aspects. In total fourteen persons in seven roles were interviewed (two representatives of each role). Table 1 shows the characteristics of the interviewees.

The semi-structured interviews took about 1-1.5 hours each and were held at the interviewees' workplace. Documentation was done with audio recordings and written notes and the interviews were held by the same interviewer. The interviews were divided into four parts, an introduction, a second section containing broader questions, followed by a section with more detailed questions, and a conclusive end section. The third part of the interview used different angles of the overall investigated issue to trigger the interviewees' thoughts in order to obtain more extensive answers. The contents of the various sections are described in Table 2.

Table 1: Characteristics of the various groups of interviewees in Study I.

Role	**Description**	**Background**	**Reactor type experience**
Reactor operator	Responsible for operation of safety-related systems	Operation	Half the group had experience of boiling water reactors, the other half had experience of pressurised water reactors
Shift supervisor	Operatively responsible for all work in the control room	Operation	See above
Instructor	Responsible for implementing training of operators	Operation	See above
Human factors specialist working for the plant owner (licensee)	Responsible for human factors issues in plant modification projects	Behavioural science and/or engineering	N.a.
Human reliability analysis specialists	Performing human reliability analyses as part of the probabilistic safety analyses	Behavioural science and/or engineering	N.a.
Human factors specialist, Swedish Radiation Safety Authority	Responsible for reviewing the fulfilment of safety requirements	Behavioural science and/or engineering	N.a.
Inspector, Swedish Radiation Safety Authority	Responsible for reviewing the fulfilment of safety requirements	Operation	Boiling water reactors, one also had limited experience of pressurised water reactors

Table 2: Contents of the various sections of the interviews in Study I.

Introduction
• Explaining the purpose of the study • Explaining to the interviewees that their view of what contributes to safe operation was sought, not 'the right answer' • Explaining that the interviewees should not only consider the physical control room design when answering the questions, but also include items such as procedures and personnel • Questions regarding the interviewees' role, the duration of their experience in that role, and their previous experiences regarding employment and education
Broader questions
• The interviewees' view of the meaning of the term 'safe operation' and the definition of the term for the study in question • What contributes to safe operation • What contributes most to safe operation • What in the control room system needs to be evaluated to assess if it supports safe operation • To interviewees with an operational background: recollection of a real-life or simulator-set event with negative/potentially negative consequences, and aspects that saved the situation or mitigated the effects
Detailed questions
• What must be possible to perform in the control room system. The approach used: look at the human-machine system from a task point of view, consider the control room system as a performer of tasks (something the system does). • What sub-functions must exist in the control room system. The approach used: look at the system from a functional point of view, consider the control room system as a compilation of abilities (something the system has the capacity to do). • What parts the control room system should consist of. The approach used: look at the system from a structural point of view, consider the control room system as a collection of physical or social parts that realise the system. • What characteristics the control room system should have. The approach used: discuss the control room system in terms of characteristics of the physical or social parts; the necessary properties of the structural elements.
Conclusion
• Summary of the interviewee's answers (to give an opportunity to correct misunderstandings) • What contributes most and least to safe operation, since the interview might have led the interviewee to think more specifically about this issue.

The qualitative material from the interviews was analysed using thematic analysis, a primarily descriptive approach to defining broad categories (themes) that describes significant features of data (Howitt, 2013). The thematic analysis procedure consists of six steps: data familiarisation, initial coding generation, search for themes based on initial coding, review of the themes, theme definition and labelling, and report writing. Going through these steps should be an iterative process and not a linear one (Howitt, 2013).

The interview data was transcribed (in full). Initial codes were generated by marking statements regarding aspects of the control room system that contribute to safe operation and summarising their content into one or a few words. The initial codes were searched for patterns that indicated themes and sub-themes. The angles utilised in the third part of the interview were used to structure the initial coding, but were modified to better fit the data.

The empirical data from Study I was used in an additional analysis to explore how aspects of the nuclear power plant control room system can be connected to the four basic abilities of resilient performance (respond, monitor, anticipate, and learn; these are further described in section 3.1). The result of this analysis is presented in Paper III. The perspective used in Study I – what contributes to safe operation, not what threatens it – is in line with the focus of investigating not only the things that go wrong, but also the things that go right argued in resilience engineering (Hollnagel, 2013). Thus the interview data from Study I was deemed relevant for this second analysis. Statements concerning aspects deemed to affect any of the four cornerstones of resilience were marked. Each of the four groups of statements was then reviewed again and the themes presented in Paper I (described in section 5.3) were used to connect concrete aspects of the control room system design to the four basic resilient abilities. The four basic resilient abilities are functions, and these are in turn made possible by underlying sub-functions. The focus of Paper III was the design of the control room system, so tasks were expressed as the functions requiring the performance of these tasks, and the structural elements and their associated characteristics needed to perform these tasks. Situations, as they were defined in Paper I, concern the system's resilient behaviour as a whole, and were not connected to specific cornerstones.

The results of Study I are presented in sections 5.3 and 5.5.

2.2 STUDY II

One path towards increasing the control room system's ability to support safe operation is to identify usability problems in the control room system design so they can be rectified. The availability of new technologies brings changes in nuclear power control room system design, which may affect the usability problems that occur. The evaluation methods used will determine the type of measures that can be implemented, meaning that finding different types of usability problems will require the use of different evaluation methods. It is therefore interesting to investigate the types of usability problems found in present and future nuclear power control room systems to be able to identify suitable measures and methods.

The aim of Study II was to suggest requirements that the human factors evaluation methods must fulfil to be useful. The requirements were to be based on possible usability problems that required attention in the design of future Swedish nuclear power control room systems.

Design trends in future Swedish nuclear power control room systems were investigated through six semi-structured phone interviews with seven professionals (one of the interviews was a group interview with two persons). The interviewees were in a position of responsibility for human factors issues in the control rooms of their respective production units. They therefore had knowledge of forthcoming control room alterations, as well as insights regarding the development of their units' control rooms in the more distant future. The interviews covered all ten reactors in Sweden. The interviews were all conducted by the same person and took about one hour each. The interviews were documented in handwritten notes.

The questions concerned the control room system changes planned for each unit, the reasons for making the changes, when the changes were planned to be implemented as well as what changes the interviewee believed would be made in the more distant future (i.e. changes the interviewee viewed as probable but not yet decided by the plant owners).

In addition to investigating modifications in today's nuclear power plants, Study II sought information on control room systems in new plant designs. For this reason an additional interview was conducted. This interview was held with a person who had knowledge about the control room designs for two generation III+ reactors. Current reactors in Sweden today are generation II, and the term generation III+ is used to denote a category of more modern reactors. Generation III+ reactors are the most modern reactors being built today. The control rooms of the new generation III+ reactor types are in theory standardised, but may be changed according to the requirements and needs of the customer in each individual implementation. Only the standardised design of the control rooms was investigated in Study II. This interview took about two hours and was carried out face to face by the same person who undertook the other six interviews. The interviewee was asked to describe forthcoming design trends in the control rooms of two specific generation III+ reactor types. This interview was also documented in handwritten notes.

The identified control room system design trends were analysed in terms of usability problems that could potentially arise. The resulting usability problems are not to be regarded as a comprehensive list of usability problems that may occur in future control rooms, but they do indicate requirements concerning human factors methods for evaluating safe operation in control room systems.

The results of Study II are presented in section 5.4.

NUCLEAR POWER PLANT CONTROL ROOM SYSTEMS

This chapter describes the application area of this licentiate thesis, nuclear power plant control room systems. Safety and safe operation are two central concepts, as is the activity of control room system modernisations.

3.1 SAFETY-I AND SAFETY-II

A traditional definition of safety is that it is freedom from unacceptable risk. A consequence of this view is that the focus is on what goes wrong, and the road to safety goes through looking for failures, trying to find their causes, and trying to eliminate causes and/or improving barriers (Hollnagel, 2013). However, socio-technical systems such as nuclear power plants are complex because the interactions between elements of the system are complex. Complex interactions bring about unfamiliar or unexpected sequences of events, sequences that are either not visible or not immediately comprehensible (Perrow, 1999). Trying to remove the possibility for all of these unexpected and unwanted outcomes in complex systems is extremely difficult (or even impossible). A complementary view of safety addresses this problem by defining safety as the ability to succeed under varying conditions, so that the number of intended and acceptable outcomes is as high as possible (Hollnagel, 2013). The traditional view of safety has been dubbed Safety-I and the complementary Safety-II. The intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that the system can sustain required operations under both expected and unexpected conditions, is called resilience (Hollnagel, 2011b). This definition emphases that a system should not only strive to avoid failures, but to adapt its functioning to handle all conditions. Resilience engineering is the field that has developed theories, methods, and tools to deliberately manage this adaptive ability of organisations in order to make them function effectively and safely (Nemeth and Herrera, 2015). Resilience engineering argues that the focus should be on increasing the number of things that go right, which as a natural consequence will decrease the number of things that go wrong.

3.2 SAFE OPERATION IN NUCLEAR POWER

Nuclear safety is defined by the International Atomic Energy Agency as "*The achievement of proper operating conditions, prevention of accidents or mitigation of accident consequences, resulting in protection of workers, the public and the environment* *from undue radiation hazards*" (International Atomic Energy Agency, 2007). A presentation of underlying objectives and principles of nuclear safety is given by the International Nuclear Safety Advisory Group (1999). The framework provided there contains three overriding safety objectives, six fundamental safety principles, and nine technical principles, which provide a general framework for a number of specific safety principles (Figure 1). The latter are grouped mainly after the main stage in a nuclear power plant's lifetime where they are applicable. The first overriding safety objective is a general nuclear safety objective: "*To protect individuals, society and the environment by establishing and maintaining in nuclear power plants an effective defence against radiological hazard*" (International Nuclear Safety Advisory Group, 1999, p.8). The other two overriding safety objectives are overlapping objectives regarding radiation protection and technical safety (such as pointing to the use of reliable components). Together, the three objectives ensure completeness.

The concept of defence in depth is a fundamental principle in nuclear safety. International Nuclear Safety Advisory Group (1999) describes the concept as implementing several levels of protection, including successive barriers preventing the release of radioactive materials to the environment, to compensate for potential human and mechanical failures. This includes all safety activities, whether organisational, behavioural or equipment related. There are five levels of defence in depth, and they range from preventing abnormal operation and failures to mitigating radiological consequences of significant releases of radioactive materials. The strategy is to first prevent accidents and if this fails, limit the potential consequences of accidents and prevent them from evolving into more serious conditions.

Because of the potentially harmful consequences nuclear power plants are regulated at the governmental level to protect people and the environment from the undesirable effects of radiation. Human factors is stated as a general technical principle in the framework by the International Nuclear Safety Advisory Group (1999). This principle proclaims that the possibility of human error should be handled by facilitating correct decisions by operators and inhibiting incorrect ones, as well as by providing means for detecting and correcting or compensating for errors. In Swedish nuclear power plants human factors issues are regulated by the Swedish Radiation Safety Authority. Chapter 3 section 3 of the regulatory code SSMFS 2008:1 stipulate that "*the design shall be adapted to the personnel's ability to, in a safe manner, monitor and manage the facility and the abnormal operation and accident conditions which can occur*". More detailed regulations for control room design and emergency control posts are given in another regulatory code, SSMFS 2008:17.

Objectives	General nuclear safety objective			Radiation protection objective			Technical safety objective	
Fundamental safety management principles	Safety objective			Responsibility of operating organisation			Regulatory control and verification	
Fundamental defence in depth principles	Defence in depth			Accident prevention			Accident mitigation	
General technical principles	1) Proven engineering practices	2) Quality assurance 3) Self-assessment 4) Peer review	5) Human factors	6) Safety assessment and verification	7) Radiation protection	8) Operating experience and safety research	9) Operational excellence	
Specific principles	Siting	Design	Manufacturing and construction	Commissioning	Operation	Accident management	Decommissioning	Emergency prepared-ness

Figure 1: Safety objectives and principles for nuclear power plants. First row: overriding safety objectives. Second and third rows: fundamental safety principles. Fourth row: technical principles. Fifth row: specific safety principles. Adapted from International Nuclear Safety Advisory Group (1999).

Nuclear power plants must not only be safe, they must be safe while producing electricity. In the long run safety and production are prerequisites for each other. Combining the demand to produce electricity with the demand to uphold nuclear safety concludes that a nuclear power plant must produce electricity without exposing workers, the public or the environment to radiation hazards. This is a definition of the term safe operation from a Safety-I perspective. A definition of safe operation from a Safety-II perspective would be that the nuclear power plant must produce electricity and operate the process within permitted operational limits during all conditions. In Sweden, clearly defined operational limits and conditions are stipulated by the Swedish Radiation Safety Authority in chapter 5 section 1 of the regulatory code SSMFS 2008:1. These should, together with procedures, provide personnel with the guidance they need to be able to conduct operations in accordance with what the plant is designed to handle, as stated in the plant's safety analysis report.

