

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Advancing Formative Control Room System Evaluation

Decision support for human factors evaluation planning and method development

EVA SIMONSEN



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Industrial and Materials Science
Division Design & Human Factors
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

Advancing formative control room system evaluation
Decision support for human factors evaluation planning and method development

Eva Simonsen

ISBN: 978-91-7597-793-5

© Eva Simonsen, 2018

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4474
ISSN 0346-718X

Published and distributed by
Department of Industrial and Materials Science
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 2018

ABSTRACT

The design of the nuclear power plant (NPP) control room system affects the operation of the plant it controls, as well as the well-being of its operators. One important activity in control room system development is Human Factors (HF) evaluation. Previous research indicates that HF evaluation practice for NPP control room systems can be improved. For example, there is a need for methods that are flexible and simple to use. In order to advance evaluation practices as part of the development process the purpose of this thesis was to increase understanding of HF evaluation of NPP control room systems. The main object of study was the evaluation activity. The first research question concerned the aspects that need to be assessed so as to be able to evaluate the control room system's ability to fulfil its intended purpose. The second research question asked if, and how, HF evaluation can better support control room system development. The methods used were two literature studies and empirical studies in the form of an interview study, three case studies, and three focus groups.

In response to the first research question, the interview study investigated those aspects of the NPP control room system that contribute to safe operation. In the first literature study these aspects were used together with aspects found in other studies to identify categories of measures relevant for assessing NPP control room systems. The identified categories – *system performance, task performance, use of resources, user experience, and identification of design discrepancies* – complement each other and should all be included in control room system evaluation during the course of the development process. In response to the second research question, the second literature study identified a gap in today's evaluation practice and the research efforts focused on formative evaluation of more general (higher-level) design decisions, preferably undertaken early in the development process. A combination of two methods, *heuristic evaluation* and *scenario-based talkthrough*, was used in the case studies and focus groups to explore the evaluation activity in practice. This method combination was found to be useful for formative assessment of higher-level design decisions in NPP control room systems. In addition, HF specialists from other domains who participated in the focus groups believed that the method combination would be useful outside the nuclear power domain too. A description of the method combination is included in the thesis to provide concrete guidance for HF practitioners. The experiences from the case studies were also used to identify *guidelines for developing HF evaluation methods that are useful in practice*.

From the knowledge gained through exploration of the research questions five perspectives that provide decision support in HF evaluation planning and method development emerged: 1) the *purpose* of the evaluation activity, 2) the *object* to be evaluated, 3) the *tactic* used in the evaluation activity, 4) the *evaluation* procedure, and 5) the *use* of the evaluation method. Individual results from the studies, such as the categories of measures and guidelines for developing methods that are useful in practice, can be used as more detailed support within these perspectives.

Keywords: *control room system, design, formative evaluation methods, human factors, nuclear power*

ACKNOWLEDGEMENTS

I would like to thank my examiner and head supervisor Professor Anna-Lisa Osvalder for her guidance, input, and confidence in me these past few years. Had you not alerted me to the existence of this PhD position in the first place, I might have missed the opportunity to dig deeper into a subject that is so much a passion of mine. I am very grateful to my co-supervisor, Dr Lars-Ola Bligård, for generously giving so much of your time to help me, and for our discussions that made my thoughts clearer. I would also like to thank my other co-supervisor, Dr Andreas Kjellin at the Swedish Radiation Safety Authority, as well as my co-supervisor until licentiate, Professor Håkan Alm (formerly of the Luleå University of Technology), for valuable comments and new perspectives.

The work presented in this thesis was funded by the Swedish Radiation Safety Authority, and I am grateful to my contact person there, Dr Yvonne Johansson, for your support and trust in my ideas.

The members of my reference group have varied during the years of my research project, but the following persons have all taken part at some point: Dr Jonas Andersson (RISE Viktoria), Agneta Bengtsson (Oskarshamns Kraftgrupp), Per Øivind Braarud (Institute for Energy Technology), Johan Holgersson (Ringhals), Jari Laarni (VTT Technical Research Centre of Finland), Professor Emerita Lena Mårtensson (Royal Institute of Technology), Maren Rø Eitheim (Institute for Energy Technology), Stefan Sördal (Swedish Radiation Safety Authority), and Associate Professor Clemens Weikert (formerly of Lund University). I thank you for sharing your experience and knowledge.

My time at Chalmers has been inspiring and fun, not least thanks to my past and present colleagues at the division of Design & Human Factors. In particular, I would like to thank Maral Babapour, Ingrid Pettersson, Dr Isabel Ordóñez Pizarro, Sara Renström, Dr Anneli Selvefors, and Dr Helena Strömberg for being my companions on this journey – I have learned so much from you. I am especially grateful to Dr Helena Strömberg for your supervision disguised as small-talk, and to Maral Babapour for being such an encouraging sounding-board and for all the emotional support. I would also like to say thank you to past and present colleagues at the Human Factors division at Vattenfall for being so generous with your knowledge and time.

I am profoundly grateful to the participants in my studies for sharing your experiences with me. Without you there would be no thesis. I also thank my pre-dissertation seminar leader, Dr Gesa Praetorius at Linnæus University, for our challenging discussions that made me think along new lines. I would also like to thank Ilya Meyer for proofreading my texts.

My research would not have been possible without support from people outside my work. I am so grateful to my friends, Cecilia and Johanna especially, for the distractions away from work, and also for listening to me talk passionately about my work. Lastly, I want give the warmest of thanks to my parents and my brother – Barbro, Lars-Erik and Gustav – who have done so much to shape me into who I am, for your love and support. And to my husband Lars, to whom I am more grateful than I can ever put into words. Thank you for making me food and for making me laugh, and for always asking how you can help.

APPENDED PUBLICATIONS

This thesis is based on the work contained in the following papers:

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, Nevada.

Simonsen planned the study with feedback from Osvalder. Simonsen collected and analysed the data. Simonsen wrote the paper with feedback from Osvalder.

PAPER II

Simonsen, E., Osvalder, A.-L. (2018) Categories of measures to guide choice of human factors methods for nuclear power plant control room evaluation. *Safety Science*, vol. 102, pp. 101-109.

Simonsen planned the study, collected the data, and analysed the data. Simonsen wrote the paper with feedback from Osvalder.

PAPER III

Simonsen, E. (2017) A comparison of human factors evaluation approaches for nuclear power plant control room assessment and their relation to levels of design decision specificity. In *Nordic Ergonomic Society 2017 "Joy at work" Conference Proceedings*; 20-23 August 2017, Lund. pp. 405-414.

Simonsen planned the study, collected the data, analysed the data, and wrote the paper.

PAPER IV

Simonsen, E., Osvalder, A.-L. (submitted 2017) Human factors methods for early evaluation of control room systems – guidelines for use in practice. *Submitted 2017-12-27 to Theoretical Issues in Ergonomics Science, under review.*

Simonsen planned the study collected the data, and analysed the data. Simonsen wrote the paper with feedback from Osvalder.

PAPER V

Simonsen, E., Bligård, L.-O., Osvalder, A.-L. (submitted 2018) Feasibility of methods for early formative control room system evaluation. *Submitted 2018-08-31 to International Journal of Industrial Ergonomics.*

Simonsen planned the study, collected the data, and analysed the data. Simonsen wrote the paper with feedback from Bligård and Osvalder.