3.3 THE CONTROL ROOM SYSTEM

A control room is a functional entity responsible for the operational control of something, for example a nuclear power plant or train dispatch. The control room, including its associated physical structure, is where the operators carry out centralised control, monitoring and administrative responsibilities (International Standard Organisation, 2000).

The nuclear power plant control room is a place where human operators exercise control over a process. Tschirner (2015) propose four prerequisites that must be fulfilled for a human operator to achieve efficient control over a process. First, the operator needs a clear goal, such as a state to reach or a condition within which a system must be maintained. Second, the operator needs a model of the process, the system, and the environment to be able to assess the current state of the system and predict future ones. Third, the operator must be able to observe the current state of the system, environment, and process. Fourth, the operator must be able to control the process.

The nuclear power plant operators' work in normal operation is typically calm and can be carried out according to predefined routines. Routines typically exist for undesired events as well, but situations where the operator has to handle an unfamiliar situation without the support of routines will also occur.

A nuclear power plant control room is operated by a team of operators, who work in shifts to allow continuous operation. Responsibilities are divided among the operators, creating different roles. In Swedish nuclear power plants these are typically shift supervisor, reactor operator, turbine operator, and field operators. An assistant reactor operator or an electrical operator is also included in the shift team, depending on the reactor type.

The physical structure of the nuclear power plant control room includes operator interfaces, which can be screen-based or analogue. The operator interfaces may be installed so they can be operated while sitting or standing, and viewed from nearby or from further away. In addition to the equipment needed to control the plant directly, more indirectly contributing parts such as a meeting area and office for the shift supervisor are often included in the control room as well.

Procedures are often used to guide operations in the control room, especially within the nuclear power domain. Traditionally they are presented on paper, but in recent years computer-based procedures have been developed as well. Procedures play a very important part in the operation of nuclear power plants and, as stated in chapter 5 section 1 of SSMFS 2008:1, are required by the Swedish regulator to provide personnel with the guidance they need.

In this licentiate thesis, the focus will be on the control room system, a socio-technical system including humans, technology, and organisational elements. This focus was chosen to emphasise that the operator interfaces and other parts of the physical structure are not enough to achieve proper control. Other components such as the operators' competence, procedures, roles in the shift team, and work routines are also vital for the function of the control room system. In the present work, a control room system is defined as a socio-technical system consisting of humans, technology, and organisational elements that exercise centralised control and monitoring over a process, as well as administrative responsibilities.

3.4 CONTROL ROOM SYSTEM MODERNISATION

A control room system rarely stays unchanged from its initial construction to its final decommissioning. Control room system modernisations can be initiated for many reasons, and examples for nuclear power can be found in a report from a Nuclear Energy Agency committee (OECD/NEA Committee on Safety of Nuclear Installations, 2005). Reasons stated in the report include, but are not restricted to, rectification of plant deficiencies, improvements in plant performance, adaptation to new regulatory requirements, and the utilisation of new technologies. Depending on the reason for change and the budget available, control room system modernisations can differ considerably in scope. They may range from the changing of a set point or the substitution of a component to a total upgrade of the entire control room system (OECD/NEA Committee on Safety of Nuclear Installations, 2005).

If the operation controlled by the control room system is safety-critical, changes to the control room system will have potential safety consequences, due to the control room system's operational significance (Norros and Nuutinen, 2005). For nuclear power, this view is shared in the report by the OECD/NEA Committee on

Safety of Nuclear Installations (2005), where it is emphasised that modifications have the potential to introduce challenges to safety if they are not carried out with the necessary caution and prudence.

The Swedish Radiation Safety Authority's regulatory code SSMFS 2008:1 chapter 3 section 3 stipulates that the design of the nuclear power plant must "*be adapted to the personnel's ability to, in a safe manner, monitor and manage the facility and the abnormal operation and accident conditions which can occur*". The general advice in the regulatory code SSMFS 2008:17 section 18 suggests that examples of methodology for the evaluation of control room modifications are to be found in documentation from the United Stated Nuclear Regulatory Commission (U.S. NRC), NUREG-0711. This document provides staff at the U.S. NRC with a review methodology that addresses the scope of a human factors engineering review of plant modifications or newly built plants (United States Nuclear Regulatory Commission, 2012). In Sweden, and in other countries, NUREG-7011 is used as guidance for deciding which human factors activities should be included in plant development projects, and how these should be performed.

DEVELOPMENT AND EVALUATION OF CONTROL ROOM SYSTEMS

This chapter describes the process for development of control room systems and the role evaluation plays in this process. It also presents an overview of evaluation measures and methods, as well as what is required of methods for evaluation in the early phases of the development process.

4.1 THE DEVELOPMENT PROCESS

Man-made things do not appear out of nowhere, and organisations designing and developing things normally follow some sort of process to do so. Ulrich and Eppinger (2003, p. 14) use the term 'product development process', which they describe as "*the sequence of steps or activities which an enterprise employs to conceive, design, and commercialize a product*". Other benefits of using a well-defined development process stated by Ulrich and Eppinger (2003) are that it supports quality assurance, coordination, planning, management, and improvement.

There are numerous suggestions for how development processes are and should be structured. They differ, among other things, in how much of the product life cycle they cover and what is included in each phase. Some end after the design is finished and others include production. One common theme however, is the gradual increase in detailing of the solution. A phase establishing a more overall design solution normally precedes a phase where a more detailed design is developed.

The planning and design process suggested by Pahl and Beitz (1996) includes four phases. The first phase, planning and clarifying the task, has the purpose of collecting information about the requirements that have to be fulfilled by the product, as well as existing constraints. The second, the conceptual design phase, determines the principal solution, and is followed by the embodiment design phase where the construction structure (overall layout) is determined. The arrangements, forms, dimensions, and surface properties of all individual parts are then decided on in the detail design phase.

The beginning of a process suggested by Ulrich and Eppinger (2003) is similar to the process proposed by Pahl and Beitz (1996), up to and including detail design, but it also includes the testing and refinement phases, and production ramp-up at the end. Testing and refinement are where preproduction versions of the product are constructed and evaluated to finalise the design. The purpose of production ramp-up is to train the work force and to work out remaining problems in the production process.

A plant, such as a nuclear power plant, is normally not viewed as a product. That does not mean that developing and modifying them does not need a structured process. The OECD/NEA Committee on Safety of Nuclear Installations (2005) stated in a report that a systematic approach to plant modifications is necessary to reduce the risk posed by modifications. They suggest that an established and documented modification process ensures consistency, repeatability, and traceability. Hale et al. (2007), in a special issue of Safety Science on safety in design, summarise six main phases in typical design processes for complex technical systems involving major accident hazards: business development; feasibility study; conceptual design; basic design; detailed design; and fabrication, installation, commissioning and start-up. The main difference between this process and the ones proposed by Pahl and Beitz (1996) and Ulrich and Eppinger (2003) is the final phase. This is a natural consequence of the fact that many complex technical systems, such as process plants or offshore platforms, are uniquely built and installed, not mass-produced. The three processes described above are summarised in Table 3.

Table 3: Overall correspondence between phases in different development processes.

Pahl and Beitz (1996)	**Ulrich and Eppinger (2003)**	**Hale et al. (2007)**
Planning and clarifying the task	Planning	Business development
		Feasibility study
Conceptual design	Conceptual design	Conceptual design
Embodiment design	System-level-design	Basic design
Detail design	Detail design	Detailed design
	Testing and refinement	
	Production ramp-up	
		Fabrication Installation Start-up Commissioning

Another issue differentiating the development process of complex technical systems from that of other products is the development of procedures and training. The operation of complex technical systems is often very dependent on both procedures and training of personnel. While not unimportant for other products, training and procedures are seldom a requirement for use. The same is true for training, and if it is a requirement for use it is often not the responsibility of the company developing the product. This emphasis on procedures and training for complex technical systems is evident in the process indicated by NUREG-0711, the nuclear power review methodology guide presented in section 3.4 (United States Nuclear Regulatory Commission, 2012). In this document, procedure development and training programme development are equal parts of the design phase together with human-system interface design.

The previously mentioned report by the OECD/NEA Committee on Safety of Nuclear Installations (2005) stated that there is a need for guidelines and tools to support the modification process in incorporating human factors assessments, among other areas. There are several standards proposing processes or ways for how human factors aspects are to be included in design, such as "*ISO 6385:2004 Ergonomic principles in the design of work systems*" (International Standard Organisation, 2004) and "*ISO 11064-1:2000 Ergonomic design of control centres – Part 1: Principles for the design of control centres*" (International Standard Organisation, 2000).

ISO 6385:2004 advocates the iteration of certain activities in various phases of the design process: analysis, synthesis, simulation and evaluation. The phases suggested are: formulation of goals (requirements analysis); analysis and allocation of functions; design concept; detailed design; realisation, implementation and validation; evaluation. ISO 11064-1:2000, which specifically concerns ergonomic design of control centres, also emphasises the iterative nature of the process. This standard presents a framework for an ergonomic design process consisting of the following phases: clarification; analysis and definition; conceptual design; detailed design; and operational feedback. The processes are largely similar, the major difference being the inclusion of realisation and implementation in the ISO 6385:2004 process that has no correspondence in the ISO 11064-1:2000 process. Realisation includes the building, production or purchase of the work system and its installation in the place of operation. Realisation should also include fine-tuning of the system in accordance with local context. Implementation includes introducing the work system to all people concerned with it, for instance through information and training (International Standard Organisation, 2004). It should be noted, however, that the development of training regimes and the like is included in detail design in the ISO 11064-1:2000 process. The overall correspondence between phases in ISO 6385:2004 and ISO 11064-1:2000 is shown in Table 4.

One difference in these processes in Table 4 when compared to the ones summarised in Table 3 is the emphasis on acquiring operational feedback after the design has been in operation for some time. The purpose is to continuously check on the validity of the design of the control centre during its lifespan (International Standard Organisation, 2000).

Table 4: Overall correspondence between phases in ISO 6385:2004 and ISO 11064-1:2000.

ISO 6385:2004 Ergonomic principles in the design of work systems	**ISO 11064-1:2000 Ergonomic design of control centres – Part 1: Principles for the design of control centres**
Formulation of goals (requirements analysis)	Clarification
Analysis and allocation of functions	Analysis and definition
Design concept	Conceptual design
Detailed design	Detailed design *[validation included here]*
Realisation, implementation and validation	*[no correspondence to realisation and implementation]*
Evaluation	Operational feedback

4.2 EVALUATION IN THE DEVELOPMENT PROCESS

A specific issue raised at the workshop that formed the foundation for the aforementioned OECD/NEA report was that the process for modification of nuclear power plats should include actions to verify the fulfilment of requirements and validate the appropriateness of the modification (OECD/NEA Committee on Safety of Nuclear Installations, 2005). The same issue, but regarding safety in complex technical systems involving major accident hazards, was raised by Hale et al. (2007), who noted that one similarity between development processes was iterative safety checks in conjunction with decisions to move on to the next design phase, to ensure the focus on safety issues as the design develops.

Ullman (1997) describes design as the successive development and application of constraints to reduce the number of potential solutions to a problem, until only one unique product remains. The majority of constraints follow as a consequence of design decisions, and sometimes as a consequence of the absence of decisions. Choosing one alternative solution over others means adding constraints that make the rejected solutions unsuitable. Comparing the contraints to alternative soloutions is called evaluation, and the best alternative can be chosen based on the result of the evaluation. This view is in line with the dictionary definition of the word evaluation, "*the making of a judgment about the amount, number, or value of something*" (Oxford Dictionaries, 2015a).

Evaluations in the development process can be categorised in several ways, one way being the purpose for which the evaluation takes place. Evaluations can be divided into formative and summative, where the former has the purpose of improving the object that is being evaluated and the latter is meant to provide a concluding quality assessment. Due to this difference in purpose formative evaluations are usually performed during the development process and summative on the finished design (Noyes, 2004).

A different categorisation of evaluations can be made between verification and validation. An ergonomics standard for the ergonomic design of control centres (International Standard Organisation, 2006, p. 1) defines the evaluation process as the "*combined effort of all verification and validation (V&V) activities in a project using selected methods and the recording of the results*" – hence using the word 'evaluation' as an overall concept containing the more specific activities verification and validation. The same standard defines verification as "*confirmation, through the provision of objective evidence, that specified requirements have been fulfilled*" (International Standard Organisation, 2006, p. 2) and validation as "*confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled*" (ibid). Engel (2010, p. 19), having studied definitions of verification and validation from multiple sources, adopted the definitions that verification is "*the process of evaluating a system to determine whether the products of a given lifecycle phase satisfy the conditions imposed at the start of that phase*" and that validation is "*the process of evaluating a system to determine whether it satisfies the stakeholders of that system*" (ibid). Verification, in both these definitions, has the purpose of assessing the fulfilment of requirements (or differently put, satisfaction of conditions). The definitions of the term 'validation' are less similar, but both embrace a more holistic assessment than the one done in verification concerning the satisfaction of requirements of use or stakeholders.