ABBREVIATIONS

HF	Human factors
HSI	Human-system interface
IAEA	International Atomic Energy Agency
IEA	International Ergonomics Association
IEC	International Electrotechnical Commission
INCOSE	International Council on Systems Engineering
INSAG	International Nuclear Safety Advisory Group
ISO	International Standard Organisation
NPP	Nuclear power plant
OECD NEA	Organisation for Economic Co-operation and Development Nuclear Energy Agency
US NRC	United States Nuclear Regulatory Commission

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 Background	1
1.2 Purpose and research questions	6
1.3 Reading instructions	7
2. FRAME OF REFERENCE	9
2.1 The control room system as a socio-technical system	9
2.2 Nuclear safety and safe operation	10
2.3 Human factors requirements in nuclear power	11
2.4 Human factors and design	12
2.5 Evaluation in design	13
2.6 The development process	15
2.7 Human factors methods	21
3. RESEARCH APPROACH	23
3.1 Research interests and worldview	23
3.2 Research design	26
4. RESULTS – SUMMARY OF STUDIES	29
4.1 Study A (Paper I)	29
4.2 Study B (Paper II)	30
4.3 Study C (Paper III)	32
4.4 Study D (Papers IV and V)	33
4.5 Study E	36
5. ANALYSIS	41
5.1 RQ1 - What to evaluate?	41
5.2 RQ2 - How to support control room system development?	43
5.3 Perspectives to guide evaluation	46
6. DISCUSSION	55
6.1 Advancing evaluation practices	55
6.2 Reflections on the research approach	56
6.3 Evaluation of systems with unpredictable behaviour	57
6.4 Evaluation of safety-critical systems	59
6.5 Evaluation of socio-technical systems	60
6.6 Practical implications	61
6.7 Relation to proposed evaluation approaches	62
6.8 Developing human factors as a design discipline	65

7. FURTHER WORK	67
8. CONCLUSIONS	69
REFERENCES	71
APPENDIX A	77

INTRODUCTION

1.1 BACKGROUND

This thesis concerns human factors evaluation of nuclear power plant control room systems. This section describes what a nuclear power plant control room system is and presents the problem that was the starting point of the work presented in this thesis.

1.1.1 The nuclear power plant control room system

A control room, according to the International Standard Organisation [ISO] (2000, p. 5), is a “*core functional entity, and its associated physical structure, where operators are stationed to carry out centralized control, monitoring and administrative responsibilities*”. In a standard for the design of nuclear power plant control rooms the International Electrotechnical Commission [IEC] (2009, p. 10) defines a control room system as “*an integration of the human-machine interface, the control room staff, operating procedures, training programme, and associated facilities or equipment which together sustain the proper functioning of the control room*”. The performance of the control room system, in turn, affects the power plant’s ability to fulfil its operational goal. IEC (2009) provides an overview of the control room system and its relation to other parts of the nuclear power plant and the goals of the plant (Figure 1).

In this thesis, 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. This kind of systems view emphasises 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. The purpose of the nuclear power plant control room system is primarily to support safe operation of the plant.

However, supporting safe operation is not the sole purpose of the control room system. An additional purpose is to support operator well-being. The control room system is the work environment for its operators, and the control room system design impacts their well-being. Savioja et al. (2014) argued this for the nuclear power domain, proposing that tools (such as operator interfaces) play a role in how

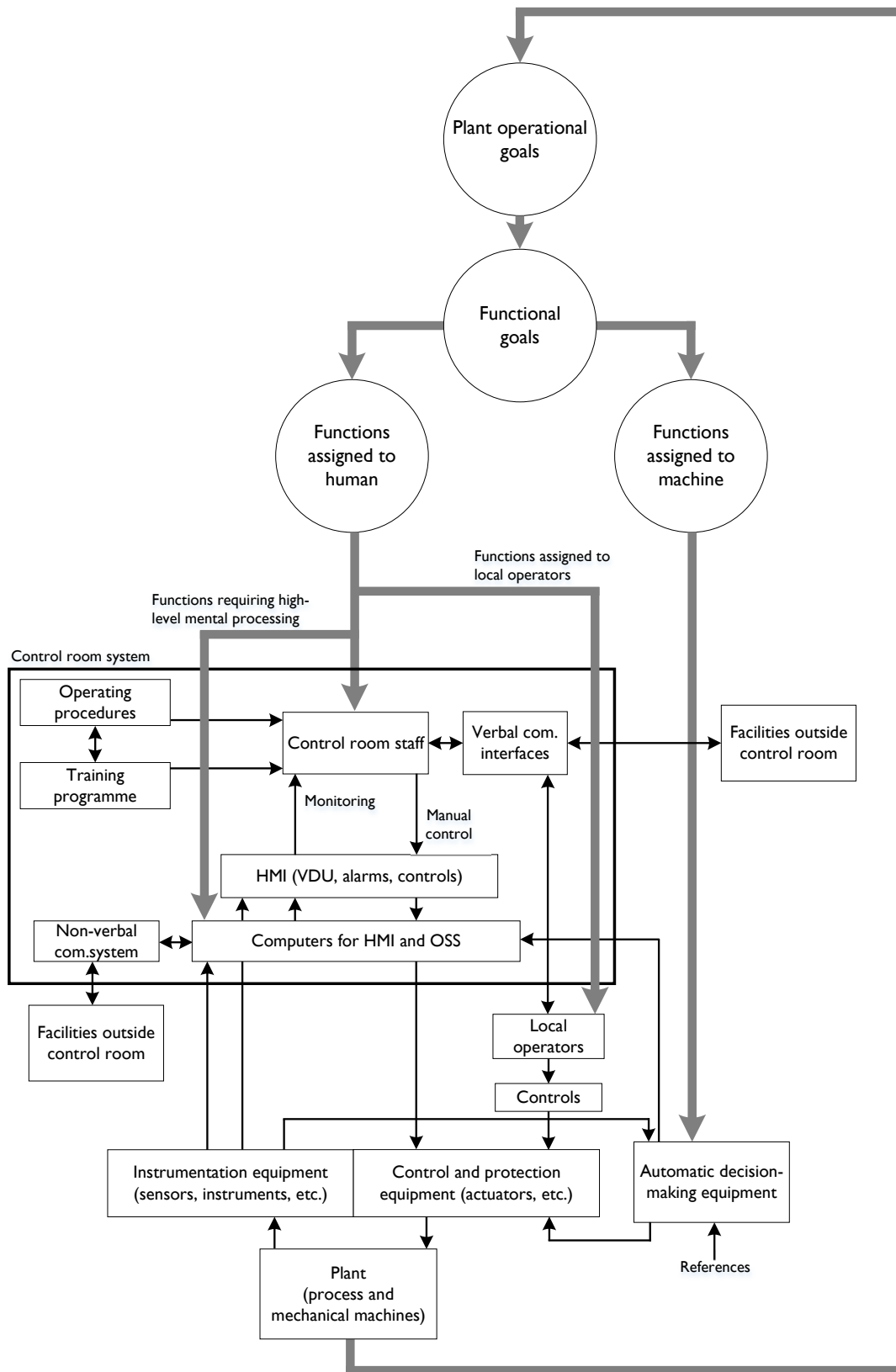


Figure 1: Overview of the nuclear power plant control room system, its relation to other parts of the plant, as well as to the plant's operational goals. Note that the organisational elements included in the definition of a control room system used in this thesis are not clearly illustrated in this overview. The present thesis also includes local operators and the controls they use in the control room system. Abbreviations: HMI – Human-Machine Interface, OSS – Operator Support System, VDU – Video Display Unit. Adapted from IEC (2009).

satisfying, exciting, and meaningful the work activity is to the worker. Operator well-being should thus not only be seen as a way to increase performance, but also as something that has a value of its own.

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 such 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 too. A control room system design must thus include both the design of operator interfaces and the placement of these and other functions in the control room space.

The nuclear power plant control room system is a place of work for professional users, who are specifically trained for their tasks and who have in-depth knowledge of the system they are to control. The control room system is operated by a team of operators who work in shifts to allow continuous operation. Responsibilities are divided among the operators in the shift team, creating different roles. In Swedish nuclear power plants these are typically shift supervisor, reactor operator, turbine operator, and local operators (the latter are part of the shift team but do not have their primary place of work in the control room). Depending on the reactor type, an assistant reactor operator or an electrical operator is also included in the shift team.