4.3 HUMAN FACTORS EVALUATION MEASURES AND METHODS

Meister (2001) defines measurement as the analytic phase that determines the questions to be asked and how measurement will provide the answers sought, as well as the following collection of data and a concluding analysis phase that examines what the measurement means. A similar description can be provided of the process of evaluation:

1. Determine the evaluation measures. What is the purpose of the evaluation and of the system, and how can this be operationalised?
2. Determine acceptance criteria. At what level is the measure regarded as good enough?
3. Collection of data. Determining the value of evaluation measures for the object to be evaluated.
4. Comparison of the collected data with acceptance criteria, making a judgement.

The issue of determining acceptance criteria is emphasised by the Institute of Electrical and Electronics Engineers (1999). Their guide for evaluation of human-system performance in nuclear power generating stations states that the interpretation of results in an evaluation requires the specification of criteria for judging the acceptability of the human-system performance. The guide differentiates between informal criteria, "*evaluator's opinion regarding the acceptability of the performance*", and formal criteria, such as "*operator diagnosis within a specific time limit*" (Institute of Electrical and Electronics Engineers, 1999, p. 3). Acceptance criteria can be determined in several ways. For example, the human factors engineering programme review model NUREG-0711 of the United States Nuclear Regulatory Commission (2012) describes four different bases for this:

- *Requirement*: Quantified performance requirements for the performance of systems, subsystems, and personnel are defined through engineering analyses.
- *Benchmark*: A benchmark system, a current system deemed to be acceptable, is used to define acceptance criteria. This can be done by evaluating the same system before and after change.
- *Norm*: Instead of a single benchmark system many predecessor systems can be used to create a norm against which the system under evaluation is compared.
- *Expert Judgment*: Subject-matter experts establish the acceptance criteria.

One tool in performing evaluations is that of evaluation methods. The dictionary definition of a method is "*a particular procedure for accomplishing or approaching something, especially a systematic or established one*" (Oxford Dictionaries, 2015b), hence an evaluation method is a systematic procedure for making a judgement about something.

Numerous human factors evaluation methods exist, and in this licentiate thesis they are categorised by what they measure and how that measurement is performed. Categorisation according to evaluation measures is inspired by compilations and reviews of human factors evaluation methods (Stanton et al., 2005, International Standard Organisation, 2006, Le Blanc et al., 2010, Savioja et al., 2014). In the present work, evaluation measures are divided into the following categories:

- *System performance*. Measures the performance of the whole system together. In the case of control room systems, this could be measuring crucial plant parameters such as tank levels and temperatures.
- *Task performance*. Measures the performance of tasks, such as the number and nature of errors, or time.
- *Teamwork*. Measures meant to assess the quality of team-based activity.
- *Use of resources*. Measures meant to assess different aspects of the operators' use of their mental and physical resources, such as situation awareness, mental workload, and physical load.
- *User experience*. Measures assessing the feelings and emotions of the operators. The definition of user experience measures utilised here is the one by Savioja et al. (2014, p. 429): measures that indicate "*the users' subjective feeling of the appropriateness of the proposed tool for the activity*".
- *Identification of design discrepancies*. The appropriateness of the system is evaluated by assessing its compliance with an ideal. This can be done explicitly by comparing the design of the system with guidelines, or implicitly by allowing experts assess the system's quality.

Advantages and disadvantages can be identified for each of the categories described above. Measures of system and task performance have the advantage of being closely connected to the system goal, but have the disadvantage that the evaluation result provides little guidance on how the identified issues should be addressed. In other words, these measures point out the problem, but tell us little about what we should do about it. In contrast, the identification of design discrepancies displays in much greater detail how issues should be remedied. Accurately measuring task performance requires a definition of what constitutes a deviation from the correct execution of tasks. Measures such as errors and time are situation- and system-dependent, and must be judged as such (Le Blanc et al., 2010).

Teamwork, use of resources, user experience, and identification of design discrepancies all have the ability to identify problem areas that may not show directly in system or task performance during an evaluation. Their advantage is that they can identify issues that may lead to insufficient performance in a slightly different context than the exact one tested (for instance with more inexperienced operators or another combination of events). The use of these measures is motivated by the assumption that a control room evaluation cannot, for practical reasons, recreate every possible situation that might occur in the control room system's lifetime.

The other categorisation of human factors evaluation methods in this licentiate thesis is by the nature of the studies in which the measures are taken (how measurement is done): empirical and analytical (Osvalder et al., 2009). Empirical studies are direct studies of use where users carry out tasks using actual (or mock-ups of) systems. Analytical studies more indirectly study use by letting different subject-matter experts investigate the system analytically without any actual use taking place. These categories are not discrete, but rather more of a spectrum. For example, a talk-through with an operator using a very early prototype would be closer to the analytical side of the spectrum since the interaction with the prototype may be very different from how interaction would be with the finished system.

The two categories described above, what and how measures are taken, can be combined to create 12 groups of evaluation methods. Finding actual evaluation methods suitable for early evaluation of control room systems does not lie within the scope of this licentiate thesis. However, examples of methods were needed to better describe the groups of evaluation methods. An initial review of existing methods was undertaken and the result can be seen in Table 5. This review was not able to distinguish examples for all groups. A more thorough review of existing methods will be performed as part of the future work within this doctoral project.

Table 5: Examples of methods for each of the 12 groups of evaluation methods.

		How measures are taken	
		Empirical (direct study of use)	**Analytical** (indirect study of use)
What measures are taken	**System performance**	Automated or manual logging of plant parameters (Le Blanc et al., 2010)	
	Task performance	Automated or manual logging of errors or response times (Le Blanc et al., 2010) Eye-tracking (Le Blanc et al., 2010)	Human Reliability Analysis techniques; such as Human Error Assessment and Reduction Technique, and Cognitive Reliability and Error Analysis Method (Stanton et al., 2005) Cognitive walk-through (Osvalder et al., 2009)
	Teamwork	Co-ordination Demands Analysis (Stanton et al., 2005) Questionnaires for Distributed Assessment of Team Mutual Awareness (Stanton et al., 2005)	
	Use of resources	NASA-Task Load Index (Le Blanc et al., 2010) Situation Awareness Control Room Inventory (Le Blanc et al., 2010) Rapid Upper Limb Assessment (Osvalder et al., 2009)	Predictive Subjective Workload Assessment Technique (Stanton et al., 2005) Rapid Upper Limb Assessment (Osvalder et al., 2009)
	User experience	UX-questionnaire from Framework for evaluating systems usability in complex work (Savioja et al., 2014) Geneva Emotion Wheel (Sacharin et al., 2012)	Anticipated experience evaluation (Gegner and Runonen, 2012)
	Identification of design discrepancies	Heuristic analysis (Osvalder et al., 2009) Physical measurement, such as of acoustics, lighting, and thermal conditions (International Standard Organisation, 2006) Questionnaire for User Interface Satisfaction (Stanton et al., 2005)	Heuristic analysis (Osvalder et al., 2009) Delphi technique (International Standard Organisation, 2006)

Reliability and validity are two concepts important in the use of methods. In a research tradition concerned with quantitative data there is a general agreement on the definition of these concepts. High reliability within this tradition is when repeated measurements of the same object deliver the same result. Validity indicates how well what is meant to be measured actually is measured (Svensson and Starrin, 1996). Within the research tradition concerned with qualitative data, however, these concepts are handled somewhat differently. Reliability in this tradition must be viewed in its context. For example, one cannot simply compare two answers from different interviews and consider the question reliable if the answers are alike. Reliability cannot be judged without also judging the validity of the question in the context in which it was asked. Identical questions in two different interviews can be considered reliable even if the answers differ (Svensson and Starrin, 1996). The definition of the concept of validity within the qualitative research tradition can be divided into two overall views. One considers the concept usable for studies with qualitative data, the other considers terms such as authenticity or trustworthiness to be more relevant. However, their approach for testing validity is often similar, namely that validity should be judged in relation to context and the persons involved (Svensson and Starrin, 1996). Hammersley and Atkinson (1983), referenced in Svensson and Starrin (1996), claim that a method in itself is neither valid nor invalid. Validity is not a characteristic inherent in a specific method, but belongs to the data, presentation, and conclusions that have been reached by using the method in a specific context for a specific purpose.

4.4 EVALUATION METHODS IN EARLY STAGES OF THE DEVELOPMENT PROCESS

In a special issue of Safety Science on safety in design Hale et al. (2007) point out that safety imposes additional requirements on the design and design process and may add to costs, decreasing profit margins and market share. To address these challenges, Hale et al. (2007) state that safety implications have to be taken into account early on, otherwise it may be necessary to implement expensive and less user-friendly safety add-ons later on. Papin (2002) advocates the same for nuclear power plants, stating that most of the situations seriously challenging the operators' performance have their origin in design decisions taken early in the development process. For nuclear power plants, these design deficiencies concern reactor system design rather than operator interface design (ibid).

Savioja (2014) stresses the importance of performing evaluations early in the design process. Boring (2014) does the same, stating that the feedback that evaluation early in the design process can provide will help ensure that errors in human-system integration are eliminated rather than incorporated into nuclear power plant control room system design. Boring (2014) concludes that there is a need to perform evaluations earlier in the design cycle. As described in section 3.4, the review guide NUREG-0711 (United States Nuclear Regulatory Commission,

2012) is widely used to guide control room system modernisations within the nuclear power domain. It states that a structured methodology, including tests and evaluations during the design phase, should be used to guide design work. However, as Boring (2014) points out, it lacks explicit guidance for human-system interaction evaluations during the design phase.

Earlier evaluation seems to be desirable, but what does this imply? As was described in section 4.1, the development process is often divided into phases with a gradually increasing level of detail in the developed system. In the development process described by Pahl and Beitz (1996) that was described in section 4.1, the level of detail in the design is divided into the following main stages:

- *Detailed product proposal* (phase Planning and clarifying the task): a clarification of the task that includes information about the requirements that have to be fulfilled by the product, as well as the existing constraints and their importance.
- *Conceptual design*: a specification of principle that establishes function structures and working principles.
- *Embodiment design*: a specification of overall layout design (general arrangement and spatial compatibility), preliminary form designs (component shapes and materials) and production processes.
- *Detail design*: a specification of arrangements, forms, dimensions, and surface properties of all individual parts.

Hollnagel (1985) classified evaluation methods in a way that connects to the different levels of detail in the design specified by Pahl and Beitz (1996). The classification is done according to how the system being evaluated is represented in the evaluation. Four types of system representation are specified:

- *Conceptual*, where the system is not represented physically but by a description of its functional characteristics.
- *Static simulation*, where the system is represented by samples taken from preliminary performance recordings, for example a series of frozen frames focused on how information is presented to the operator.
- *Dynamic simulation*, where the entire process is simulated and the operators react to the simulated process.
- *The real system*, the design implemented in its intended context.

Comparing them suggests that conceptual system representations would be available when the design is in the conceptual design stage, and possibly also in the detailed product proposal phase. Static simulation system representations could be created when the design is in the embodiment design and detail design phase. Dynamic simulations could be done when the design is in the detailed design phase.

The use of system representations available in the detailed design phase will in some aspects differ from the use of system representations available earlier in the development process. One example of such aspects is the time to complete tasks, the time it takes to perform a task using a paper mock-up is not necessarily representative of the time it takes with the final product. A method designed to assess aspects of actual use might therefore not be suitable when system representations are more conceptual. Thus analytical methods that utilise indirect studies of use would be more suitable for earlier evaluations.

EVALUATION MEASURES FOR NUCLEAR POWER PLANT CONTROL ROOM SYSTEMS

The first step in the evaluation process is to determine the evaluation measures. With regard to control room systems, if the purpose of the evaluation is to assess the control room system's ability to support safe operation – how can this be measured? This chapter will explore which of the categories of evaluation measures presented in section 4.3 are relevant for control room systems.

5.1 MEASURES IN EMPIRICAL CONTROL ROOM SYSTEM EVALUATIONS

Evaluations to assess the suitability of nuclear power plant control room systems are nothing new, since after being built they must be maintained as well as modified and modernised. This section reviews performed and planned nuclear power plant control room system evaluations presented in academic literature. Since the focus of this licentiate thesis is evaluation of the control room system as a whole studies evaluating only smaller parts of the control room are not included in this review. The purpose is to gain an overview of the measures used to evaluate control room systems and to compare these to the categories presented in section 4.3.