The nuclear power plant operators' work in normal operation is typically calm and can be carried out according to predefined routines. Routines exist for undesired events as well, but situations where the operators have to handle an unfamiliar situation without the support of routines will also occur. Procedures, the written and formalised account of routines, are used to guide operations in the control room. Procedures play a very important part in the operation of nuclear power plants and are required by the Swedish Radiation Safety Authority (2008a). Traditionally they are presented on paper, but in recent years computer-based procedures have been developed as well.

The behaviour of nuclear power plant control room systems may be characterised in many ways, but one characteristic that is of importance for the content of this thesis is the difficulty in fully predicting how the control room system will behave. This characteristic is the result of the nature of both the control room system itself and its environment. A nuclear power plant control room system includes human agents (the operators), an element whose behaviour cannot be fully predicted beforehand. Some aspects of operator behaviour are more predictable due to common human cognitive abilities, but operator behaviour in the control room will vary due to factors such as experience, training and personality. The control room system also interacts with an environment that is not fully predictable, such as the weather or other human agents (for example maintenance personnel). This difficulty in predicting the form and timing of input from the environment makes

it difficult to predict the control room system's behaviour. Another element in the environment of the nuclear power plant control room system characterised by unpredictable behaviour is the plant itself. Nuclear power plants are typically described as complex systems. Perrow (1999) described nuclear power plants as systems with highly complex interactions, meaning interactions that are unfamiliar, unplanned, or unexpected as well as invisible or not immediately comprehensible. Another example is Flach (2012), who defined a system as complex if its future is uncertain, and noted nuclear power plant as systems of medium complexity. Even if there are different opinions about exactly *how* complex nuclear power plants are, if their behaviours are impossible or just difficult to predict, it can be stated that the behaviours of nuclear power plants are not simple to predict. Thus the difficulty in predicting the behaviour of the nuclear power plant control room system is also due to the fact that it is tasked with controlling another system whose behaviour is difficult to fully foresee.

1.1.2 Problem description

Nuclear power is a safety-critical domain, and successful performance is important from both a safety and an economic perspective. The nuclear power plant accident at the Three Mile Island power plant in 1979 taught the nuclear power domain that attention to human factors in the design of the control room system was important (Vicente, 2004). The International Ergonomics Association [IEA] (2018) defines human factors (or ergonomics, but human factors is the term that will be used in this thesis) as “*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*”.

Much work has been done since the Three Mile Island accident to improve nuclear power plant control room systems from a human factors point of view, and its importance is also reflected in requirements and recommendations (this is elaborated in Section 2.3). Performance, and to be specific, safety, has been and still is the primary rationale for human factors work within the nuclear power domain. However, as is emphasised in the definition of human factors given by IEA (2018) and in Section 1.1.1, operator well-being is also an important rationale.

Swedish nuclear power plants were built in a period from the mid-1970s to the mid-'80s. Maintenance and modernisation demands have led to the initiation of a number of plant development projects. Either directly or indirectly, this has led to changes in the plants' control room systems as well. The modification of control room systems creates a need to evaluate whether the changed design continues to support safety, productivity and operator well-being. Against this background, the Swedish Radiation Safety Authority initiated a study (Osvalder and Alm, 2012). The purpose was to study and critically review human factors methods and

procedures used today to evaluate changes in control rooms and their possible impact on safety, productivity and operator well-being, and also to discuss the need for modified or new methods.

The study by Osvelder 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 from a human factors point of view. 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. Osvelder 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 need to further develop control room evaluation has been identified by other researchers too. Boring et al. (2015) pointed out that regulations for nuclear power plant control room modernisation put an emphasis on late-stage evaluation. Even though this is only natural from a regulatory perspective, Boring et al. (2015) noted that this may be interpreted by system designers to mean this is the only required or, indeed, preferred type of evaluation. Laarni et al. (2014) described challenges of human factors verification and validation in projects that are realised in multiple stages, over several years, and are closely linked to modernisation of the instrumentation and control of the plant. While the presented needs for evaluation have been discovered in the context of undertaking modernisations of nuclear power plants, evaluation is likewise needed when building new nuclear power plants.

The main object of study in this thesis was the evaluation activity. The term 'activity' is used here to denote something that is done with a certain intent, in this case evaluation. Another term that has been used in the thesis is 'practice'. This term is used when it is important to differentiate between ideas about what should be done and what is actually being done in the real world. 'evaluation practice' is thus how the evaluation activity is actually carried out in the real world, as opposed to ideas about how it should be undertaken.

In a development project, if an evaluation is meant to answer how well a proposed control room system design will be able to fulfil a certain purpose, then that question must be operationalised into parts that it is possible to assess. This is a key concept in evaluation, breaking down a large question into smaller underlying questions that can be studied with greater ease, and answering the larger question through the aggregation of the answers to the underlying questions. Finding underlying questions that can be answered may be expressed in terms of finding a suitable way to view the object of interest. An important angle when studying the evaluation activity in this thesis is thus how the design of the control room system and its impact are viewed by the evaluation activity.

The evaluation activity studied in this thesis is a part of the development process, and is given its purpose through that relation. Evaluation is not undertaken for its own sake, the evaluation activity should support control room system development. Further development of evaluation practices should thus be done in a way that improves the impact that the evaluation has on development.

1.2 PURPOSE AND RESEARCH QUESTIONS

To summarise the background of this thesis, the design of the control room system affects the operation of the plant it controls as well as the well-being of its operators, and human factors evaluation is an important activity in control room system development. As indicated by previous research, the practice of human factors evaluation of control room systems in the nuclear power domain can be improved. In response to this background, to be able to advance evaluation practices as part of the development process, the purpose of this thesis is to increase understanding of human factors evaluation of nuclear power plant control room systems.

The main object of study is the evaluation activity, which was considered from two different angles: how the design of the control room system and its impact are regarded by the evaluation activity; and the relation between the evaluation activity and the development process as a whole.

With regard to the first angle, the focus in this thesis was evaluation of a safety-critical system (the nuclear power plant control room system) – not how to assess safety. Evaluating a safety-critical system means assessing whether the system fulfils its intended purpose – which is to support safe operation and operator well-being. To explore the angle of how the evaluation activity should view the control room system and its impact, the first research question to guide the research efforts in this thesis was:

RQ1: What must be evaluated, from a human factors perspective, to assess a nuclear power plant control room system's ability to fulfil its intended purpose?

To explore the second angle, the relation between the evaluation activity and the development process as a whole, the second research question to guide the research efforts in this thesis was:

RQ2: Can human factors evaluation better support nuclear power plant control room system development? If so, how?

1.3 READING INSTRUCTIONS

This first chapter contains an overview of the background to the work presented in this study as well as a presentation of the purpose and research questions. Chapter two presents the theories and concepts relevant to the work presented in this thesis, and chapter three describes the research approach used. Chapter four summarises the studies included in the thesis. Chapter five analyses the findings from the studies to answer the research questions, and also presents the additional insights that emerge from this analysis. Chapter six discusses the findings and chapter seven describes how the findings in this thesis could be developed further in the future. Chapter eight summarises the conclusions of this thesis.

FRAME OF REFERENCE

This chapter describes the theories and concepts relevant for the work presented in this thesis. It elaborates on the nature of the object to be evaluated (2.1 The control room system as a socio-technical system) and the primary rationale for evaluation of nuclear power plant control room systems (2.2 Nuclear safety and safe operation). This chapter also describes requirements for human factors activities in the nuclear power domain (Section 2.3), as well as the relation between human factors and design (Section 2.4). It also presents important aspects of the evaluation activity; more specifically the concept of evaluation (Section 2.5), the context of the evaluation activity studied in this thesis (2.6 The development process), and the means used in the evaluation activity (2.7 Human factors methods).