In her doctoral thesis, Savioja (2014) undertook a literature review of empirical studies of control rooms in the nuclear power domain. 22 empirical studies were reviewed, and six of them were labelled as studying the totality of the control room. However, one of the studies was excluded here since its purpose was to investigate human error probabilities for the purpose of developing Human Reliability Analysis methods, not to evaluate the control room. In the studies, a number of different measures were used to assess the control room design. Savioja (2014) summarises them into the following categories: plant performance; task performance (time); task performance (errors); situation awareness; workload/task load; teamwork and communication; anthropometric measures; physiological measures; usability; expert opinion concerning error probability; and expert opinion concerning safety (Table 6). The most popular categories of measures in the reviewed studies were task performance (errors), situation awareness, workload/task load, and teamwork and communication. The expert opinion in the study by Hwang et al. (2009) was collected by prompting the operators to discuss design of interface, utilization of procedures, task processes, members' communication, situation awareness, source of information, and mental workload from the point of view of error probabilities and errors that had occurred during the operators' training period. The expert opinion concerning safety in the study by Luquetti dos Santos et al. (2009) stems from a questionnaire where each feature of the control desk related to the panel layout, panel label, information display, controls and alarms, was rated according to its conformance to a certain human factors guideline and weight of importance.

An interesting issue is the reasoning behind the choice of measures in the studies reviewed by Savioja (2014). Two studies (Ha et al. and Luquetti dos Santos et al.) based their choice of measures on guidance documents from the United States Nuclear Regulatory Commission (1997, 2002, 2012). The human factors engineering program review model NUREG-0711 (United States Nuclear Regulatory Commission, 2012) and the more detailed integrated system validation document NUREG/CR-6393 (United States Nuclear Regulatory Commission, 1997) proposes the following categories of measures: plant performance; primary and secondary task performance; situation awareness; workload; and anthropometry and physiology. In NUREG/CR-6393, each of the categories is described in the context of its connection to the control room system's support of safe operation. NUREG-0700 (United States Nuclear Regulatory Commission, 2002) contains human-system interface design review guidelines.

Gatto et al. (2013) used operating time to assess the physical layout of the control room since the location of adequate interfaces was considered crucial for promptly identifying abnormalities and responding appropriately. Hwang et al. (2009) and Chuang and Chou (2008) provide no clear reasoning behind their choice of measures. Table 6 summarises the studies in terms of the measures that were used and the source of or reasoning behind these measures.

In addition to the studies in the literature review by Savioja (2014), a search was conducted for additional control room evaluation studies in academic literature. Preparation (i.e. not yet empirically tested) for an integrated system validation of a nuclear power plant control room by Li et al. (2012) was identified in this search. The rationale for the selection of measures was not explicitly stated, but the process used is described as corresponding to the elements in the human factors engineering program review model NUREG-0711 (United States Nuclear Regulatory Commission, 2012). The same argument is made in a paper by Liu et al. (2012), although the measures taken are slightly different. Table 7 presents a summary of these studies.

These examples show that a range of measures are used in empirical control room evaluations. All but one of the categories of measures presented in section 4.3 are represented. The plant performance, task performance, and teamwork and communication measures categories of Savioja (2014) correspond well to the system performance, task performance, and teamwork measures categories in section 4.3. Situation awareness and workload/task load can be accommodated within the use of resources category. Anthropometric and physiological measures as they were used by Ha et al. (2007) and Li et al. (2012) fit into the identification of design discrepancies category. The usability measures used by Chuang and Chou (2008) and Li et al. (2012) are not adequately detailed in their papers, but seems to be measures belonging to the identification of design discrepancies category as well.

Table 6: Summary of empirical studies of NPPs evaluating the totality of control room adapted from Savioja (2014) with the addition of source of/reasoning behind measures or explored concepts. *References for NUREG-documents: NUREG/CR-6393 - United States Nuclear Regulatory Commission (1997);NUREG-0711 - United States Nuclear Regulatory Commission (2012); NUREG-0700 - United States Nuclear Regulatory Commission (2002).

Study	**Plant performance**	**Task performance, time**	**Task performance, errors**	**Situation awareness**	**Workload/task load**	**Teamwork and communication**	**Anthropometric measures**	**Physiological measures**	**Usability**	**Expert opinion concerning error probability**	**Expert opinion concerning safety**	**Rationale for selection of utilised measures or explored concepts**
Ha et al. (2007)	x		x	x	x	x	x	x				Each utilised measure is described in the context of its connection to the control room system's support of safe operation. NUREG-0711* and NUREG/CR-6393* are given as references as well.
Chuang and Chou (2008)				x	x	x			x			No explanation is given.
Hwang et al. (2009)			x							x		No explanation is given.
Luquetti dos Santos et al. (2009)											x	NUREG 0700*
Gatto et al. (2013)		x										Used operating time to assess the physical layout of the control room since the location of adequate interfaces was considered crucial for promptly identifying abnormalities and responding appropriately.

Table 7: Summary of planned and executed empirical studies of NPPs evaluating the totality of control room **not part of the literature review by Savioja (2014)**.
**Reference for NUREG-0711: United States Nuclear Regulatory Commission (2012).

Study	Plant performance	Task performance, time	Task performance, errors	Situation awareness	Workload/task load	Teamwork and communication	Anthropometric measures	Physiological measures	Usability	Expert opinion concerning error probability	Expert opinion concerning safety	Additional measures/comments	Rationale for selection of utilised measures or explored concepts
Li et al. (2012)	x		x	x	x	x		x	x			Goal achivement also measured. Only plan for evaluation, no description of execution.	Not explicitly stated, but majority of measures are identical to those suggested in NUREG-0711**
Liu et al. (2012)	x		x*		x*	x*						*Acceptance criteria reviewed by experts: control the unit smoothly; have access to monitor the specified parameters; can deal with relative alarms; and communication between the team shifts is unhindered. *Operators queried on: communication; monitoring and control of unit; and subjective workload.	Not explicitly stated, but the process used is described as similar to NUREG-0711**

The measure 'expert opinion concerning error probability' by Hwang et al. (2009) corresponds to the task performance category. Lastly, the measure 'expert opinion concerning safety' is very general and its rationale for belonging to a category depends on the use of the measure in the study in question. The way Luquetti dos Santos et al. (2009) use this measure makes it a part of the identification of design discrepancies category. The only category of measures from section 4.3 not utilised in the reviewed evaluations was user experience.

As described above, the rationale behind the choice of evaluation measures in the reviewed studies was not always clear. Several of the studies used guidance documents from the United States Regulatory Commission as a reference, which in turn motivates its categories of evaluation measures with their connection to the control room system's support of safe operation. It is worth noting here that the documents NUREG/CR-6393 and NUREG-0711 (United States Nuclear Regulatory Commission, 1997, 2012) do not specifically mention teamwork as a performance measure, although this was assessed in the studies referring to them.

5.2 FRAMEWORKS FOR SELECTION OF EVALUATION MEASURES

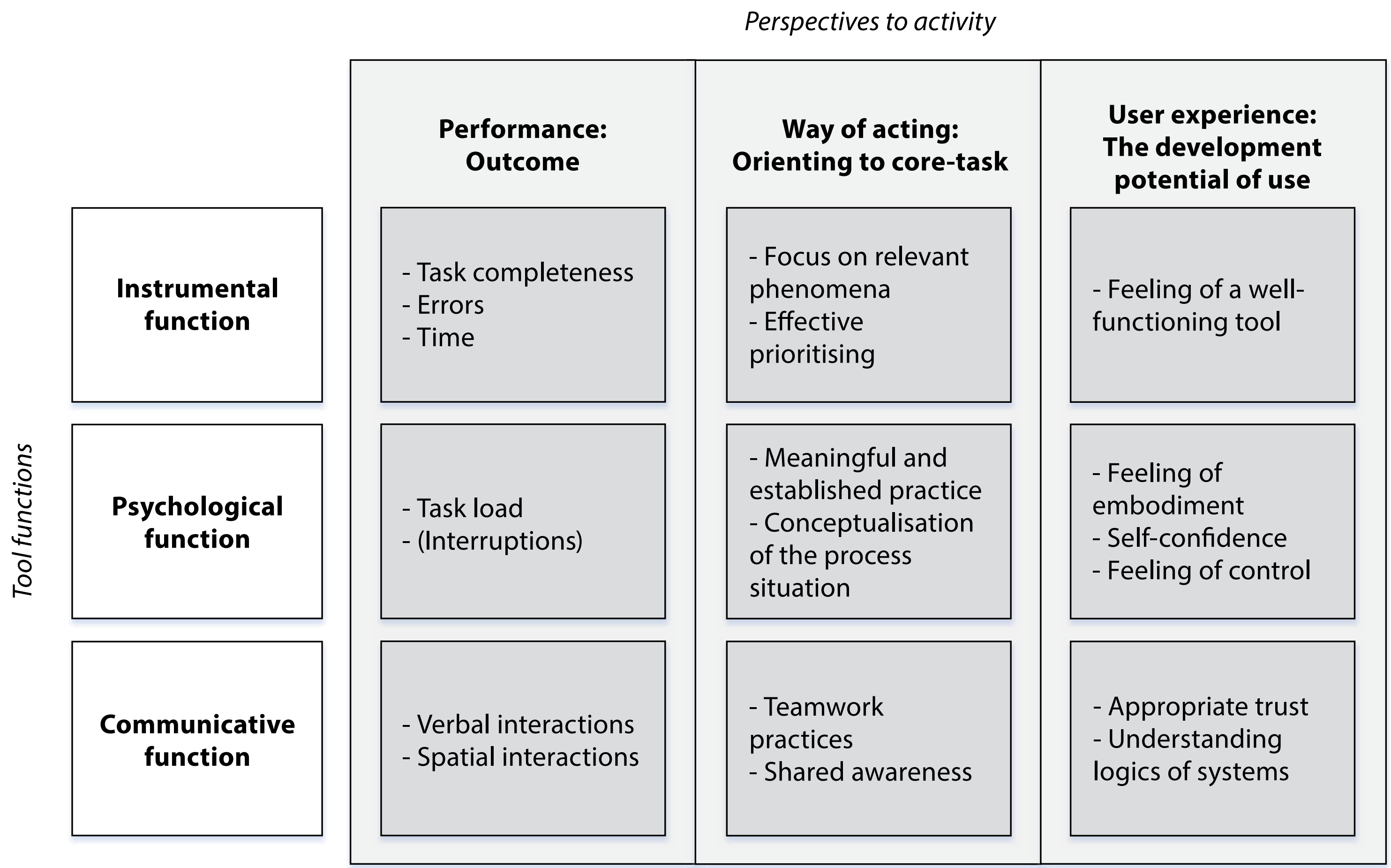
As was seen in the review of empirical control room evaluations in section 5.1, systematic operationalisation of the control room's system purpose as a basis for determining evaluation criteria is not always undertaken. To support this activity, researchers have developed frameworks to guide the selection of evaluation measures for control rooms. This section presents two such frameworks. These two frameworks have been developed in the context of nuclear power plant control room evaluation, which is the reason for their inclusion in this licentiate thesis.

When discussing reviewed empirical control room evaluations, Savioja (2014) concluded that the evaluation methodologies used tend to simplify and generalise the operating work, which may weaken the relevance of the results of these evaluations. She continued by stating that "*In a control room study, the evaluation framework should preferably be such that the complexities of everyday work of operating crews in NPPs can also be addressed*" (Savioja, 2014, p. 53). In response to this, she presents an evaluation methodology to assess systems usability in control rooms. The methodology is based on activity theory, and its framework is meant to be a conceptual tool that aids in finding relevant measures for a comprehensive evaluation of systems usability in the control room. The scope of the research covers only empirical user tests assessing use of the system. Systems usability is described as the capability of the technology to fulfil the instrumental, psychological, and communicative functions of the tool in the activity, and to support the fulfilment of core-task demands in the work. The instrumental function regards the tool's ability to bring about the desired effect on the plant's process. The psychological function is the tool's ability to enable psychological processes and provide the operators with external means to control their own behaviour. The communicative function is the social aspects of using the tool, the role of the community that uses the tool for the same activity. Core-task demands are demands of the main content of the work that relate to the actors' skills, knowledge, and collaboration.

Savioja (2014) states that systems usability is evidenced in technology usage in the appropriate performance outcome, way of acting and user experience. To assess only performance outcome is not deemed sufficient since the multitude of barriers (technical, organisational etc.) in complex socio-technical systems are designed to neutralise the effect of performance variance on the outcome, which means that the performance outcome measure will not be sensitive enough to assess variation in the tools (i.e. design changes in the control room). Therefore, the more indirect measure 'way of acting' must also be evaluated. The methodology also assesses the concept of user experience. This is motivated by the difficulty of analytically understanding all possible implications the tool (the control room) will have on the activity, thus making the utilisation of professional operators' experiences important. In addition, Savioja (2014) refers to the view of Kaptelinin and Nardi (2012), who state that the user's emotions are indicators of the status of the activity as a whole. It is also stated that user experience is an indication concerning the development potential of the tool. User experience is defined as "*the users' subjective feeling of the appropriateness of the proposed tool for the activity*" (Savioja et al., 2014, p. 429). Savioja (2014) states that the overbearing quality of user experience is that the user feels that the technology has the potential to develop into a meaningful tool for the activity and benefits interaction with the object of the activity. In short, the methodology evaluates the concept of systems usability by assessing the instrumental, psychological and communicative functions of a tool from the different perspectives of performance, way of acting and user experience. Figure 2 shows an example of how the framework has been used to identify measures.

Another example of a framework to support the development of evaluation criteria is suggested by Braarud and Rø Eitrheim (2013). They state that "*current models of control room work lack integrated descriptions that consider both the physical representations, couplings to cognitive support, discrimination between individual and team demands, and relation to the current situation and process state*" (Braarud and Rø Eitrheim, 2013, p. 17). Because of this, they suggest a framework with a model that covers the control room functionalities, possible physical representations and support for safe and effective performance of the tasks by the team. It is meant to be used as a basis for a criterion-referenced validation of the control room system. A criterion-referenced approached is described as sharing elements with the requirement and expert judgement-based determination of acceptance criteria by the United States Nuclear Regulatory Commission (2012) described in section 4.3. The difference is said to be that the criterion-referenced approach does not necessarily need to be based on a formal engineering analysis. However, the authors state that the framework in its current state could equally well be used as a basis for improvements of the benchmark approach or other approaches

Figure 2: An example of how the framework of Savioja (2014) has been used in studies in Finnish nuclear power plants to identify measures. Figure adapted from Savioja (2014).



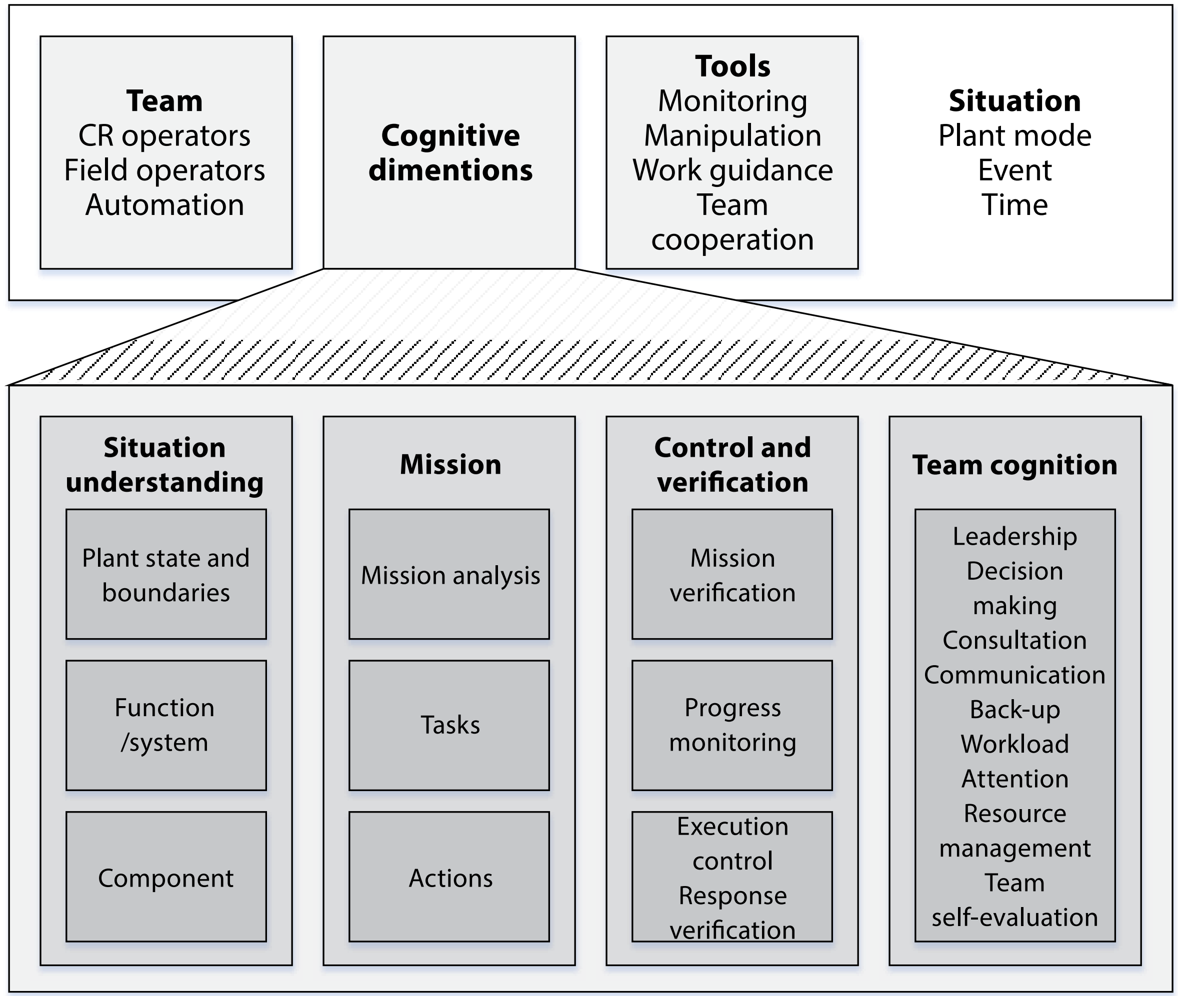
to determining acceptance criteria. The framework focuses on phenomena that can be observed in performance based testing. The purpose of the framework is to guide the categories and dimensions for which to develop evaluation and acceptance criteria.

The model in their framework consists of four main parts: team, cognitive dimensions, tools, and situation (see Figure 3). The *team* are the 'agents' interacting directly with the plant process. *Cognitive dimensions* contain characteristics of how process control is performed by the agents. The *team cognition* part consider characteristics of how interaction between team members is performed, the other three cognitive dimensions parts mainly regard work more directly connected to controlling the process. The inner levels of the *situation understanding*, *mission*, and *control and validation* parts are organised according to an abstraction hierarchy. To be able to perform process control, the agents need means and support, the *tools*. *Situation* is the setting and the plant process, including the initiation of and outcome of the agents' process control.

The framework by Braarud and Rø Eitrheim (2013) is still under development. The forming of the current version was focused on the theoretical background and the main elements of the framework. The development of performance criteria, measurements and observational techniques for evaluation purposes will be part of the authors' future work.

The framework of Savioja (2014) includes measures from all the categories presented in section 4.3, with the exception of the identification of design discrepancies category. Instrumental performance correspond to the system and task performance categories. The latter also cover instrumental way of acting. Psychological performance and way of acting relate to measures within the use of resources category. The teamwork category cover measures for all three perspectives of the communicative function. Lastly, the user experience category of measures corresponds to all three functions within the user experience perspective of Savioja. The framework of Braarud and Rø Eitrheim (2013) clearly emphasises the importance of the teamwork category by including the team cognition part in their model. The description of the cognitive dimensions suggests that it is important to evaluate system performance, task performance, and use of resources. However, this framework does not propose measures as clearly as Savioja's, nor is it possible to map its contents as directly to the categories of measures.

Figure 3: The model in the framework suggested by Braarud and Rø Eitrheim (2013). Figure adapted from Braarud and Rø Eitrheim (2013).



5.3 PAPER I – SAFE OPERATION ASPECTS IN SWEDISH NUCLEAR POWER PLANTS

This section summarises the result of Study I (described in section 2.1). The aim of Study I was to find aspects of the control room system that contribute to safe operation from a human factors perspective, to be able to identify a foundation for evaluation measures. The study in its totality is described in Paper I.

The data from the interviews in Study I was organised into five overall themes: *situations*, *functions*, *tasks*, *characteristics* and *structural elements*. The situations theme consisted of states in which the control room system may be in and events in the environment surrounding the control room system. These are situations that the control room system must be able to handle, such as different operational modes (startup, power operation, shutdown, and outage) and different types of disturbances. Functions were the abilities the control room system must have, the abstract capabilities of the control room system that are realised by the control room system design. Examples of functions stated were: presentation of information; having established codes of conduct (rules); and having a distribution of responsibility (roles). Tasks were what the control room system, either its operators or technical systems, must be able to perform. Tasks could be both 'as imagined' (such as actions described in procedures) and 'as done' (what the operators actually do), and were divided into: primary tasks (those directly connected to supervising and controlling the plant's process), way of working, communication, and cooperation. Structural elements were the physical or social entities that realise the control room system. The statements of the structural elements theme were categorised into five sub-themes: operator interfaces; physical control room design; process and instrumentation and control (I&C) systems; support systems; and personnel. Support systems were for example procedures and routines. Finally, characteristics were conditions establishing how artefacts should be designed and how personnel should be and behave. Characteristics of the design included traits such as clarity, consistency, and error tolerance. Characteristics of the operators were for example competence (especially understanding the plant and its process), curiosity, and flexibility. Further description of the contents of the themes is found in Paper I.

The interviewees' prioritisation between aspects that contributed to safe operation differed. Apart from aspects concerning plant status (overview, understanding, and the operator interface presenting it), which was stated by many as contributing most, the interviewees' answers had little in common. Neither were they in agreement on which aspects contributed least to safe operation.

The results of Study I are largely confirmed by the framework by Braarud and Rø Eitrheim (2013). The team and tool categories of their model describe the same entities that were included in the structural elements theme in Study I (the only exception being that the structural element of process and I&C systems is not as emphasised in the Braarud and Rø Eitrheim framework). Furthermore, the team category consists of subgroups according to the functionality of the tools, which mirrors some of the functions noted in Study I. Situations for which the control room system should be designed and used are present in both the Study I themes and the Braarud and Rø Eitrheim framework. The cognitive dimensions of the model details functions and tasks stated in Study I.

The framework by Savioja (2014) also confirms parts of Study I. The user experience perspective on activity strongly emphasised by Savioja was mentioned by the interviewees in Study I. The way of acting perspective in Savioja's framework supports the interviewees' statements regarding the importance of the way of working in Study I. The emphasis on the importance of communication and cooperation in Study I is mirrored in the communicative function of the tool in Savioja's framework.

Assessing the aspects contributing to safe operation resulting from Study I would require measures from the categories of system performance, task performance, teamwork, use of resources, and identification of design discrepancies. The user experience perspective was mentioned, but not emphasised.

5.4 PAPER II – TRENDS IN CONTROL ROOM SYSTEM DESIGN

The study by Osvalder and Alm (2012) initiating the work in this doctoral project questioned whether methods used today are generally adapted to conditions in traditional control rooms and not suitably adapted to changes in modernised and future control rooms. Study II (2.2) was conducted to explore possible usability problems that it is important to attend to in the design of future Swedish nuclear power control room systems and suggest requirements that human factors evaluation methods must fulfil. The control room system design trends that were identified in Study II (and presented in Paper II) were:

- More software based presentation
- Variety in operator interfaces
- Centralisation of information
- Increase in level of automation
- Addition of new process systems or functions
- Implementation of digital operating procedures
- Increase in turnover of personnel

An increase in software based presentation was also identified at the previously mentioned workshop organised by a Nuclear Energy Agency committee (OECD/NEA Committee on Safety of Nuclear Installations, 2005), where it was also concluded that access to operator interfaces are becoming more serial. The latter is known as the keyhole effect and was also noted by Carvalho et al. (2008) as one of the most important consequences of a transition from the traditional analogue instrumentation and control technology to digital technology. The keyhole effect is a consequence of the smaller space for presentation of information that a normal computer screen provides, and information must be used in a sequential manner as opposed to parallel access as is possible with analogue control panels.

The OECD/NEA Committee on Safety of Nuclear Installations (2005) also concluded that interactions are increasingly conducted through computer systems as opposed to through crew interaction with plant systems and components. This finding can be connected to the trend of information centralisation, to locate even more supervision and control to the main control room, identified in Study II. The OECD/NEA workshop also noted that the functionality of operator interfaces is being expanded, something that also was concluded in Study II.

As a consequence of these control room system design trends, Study II suggested a number of demands on human factors evaluation methods. The correspondence between these and the categories of evaluation measures from section 4.3 is shown in Table 8.

Two of the requirements from Study II, identification of usability problems connected to automation and evaluation of digital operating procedures, point to more complex problems that cannot be assessed by single categories of evaluation measures. As in the case with larger complex socio-technical systems, such as control room systems, they require the use of measures from several categories. For example, this was the case in two evaluation studies of computer-based operating procedures by Le Blanc and Oxstrand (2013).