2.1 THE CONTROL ROOM SYSTEM AS A SOCIO-TECHNICAL SYSTEM

A nuclear power plant control room system is an open system (it interacts with its environment) that purposefully transforms information *from* the process that is to be controlled into commands *to* the system that is to be controlled. A nuclear power plant control room system is also a socio-technical system. A socio-technical system is an open system where mutually interdependent elements – the technological subsystem, the personnel subsystem, and the work system design (the organisational structure and processes) – interact with one another and the external environment to *jointly* transform inputs into outputs (Hendrick and Kleiner, 2001). Because of the interaction between the elements, changes to one element will cause ripple effects in the system as a whole. Thus, a successful design of a socio-technical system cannot consider single elements in isolation, the effect planned changes may have on the other elements in the system must be considered. This concept is called joint optimisation.

The technological subsystem in a nuclear power plant control room system consists for example of operator interfaces presenting information from the process and making it possible to issue commands to control the process. The personnel subsystem consists of the operators in the shift team working in the control room, as well as the local operators who are part of the shift team but who do not have their primary place of work in the control room. The work design element consists for example of the roles and responsibilities within the shift team and the routines for how to work during major disturbances. An example of input from the environment is orders from the national grid authority to increase the power load.

The fact that the control room system is a socio-technical system has consequences for evaluation. The concept of joint optimisation states that modifications of parts of the control room system (changes in one element) will affect the function of the system as a whole (the joint function of all elements), so the various parts of the control room system must be evaluated as part of a whole. Even if only parts of a control room system are changed, the effect of these changes on the rest of the control room must be assessed, and vice-versa. In the nuclear power industry, this is most evident in the custom of thoroughly evaluating large control room modifications in the training simulator with real shift teams (so-called Integrated System Validation). In this activity, it is emphasised that the new design must be evaluated when it is integrated in the control room system as a whole (United States Nuclear Regulatory Commission [US NRC], 2012).

2.2 NUCLEAR SAFETY AND SAFE OPERATION

Safety can be defined in different ways. A traditional definition of safety is that it is freedom from unacceptable risk. One 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 difficult to predict. Trying to remove the possibility of all 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, 2011). This definition emphasises that a system should not only strive to avoid failures, it should aim to adapt its functioning to handle all conditions. Resilience engineering is the field that develops 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 (Hollnagel, 2011).

Nuclear safety is defined by the International Atomic Energy Agency (IAEA) 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 [IAEA], 2007). Nuclear power plants must not only be safe, they must be safe while producing electricity. Combining the demand to produce electricity with the

demand to uphold nuclear safety results in a demand 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, 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.

2.3 HUMAN FACTORS REQUIREMENTS IN NUCLEAR POWER

Human factors is stated as a general technical principle in a framework by the IAEA International Nuclear Safety Advisory Group [INSAG] (1999). This framework presents underlying objectives and principles of nuclear safety, and states 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.

IAEA has also issued a document (IAEA, 2016) containing requirements applicable to the design of nuclear power plants. It is meant to support organisations involved in the design, manufacture, construction, modification, maintenance, operation and decommissioning of nuclear power plants, as well as regulatory bodies. In this document, requirement no. 32 states that "*systematic consideration of human factors, including the human-machine interface, shall be included at an early stage in the design process for a nuclear power plant and shall be continued throughout the entire design process*" (ibid, p. 31). According to this requirement, human factors issues should be considered systematically, starting early and continuing throughout the development process.

The importance of human factors issues is highlighted in Swedish regulations too. Chapter 3 Section 3 of the regulatory code SSMFS 2008:1, by the Swedish Radiation Safety Authority, stipulates 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.

In addition to requirements and recommendations regarding the consideration of human factors in design in general, there are also requirements advocating human factors evaluation of nuclear power plant control room systems. The document containing requirements applicable to the design of nuclear power plants by IAEA (2016), also contains an additional human factors-related requirement

(no. 5.62) which states that “*verification and validation, including by the use of simulators, of features relating to human factors shall be included at appropriate stages to confirm that necessary actions by the operator have been identified and can be correctly performed*” (ibid, p. 32). The concepts of verification and validation are further defined in Section 2.5.