5.5 PAPER III – THE RESILIENCE ENGINEERING PERSPECTIVE

As was described in section 3.2, resilience engineering has been proposed as an approach that complements the traditional view of safety to increase safety in socio-technical systems. However, this is not an approach that is used in the empirical control room system evaluations reviewed in section 5.1. Both Savioja (2014) and Braarud and Rø Eitrheim (2013) on the other side, acknowledge the importance of the resilience perspective when evaluating safety in complex socio-technical systems. Savioja (2014) inparticular points out how evaluating the way of acting and user experience are methods of assessing a system's potential for resilient performance.

Another way of assessing a system's potential for resilient performance is the Resilience Assessment Grid, abbreviated RAG (Hollnagel, 2011a). RAG is a question based tool that support resilience management by assessing four basic abilities of a system. The four basic abilities that make resilient performance possible, proposed by Hollnagel (2011b) and also known as the cornerstones of resilience, are: respond, monitor, anticipate, and learn. Respond is the ability to know what to do, i.e. how to respond to expected and unexpected events using pre-defined responses or by adjusting the system's normal functioning. Monitor is to know what to look for, i.e. that which is or can be a threat in the near future.

TABLE 8: Correspondence between demands on human factors methods from Study II and categories of evaluation measures from section 4.3

Demands on human factors methods from Study II, they must be able to:	**Corresponding category of evaluation measures from section 4.3**
Identify inconsistencies between operator interfaces, both in appearance and interaction	Identification of design discrepancies
Identify work tasks that create extreme levels of mental workload	Use of resources
Identify inefficient work tasks	Task performance
Identify conflicting work tasks	Task performance
Identify information and control devices necessary for different work tasks and users with various experience levels	Task performance
Identify usability problems connected to the level of automation and presentation of system automation	Would require a combination of measures from several categories
Identify inconsistencies between the operators' mental models of the system and the system itself	Task performance, use of resources
Evaluate situation awareness (both collective and individual) and identify potential usability problems that might decrease awareness of the status of the process	Use of resources
Evaluate operator workload to identify related potential usability problems	Use of resources
Evaluate cooperation within the shift team to identify related potential usability problems	Teamwork
Evaluate the design and use of digital operating procedures to identify potential usability problems	Would require a combination of measures from several categories

Anticipate is to know what to expect, i.e. to anticipate future changes and their consequences. Learn is to know what has happened and to learn the right lesson from experiences of successes and failures alike. According to Hollnagel (2011a), the first step when applying RAG is to define and describe the system to be assessed. The second step is to develop a set of questions for each of the four basic abilities. Persons with experience of the domain in question are then asked to rate the answer to each question, providing the assessment of the system's potential for resilient performance.

Paper III present a study whose purpose was to explore how concrete aspects of the nuclear power plant control room system can be connected to the four basic abilities of resilient performance. In particular, it shows how aspects contributing to safe operation from Study I (functions, structural elements, and characteristics) can be used to make this connection. Examples of this, based on statements from Study I, are shown in Figure 4.

One requirement when applying RAG is to define and describe the system to be assessed to be able to develop a suitable set of questions for each cornerstone. The results of Paper III can be used as a foundation when preparing RAG questions for a nuclear power plant control room system. Paper III showed how known aspects of control room system design contributing to safe operation can be viewed in the light of their contribution to resilient behaviour. The functions, characteristics, and structural elements themes have the potential to describe the nature of concrete aspects of control room system design and to connect them to the four basic resilient abilities respond, monitor, anticipate, and learn.

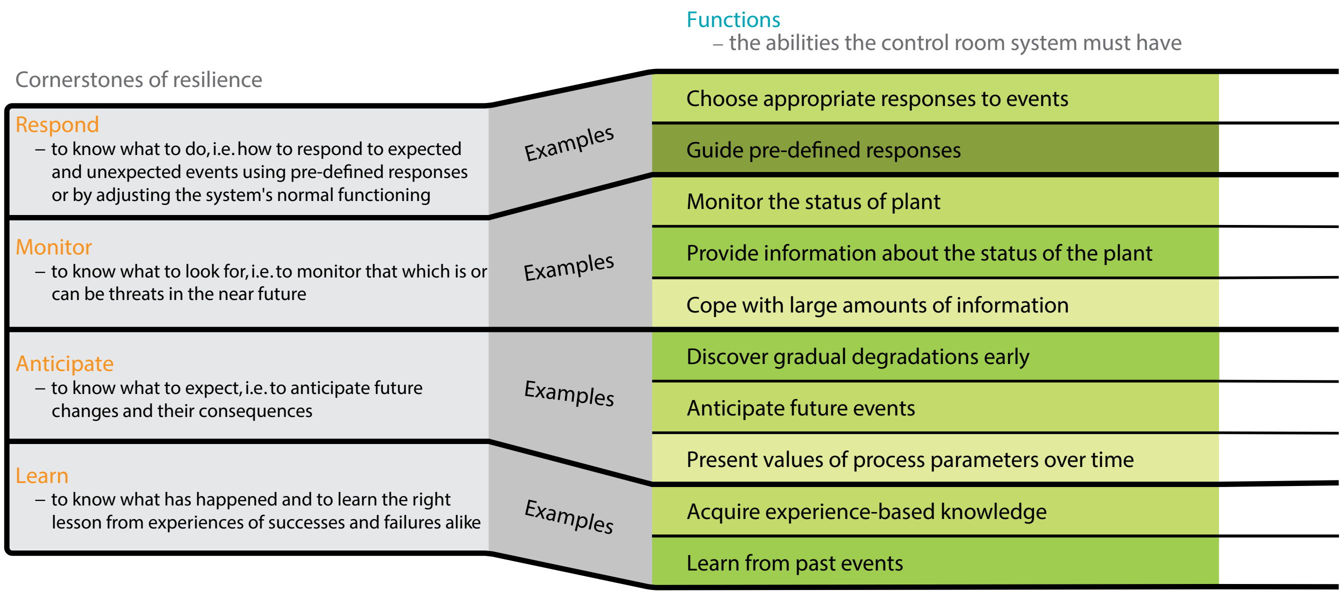
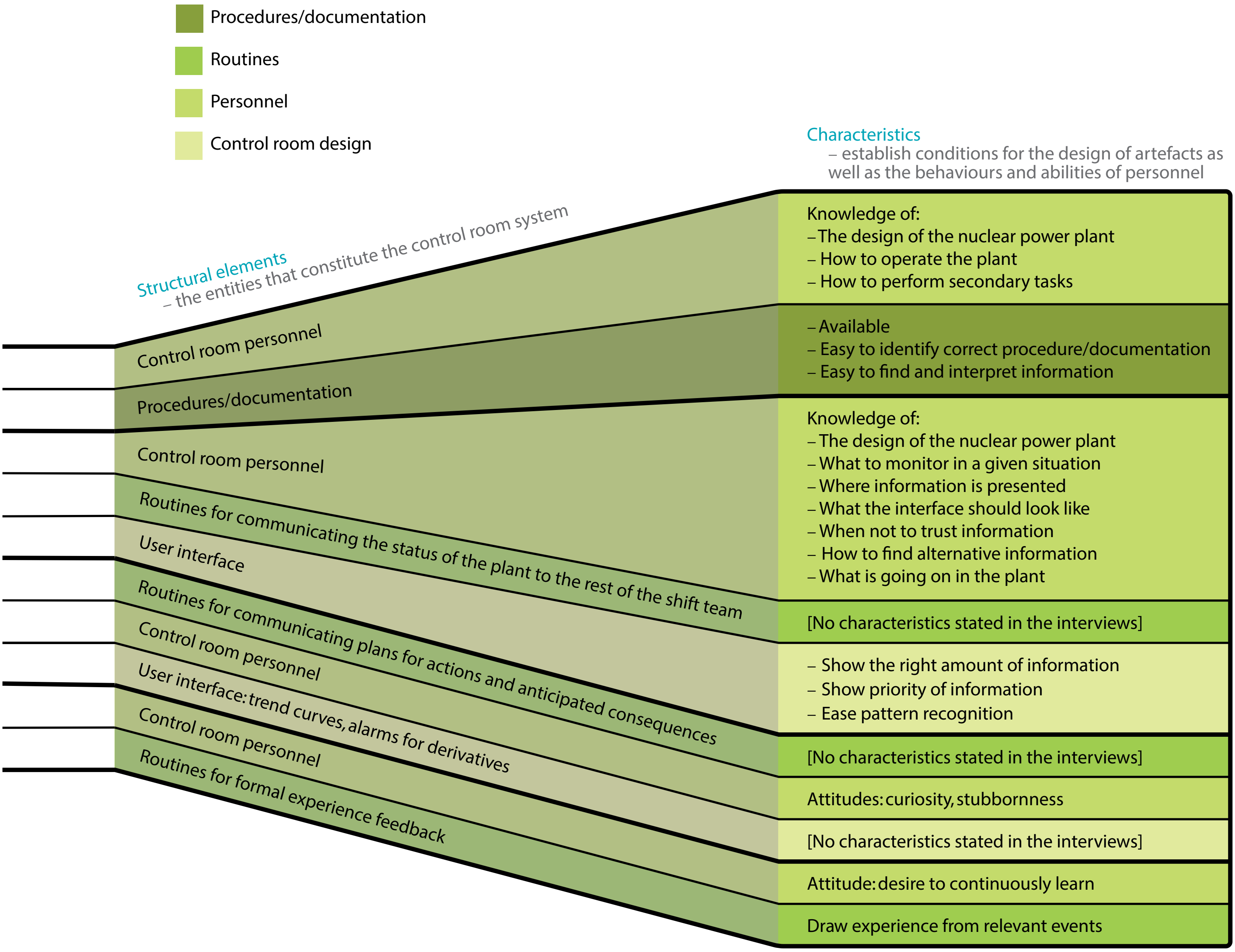


Figure 4: Examples showing how concrete aspects of the control room system (functions, structural elements, and characteristics) can be connected to the cornerstones of resilience.

5.6 CONCLUDING EVALUATION MEASURES

This section presents conclusions regarding which measures should be used to evaluate control room systems, based on the contents of sections 5.1 to 5.5. Specifically, which of the categories of evaluation measures presented in section 4.3 are suitable for nuclear power plant control room systems.

System and task performance of some kind were used as measures in several of the reviewed empirical control room evaluations (section 5.1), and they are supported by the frameworks of Savioja (2014) and Braarud and Rø Eitrheim (2013). The aspects contributing to safe operation proposed by the interviewees in Study I also suggest that system and task performance are relevant measures for nuclear power plant control rooms. In addition, task performance would be useful for assessing future nuclear power plant control room systems, according to Study II.



Teamwork and use of resources were also assessed in several of the reviewed evaluations in section 5.1, as well as being supported by Savioja (2014) and Braarud and Rø Eitrheim (2013). The communication and cooperation aspects of teamwork were heavily emphasised by the interviewees in Studies I and II. The interviewees also often regarded use of resources in terms of situation awareness and to some extent workload.

User experience was not measured in any of the reviewed empirical control room evaluations (section 5.1), it was mentioned but not emphasised in Study I, and it was not identified as an important measure category in Study II. The contents of the model by Braarud and Rø Eitrheim (2013) do not specifically suggest that it is important to evaluate user experience. However, measuring user experience is a very important part of Savioja's (2014) framework. She argues that the difficulty of analytically fully understanding the effect the tool will have on the activity makes utilisation of the operators' experiences important. Furthermore, she proposes that user experience measures are useful for evaluation of early design concepts. Less mature designs might be difficult to assess through the use of performance measures (such as time and errors), but the expert users' intuitive feeling of the tool's appropriateness may be useful in this phase.

The category identification of design discrepancies has a somewhat different focus that the other categories. It focuses on the design of the artefacts rather than the use and user of the system. Measures falling into the identification of design discrepancies category were utilised in four of the reviewed evaluations in section 5.1. Many aspects from Study I, especially those belonging to the characteristics theme, would also be suitable to assess using this category. In addition, Study II showed the category to be useful when evaluating future nuclear power plant control room systems. Neither Savioja (2014) nor Braarud and Rø Eitrheim (2013), however, include measures from this category in their frameworks. One concluding question is whether measures in this category are redundant. As was stated in section 4.3, an advantage of the identification of design discrepancies category is that it provides information regarding the nature of a problem, not only that there is a problem. As such, it serves as a very useful complement to other measures, especially in formative evaluations. This view is seemingly shared by the NUREG/CR-6393 and NUREG-0711 documents (United States Nuclear Regulatory Commission, 1997, 2012).The list of suggested measures in these documents focuses both on use and user (plant performance, task performance, situation awareness, and workload) as well as on the artefact (anthropometric and physiological factors), the latter measures belonging to the identification of design discrepancies category.

To conclude, system performance, task performance, teamwork, and use of resources are widely used and proposed categories of measures for control room system evaluations. System performance is a category of measures closely related to system goals. Assessing the system's performance can be seen as the lower level of acceptance – if this is not met, there is little point in assessing aspects from other categories to determine if the design is sufficient. Task performance, teamwork, and use of resources has the advantage of being more sensitive to variations in performance than measures in the system performance category. For example, experienced operators might be able to keep plant parameters within their limits, but discrepancies in their task performance can indicate where a less experienced operator would have problems handling the situation. The task performance, teamwork, and use of resources categories all assess unique aspects of performance variation and should all be included in control room system evaluations. The user experience category has the advantage of utilising the subjective experience of the operators as well as being less dependent on actual performance. The former adds a valuable perspective to the evaluation, and the latter makes the category useful for less mature design concepts. The last category, identification of design discrepancies, might not necessarily find unique problems that are impossible to find using the other categories. However, this category has the advantage of providing information more easily converted into design changes that need to be implemented. For formative evaluations, with the purpose of identifying ways to improve the design, this is a valuable characteristic.

As for the resilience perspective, it is a view of safety that has the potential to increase safety in ways that are unobtainable using the traditional view of safety. Having said that, the potential for resilient performance of a system is a systemic aspect, and is not something that is possible to contain in a single measure. Savioja (2014) describes how aspects of resilient performance are covered by her framework, and the Resilience Assessment Grid tool in combination with the results of Paper III offers another way to assess a nuclear power plant control room system from the resilience perspective.

The conclusion of chapter 5 is that all categories of evaluation measures presented in section 4.3 are relevant for evaluating nuclear power plant control room systems. In addition, adopting the resilience perspective in the evaluation has the potential to further prepare the nuclear power plant control room system for the unexpected.

DISCUSSION

The contents of this licentiate thesis have contributed to the purpose of the overall research project in that it has improved and further developed the knowledge of methods for evaluation of modified and newly designed control rooms. This contribution is shown here by presenting how the present work answers research question 1 and 2a.

Research question 1 (RQ1) is about finding the aspects that must be evaluated to assess the control room system's ability to support safe operation of the plant. The aspects identified are six categories of measures: system performance, task performance, teamwork, use of resources, user experience, and identification of design discrepancies. These different categories of measures each contribute differently to the assessment of the control room system. The identification of design discrepancies category differs from the others in that it may not necessarily identify unique problems. However, this category is useful in formative evaluations since it provides better information on how identified problems may be solved.

Another finding connected to RQ1 is application of the resilience engineering perspective to safety when evaluating control room systems. This is a valuable complement when assessing complex socio-technical systems where all future events cannot possibly be predicted. The framework of Savioja (2014) claims to consider the resilience engineering perspective through its focus on user experience and way of acting. However, it was developed for use in empirical control room system evaluations and not for analytical evaluations. An alternate path might be to explore analytical methods that evaluate the perspectives in Savioja's framework which examine resilience engineering. This would primarily encompass methods that evaluate user experience and task performance (the latter being the category of measures closest to 'way of acting'). The Resilience Assessment Grid (RAG) is an analytical method whose measures fall into different categories depending on the questions asked. The cornerstones of resilience consist of abilities desirable in the control room system, and using RAG is a way of determining if the system design supports them. In particular, using the result of Paper III as a foundation, RAG can be used to assess if a nuclear power plant control room system support these abilities. However, resilience engineering is still a relatively new field of research. The cornerstones of resilience encompass one set of abilities believed to be important for resilient behaviour, although other paths should also be sought.

This licentiate thesis also aimed at answering research question 2a (RQ2a) – to find the characteristics required of evaluation methods. One identified requirement is that evaluation methods for early evaluation must allow the use of more conceptual system representations. This requirement makes analytical evaluation methods that utilise indirect studies of use more suitable for early evaluation. An initial review of existing evaluation methods was performed, and it indicated that there are many methods suited for direct studies of use, but fewer analytical methods (see section 4.3). For some of the identified categories of measures no analytical evaluation methods were found. For user experience measures, Savioja et al. (2014) explain this lack of proven evaluation methods by the fact that user experience has most often been investigated in the context of consumer products and applications. As a result, the number of methods suitable for safety-critical work systems such as nuclear power plant control rooms is low. Apart from conducting a more thorough search for methods, it would be interesting to investigate further if empirical evaluation methods can be modified to suit indirect studies of use as well. A natural first step would be to explore if existing empirical methods can be used for imagined use rather than actual use. For example, walk- and talk-throughs are two approaches using this principle, making them suitable for more conceptual system representations.

A development process might look linear on paper, but is seldom strictly so in reality. However, design decisions higher up the abstraction hierarchy are usually finalised before a design decision at lower levels. Even though the detailed design of specific components might need to be discussed to be able to reach a conclusion on more overall functions, design decisions often need to be finalised from the higher abstraction levels and down. This means that the design of higher-level functions of the control room systems are normally decided before detailed interface design is developed, and can be evaluated earlier. Evaluation methods focused on evaluating functions rather than actual realisations of functions might thus be better suited for early evaluation. This is yet another topic that will be further investigated in future work within this doctoral project.

Many of the identified categories of measures are performance-related. An issue worth investigating further is whether or not performance measures are even possible to evaluate using indirect studies of work. Can imagined use or other approaches say something meaningful about measures closely related to performance? Human Reliability Analysis (HRA) methods are widely used techniques for predictively assessing task performance. Retrospective methods for assessing mental workload, such as SWAT, Subjective Workload Assessment Technique (Reid et al., 1981, Reid and Nygren, 1988), and SWORD, Subjective Workload Dominance Technique (Vidulich, 1989), exist in predictive versions as well; PRO-SWAT (Eggleston, 1984, Reid and Shingledecker, 1984) and PRO-SWORD (Vidulich et al., 1991). Analytical methods for evaluating task performance and use of resources evidently exist today, implying that methods for indirect studies of use might be possible for other performance-related measures as well.

Required evaluation method characteristics do not only stem from what they must be able to evaluate, but also from how useable they must be. In the study initiating this doctoral project Osvalder and Alm (2012) highlight the fact that methods need to be flexible and simple to use. Waterson et al. (2015) examined current sociotechnical methods and assessed their suitability from a theoretical and practical standpoint. One of the identified issues was that many methods "*proved to be difficult to use, time consuming and require a lot of training*" (Waterson et al., 2015, p. 7). This need for usability in human factors methods is in agreement with the author's own practical experience. To identify requirements on industrially viable human factors engineering methods, Andersson and Osvalder (2015) interviewed human factors engineers to investigate the extent to which existing human factors engineering methods match practitioners' needs in industrial contexts. Nine requirements or features were found for a human factors engineering method to be useful in practice, namely that the method should: be tweakable to fit the working context, be systematic, be inspiring and fun to use, be adaptable for use in multidisciplinary teams, work for varying levels of ambition, be transparent for all stakeholders, be explicit in how to use it, support measurability, and fit into the development process.

The scope of the present work has been limited to the nuclear power plant control room system. There may however be parts of the plant's process whose control is not centralised. Locally placed control interfaces might therefore exist in addition to the central control room. In recent years the Swedish Radiation Safety Authority has strived to widen the human factors focus for modifications in Swedish nuclear power plants, from only concerning the central control room to including control interfaces in other parts of the plant as well. Locally placed control interfaces in a nuclear power plant have the same overall purpose as the centralised control room system, namely safe operation. The main difference between the two lies in the locally placed control interface not being continuously manned. Furthermore, they differ in scope, both the scope of what is being controlled and also the scope of the physical and organisational structural elements that constitute them. These differences will affect the scope of the evaluation being undertaken, but not the categories of measures relevant to use. The only difference is the teamwork category. If a local control interface is always operated by a single operator, and if this operator does not need to communicate with other operators, then teamwork evaluation is not relevant. There are also, of course, control room systems that do not control a process in the same way as a nuclear power plant. One such example is the nuclear security control room, responsible for monitoring the nuclear power plant for malicious acts such as sabotage and unauthorised access, and coordinating responses to such events. For control rooms or locally placed control interfaces not controlling a process, the purpose might be different from that of the nuclear power plant control room system. The categories of measures will still be relevant, but the exact measures to be used in the evaluation must be operationalised from the purpose of the system in question.

The focus of the overall doctoral project was on methods for evaluating safe operation. Consequently the aim of Study I was to find aspects of the control room system that contribute to safe operation from a human factors perspective. When planning the study safe operation was seen as a performance and safety issue. Apart from the aspects contributing to safe operation, an additional conclusion of Paper I was that operator well-being should be regarded as a goal for control room systems in addition to safe operation. While a positive user experience certainly contributes to performance, operator well-being also has a value of its own. This is also an additional argument for using the user experience category of measures when evaluating control room systems. Another issue to discuss with regard to the goal for the control room system is to avoid financial damage, for example through damage to equipment. The definitions of the term safe operation proposed in section 3.2 (from a Safety-I and a Safety-II perspective) both emphasise the production of electricity. Production of electricity naturally requires undamaged and functioning equipment. While damage to equipment will involve financial costs in addition to the costs of lost production, the goal of the nuclear power plant control room system remains the same. Striving to uphold safe operation will mean avoiding damage to equipment regardless of the purpose.

The studies performed and included in the present work were interview studies. Study I sought aspects that contribute to safe operation and Study II explored future nuclear power plant control room trends and usability problems that these trends may promote. The statements given by the interviewees in Study I are subjective. Their knowledge and education may be grounded in objective studies of what makes the nuclear power plant control room better suited for safe operation. This foundation for their knowledge can however not be guaranteed. Their experience, on the other hand, is first-hand and should be regarded as a source of knowledge in its own right. In addition, the results of Study I do not contradict the measures used in the empirical control room system evaluations reviewed in section 5.1. Nor do they contradict the contents of the frameworks for selection of evaluation measures presented in section 5.2. The results of Study I are thus considered trustworthy. As for the results of Study II, the short-term control room system modification trends were based on the plant owners' actual plans for future changes and can thus be considered sound. The long-term trends were based on the interviewees' own predictions and may thus not be as reliable. However, the possible usability problems and required categories of measures connected to these long-term trends did not differ from the ones connected to the short-term trends. Therefore, this lesser reliability is of little importance.

The contents of this licentiate thesis have contributed to the purpose of the overall research project through contribution to fulfilment of the first and third goals. The identified categories of measures can be used in reviews of the plant owners' evaluations. The present work thus contributes to the first goal, to provide the Swedish Radiation Safety Authority with knowledge, methods and guidelines to support them in their supervisory and licencing role. The categories of measures identified as relevant for nuclear power plant control room systems also constitute useful knowledge for professionals in the domain of planning and executing evaluations. This knowledge, combined with the other required characteristics of methods identified here, will be able to further improve methods for early evaluation, providing professionals with tools useful in their evaluation work. The present work thus contribute to fulfilment of the third goal as well, to support and improve national competence in the domain and academia.

The second goal of the overall doctoral project is to modify existing and develop new methods, guidelines, and principles for evaluation of modified and newly designed control rooms for process control, with a focus on safe operation. This goal connects to research question 2b and c (RQ2b and RQ2c): are there suitable evaluation methods? If there are no suitable methods, how must existing evaluation methods be modified in order to be suitable? These questions summarise the future work to be done in this doctoral project. The present work has established the requirements, and existing methods fitting these requirements will be sought in the continuing work. If no suitable methods are found, these requirements will be used to develop new methods. These methods will then be tested, preferably in real-world projects, to assess whether or not they are able to assess control room systems' ability to support safe operation in early phases of the development process.

CONCLUSIONS

This licentiate thesis has shown that six different categories of measures are relevant in the evaluation of safe operation in nuclear power plant control room systems: system performance, task performance, teamwork, use of resources, user experience, and identification of design discrepancies. The use of a combination of measures from the different categories is necessary in order to fully assess a complex socio-technical system such as the control room system.

Apart from using a combination of measures, evaluation of the control room system should also consider the resilience engineering perspective. Simply anticipating future risks and taking precautions to avoid them will not be enough to pursue safe operation for a complex socio-technical system such as a control room system. Applying resilience engineering in the design of control room systems will make them better suited to handle unanticipated events. In this, evaluation from the resilience engineering perspective is important so as to assess whether or not the system's ability for resilient behaviour is sufficient. Suitable methods for early assessment of the capacity for resilient behaviour in control room systems is a topic that needs further exploration. The Resilience Assessment Grid is one such method that could be further investigated, together with methods that evaluate user experience and task performance.

Apart from investigating categories of measures, this licentiate thesis has also explored method characteristics required for early evaluation of control room systems. The usability of methods was one such characteristic; if methods are to be utilised and have an actual impact in industry, practitioners must find them useful in practice. The other method characteristic for early evaluations identified was that system representations in these phases are more conceptual, and that using these representations to perform tasks differs in some aspects from use of the final system. Empirical methods that directly study user interaction with the control room system are therefore less suitable for evaluations early in the development process. Analytical methods that study use indirectly are a better choice since they allow the use of more conceptual system representations. An initial review of existing human factors evaluation methods indicates that while empirical methods are abundant for all categories of measures, analytical methods are not as common. For two of the identified categories of measures, system performance and teamwork, no analytical methods were found in this initial search. Methods were identified for the other categories of measures, but not in large numbers. Further investigating and addressing this gap for control room systems will constitute the continuing work in this doctoral project.

REFERENCES

Andersson, J., Osvalder, A.-L. (2015) *Method characteristics for viable human factors engineering practice*. Göteborg: Department of Product and production development Chalmers University of Technology. (Research series from Chalmers University of Technology, Department of Product and Production Development: 95)

Boring, R. L. (2014) Formative evaluation for optimal upgrades in nuclear power plant control rooms. *In PSAM 2014 - Probabilistic Safety Assessment and Management*; 22-27 June 2014, Honolulu, Hawaii.

Braarud, P. Ø., Rø Eitrheim, M. H. (2013) *A Measurement Framework for Human Factors Integrated System Validation of NPP Control Rooms*. Halden: Institutt for Energiteknikk. (OECD Halden Reactor project: HWR-1063)

Carvalho, P. V. R., dos Santos, I. L., Gomes, J. O., Borges, M. R. S., Guerlain, S. (2008) Human factors approach for evaluation and redesign of human-system interfaces of a nuclear power plant simulator. *Displays*, vol. 29, no. 3, pp. 273-284.

Chuang, C. F., Chou, H. P. (2008) Design Development and Implementation of the Human-System Interface for Lungmen Nuclear Project. *IEEE Transactions on Nuclear Science*, vol. 55, no. 5, pp. 2654-2661.

Eggleston, R. G. (1984) A comparison of projected and measured workload ratings using the subjective workload assessment technique (SWAT). In *Proceedings of the IEEE 1984 National Aerospace and Electronics Conference*; 21-25 May 1984, Dayton, Ohio.

Engel, A. (2010) *Verification, validation, and testing of engineered systems*. [E-book] Wiley & Sons.

Evaluation. (2015a) In *Oxford Dictionaries*. http://oxforddictionaries.com (9 October 2015).

Gatto, L. B. S., Mól, A. C. A., Luquetti dos Santos, I. J. A., Jorge, C. A. F., Legey, A. P. (2013) Virtual simulation of a nuclear power plant's control room as a tool for ergonomic evaluation. *Progress in Nuclear Energy*, vol. 64, pp. 8-15.

Gegner, L., Runonen, M. (2012) For what it is worth: Anticipated experience evaluation. In *Proceedings of the 8th International Design and Emotion Conference*; 11-14 September 2012, London.

Ha, J. S., Seong, P. H., Lee, M. S., Hong, J. H. (2007) Development of human performance measures for human factors validation in the advanced MCR of APR-1400. *IEEE Transactions on Nuclear Science*, vol. 54, no. 6, pp. 2687-2700.

Hale, A., Kirwan, B., Kjellén, U. (2007) Safe by design: where are we now? *Safety*

Science, vol. 45, no. 1–2, pp. 305-327.

Hammersley, M., Atkinson, P. (1983) *Ethnography: Principles in Practice*. London: Travistock Publications.

Hollnagel, E. (1985) *A survey of man–machine system evaluation methods*. Halden: Institutt for Energiteknikk. (OECD Halden Reactor Project: HWR-148)

Hollnagel, E. (2011a) Epilogue: RAG - The Resilience Analysis Grid. In *Resilience engineering in practice : A guidebook*, eds. E. Hollnagel, J. Pariès, D. Woods and J. Wreathall, pp. 275-296. Surrey: Ashgate Publishing Limited.

Hollnagel, E. (2011b) Prologue: The Scope of Resilience Engineering. In *Resilience Engineering in Practice : A Guidebook*, eds. E. Hollnagel, J. Pariès, D. Woods and J. Wreathall, pp. xxix-xxxix. Surrey: Ashgate Publishing Limited.

Hollnagel, E. (2013) A tale of two safeties. *Nuclear Safety and Simulation*, vol. 4, no. 1, pp. 1-9.

Howitt, D. (2013) *Introduction to qualitative methods in psychology*. Harlow: Pearson.

Hwang, S.-L., Liang, S.-F. M., Liu, T.-Y. Y., Yang, Y.-J., Chen, P.-Y., Chuang, C.-F. (2009) Evaluation of human factors in interface design in main control rooms. *Nuclear Engineering and Design*, vol. 239, no. 12, pp. 3069-3075.

Institute of Electrical and Electronics Engineers (1999) *IEEE Std 845-1999 Guide for the Evaluation of Human-System Performance in Nuclear Power Generating Stations*. New York: The Institute of Electrical and Electronics Engineers.

International Atomic Energy Agency (2007) *IAEA Safety Glossary Terminology Used in Nuclear Safety and Radiation Protection - 2007 Edition*. Vienna: International Atomic Energy Agency.

International Ergonomics Association. (2016) *Definition and Domains of Ergonomics*. http://www.iea.cc/whats/ (9 January 2016).

International Nuclear Safety Advisory Group (1999) *Basic Safety Principles for Nuclear Power Plants*. Vienna: International Atomic Energy Agency. (INSAG: 75-INSAG-3)

International Standard Organisation (2000) *ISO 11064-1:2000 Ergonomic design of control centres – Part 1: Principles for the design of control centres*. Geneva: International Standard Organisation.

International Standard Organisation (2004) *ISO 6385:2004 Ergonomic principles in the design of work systems*. Geneva: International Standard Organisation.

International Standard Organisation (2006) *ISO 11064-7:2006 Ergonomic design of control centres – Part 7: Principles for the evaluation of control centres*. Geneva: International Standard Organisation.

Kaptelinin, V., Nardi, B. (2012) A*ctivity theory in HCI: Fundamentals and Reflections. Synthesis Lectures on Human-Centered Informatics*. Morgan & Claypool.

Le Blanc, K., Gertman, D., Boring, R. (2010) Review of methods related to assessing human performance in nuclear power plant control room simulations.

In *7th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies 2010*; 7-11 November 2010, Las Vegas, Nevada. pp. 411-422.

Le Blanc, K. L., Oxstrand, J. H. (2013) Computer-based procedures for nuclear power plant field workers: Preliminary results from two evaluation studies. In *Proceedings of the Human Factors and Ergonomics Society*; 30 September-4 October 2013, San Diego, CA. pp. 1722-1726.

Li, Z., Reed, J. I., Fuld, R. B. (2012) Preparations for Integrated System Validation of the AP1000 Main Control Room and human system interface design. In *8th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies 2012, NPIC and HMIT 2012: Enabling the Future of Nuclear Energy*; 22-26 July 2012, San Diego, CA. pp. 498-504.

Liu, Y., Zhanq, J., Jia, M., Guo, A., Shi, J., Sun, Y. (2012) Human factors engineering verification and validation of human system interface design of the main control room in CPR1000 nuclear power plants. In *8th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies 2012, NPIC and HMIT 2012: Enabling the Future of Nuclear Energy*; 22-26 July 2012, San Diego, CA. pp. 516-524.

Luquetti dos Santos, I. J. A., dos Santos Grecco, C. H., Abreu Mol, A. C., Rodrigues Carvalho, P. V. (2009) The use of questionnaire and virtual reality in the verification of the human factors issues in the design of nuclear control desk. *International Journal of Industrial Ergonomics*, vol. 39, no. 1, pp. 159-166.

Meister, D. (2001) Basic premises and principles of human factors measurement. *Theoretical Issues in Ergonomics Science*, vol. 2, no. 1, pp. 1-22.

Method. (2015b) In *Oxford Dictionaries*. http://oxforddictionaries.com (3 November 2015).

Nemeth, C. P., Herrera, I. (2015) Building change: Resilience Engineering after ten years. *Reliability Engineering & System Safety*, vol. 141, pp. 1-4.

Norros, L., Nuutinen, M. (2005) Performance-based usability evaluation of a safety information and alarm system. *International Journal of Human-Computer Studies*, vol. 63, no. 3, pp. 328-361.

Noyes, J. (2004) The human factors toolkit. In *Human Factors for Engineers*, eds. C. Sandom and R. S. Harvey, pp. 57-79. London: The Institution of Electrical Engineers.

OECD/NEA Committee on Safety of Nuclear Installations (2005) *Safety of Modifications at Nuclear Power Plants: The Role of Minor Modifications and Human and Organisational Factors*. Paris: OECD Publications. NEA/CSNI/R(2005)10)

Osvalder, A.-L., Alm, H. (2012) *Methods for Evaluation of Safety in Complex Process Control. Critical review of processes and methods used in the nuclear power domain and suggestions for research activities*. Solna: Swedish Radiation Safety Authority.

Osvalder, A.-L., Rose, L., Karlsson, S. (2009) Methods. In *Work and technology on human terms*, eds. M. Bohgard, S. Karlsson, E. Lovén, L.-Å. Mikaelsson, L. Mårtensson, A.-L. Osvalder, L. Rose and P. Ulfvengren, pp. 436-566. Stockholm: Prevent.

Pahl, G., Beitz, W. (1996) *Engineering Design - A Systematic Approach*. London: Springer.

Papin, B. (2002) Integration of Human Factors Requirements in the Design of Future Plants. In *Proceedings of the Enlarged Halden Program Group Meeting*; 8-13 September 2002, Storefjell.

Perrow, C. (1999) *Normal Accidents: Living with High-Risk Technologies*. Princeton, NJ: Princeton University Press.

Reid, G. B., Nygren, T. E. (1988) The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Human Mental Workload*, eds. P. S. Hancock and N. Meshkati, pp. 185-218. Amsterdam: Elsevier.

Reid, G. B., Shingledecker, C. A. (1984) A projective application of the subjective workload assessment technique. In *Proceedings of the IEEE 1984 National Aerospace and Electronics Conference*; 21-25 May 1984, Dayton, OH.

Reid, G. B., Shingledecker, C. A., Eggemeier, F. T. (1981) Application of Conjoint Measurement to Workload Scale Development. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; October 1981, Rochester, NY. pp. 522-526.

Sacharin, V., Schlegel, K., Scherer, K. R. (2012) *Geneva Emotion Wheel rating study*. Geneva: University of Geneva Swiss Center for Affective Sciences.

Savioja, P. (2014) *Evaluating systems usability in complex work - Development of a systemic usability concept to benefit control room design*. Espoo; Aalto University School of Science. (Dissertation).

Savioja, P., Liinasuo, M., Koskinen, H. (2014) User experience: does it matter in complex systems? *Cognition, Technology and Work*, vol. 16, no. 4, pp. 429–449.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., Jenkins, D. P. (2005) *Human Factors Methods : A practical Guide for Engineering and design*. Surrey: Ashgate Publishing Limited.

Svensson, P.-G., Starrin, B. (1996) *Kvalitativa studier i teori och praktik*. Lund: Studentlitteratur.

Tschirner, S. (2015) *The GMOC Model: Supporting Development of Systems for Human Control*. Uppsala; Uppsala University. (Dissertation).

Ullman, D. G. (1997) *The Mechanical Design Process*. New York: McGraw-Hill.

Ulrich, K. T., Eppinger, S. D. (2003) *Product Design and Development*. New York: McGraw-Hill.

United States Nuclear Regulatory Commission (1997) *NUREG/CR-6393 Integrated System Validation: Methodology and Review Criteria*. Washington, DC: United States Nuclear Regulatory Commission.

United States Nuclear Regulatory Commission (2002) *NUREG-0700 Human-System Interface Design Review Guidelines Revision 2*. Washington, DC: United States Nuclear Regulatory Commission.

United States Nuclear Regulatory Commission (2012) *NUREG-0711 Human Factors Engineering Program Review Model Revision 3*. Washington, DC: United States Nuclear Regulatory Commission.

Waterson, P., Robertson, M. M., Cooke, N. J., Militello, L., Roth, E., Stanton, N. A. (2015) Defining the methodological challenges and opportunities for an effective science of sociotechnical systems and safety. *Ergonomics*, vol. 58, no. 4, pp. 1-35.

Vidulich, M. A. (1989) The use of judgement matrices in subjective workload assessment: The subjective workload dominance (SWORD) technique. In *Proceedings of the Human Factors Society 33rd Annual Meeting*; October 1989, Santa Monica, California. pp. 1406-1410.

Vidulich, M. A., Ward, G. F., Schueren, J. (1991) Using the Subjective Workload Dominance (SWORD) technique for projective workload assessment. *Human Factors*, vol. 33, no. 6, pp. 677-691.