A document widely used in the nuclear power domain is the Human Factors Engineering Program Review model, NUREG-0711 (US NRC, 2012). It is published by the US NRC to provide their staff with guidance for the review of human factors activities related to the construction of new plants and modifications of existing ones. Regarding evaluations NUREG-0711 contains review criteria for so-called ‘HSI tests and evaluations’ (HSI, human-system interface), as part of HSI design, and for ‘Human factors verification and validation’. There is special emphasis on the activity referred to as ‘Integrated System Validation’, which is the final assessment of the control room. Although not intended as a process description, NUREG-0711 is used as one by many actors in the nuclear power plant domain to guide human factors activities. For example, the general advice for regulations provided by the Swedish Radiation Safety Authority (Swedish Radiation Safety Authority, 2008b) refers to NUREG-0711 for “*examples of methodology for the evaluation of control room modifications*”.

2.4 HUMAN FACTORS AND DESIGN

Human factors has a close relation to design. From the definition of human factors given by IEA (2018), Dul et al. (2012) derived three fundamental characteristics of human factors: 1) human factors takes a systems approach, 2) human factors is design driven, and 3) human factors focuses on two related outcomes – performance and well-being. This section will focus on the second characteristic, outlining aspects characterising design work that are relevant for the contents of this thesis.

Design is difficult to define, but one common theme for designers, as stated by Buchanan (1992, p. 14), is that they share a mutual interest in the conception and planning of the artificial. This statement is similar to the famous quote by Simon (1996, p. 114), that design is “*concerned with how things ought to be, with devising artifacts to attain goals*”. The implication for this thesis is that evaluation during the development process is about assessing something that does not yet exist. Evaluating something that does not yet exist requires projecting the impact of the final design, for example by imagining future use. Section 2.5 further elaborates on theory regarding evaluation.

Vicente et al. (1997) noted that human factors had a limited impact on design and that one of the reasons for this was that human factors researchers did not consider the problems and constraints that human factors practitioners face in their work. By characterising the nature of human factors design problems and

the practitioners' strategies developed to cope with them they hoped to stimulate human factors researchers to address the real challenges that practitioners face. According to Vicente et al. (1997), human factors design problems are so-called 'wicked problems' (Rittel and Webber, 1973).

Wicked problems is a central concept in design. As Buchanan (1992) describes it, the wicked problems approach suggests that in all but the most trivial design problems, there is a fundamental indeterminacy. An indeterminate problem has no definite conditions or limits to it (Buchanan, 1992). Rittel and Webber (1973) stated ten distinguishing properties of wicked problems, not all of which will be described here, but some are especially interesting to highlight in the context of this thesis. One property of wicked problems is that there is no definitive formulation of a wicked problem – understanding the problem is deeply intertwined with its resolution. Specifying the problem is not possible without specifying the direction for the resolution of the problem. Another property of wicked problems is that they have no stopping rule. Causal links in interacting open systems do not end, it is always possible to find a better resolution to a wicked problem. Work on a wicked problem is terminated for reasons external to the problem, such as lack of time or monetary resources. Yet another property of wicked problems is that their resolutions are never true or false, only good or bad. Wicked problems have many stakeholders, the viewpoint of each stakeholder differ, and no all-encompassing objective criteria for determining correctness exists. Furthermore, there is no definite way to test a wicked problem. Because of their complexity, all repercussions of a proposed resolution can never be comprehensively evaluated. The last property of wicked problems to be described here is that there is no enumerable set of potential resolutions to a wicked problem, nor a well-described set of permissible operations. Because they are open-ended, it is not possible to list all possible resolutions or admissible operations.

The implication of the wicked problems concept for this thesis is that when dealing with wicked problems that have no definitive formulation, evaluation is an important activity for examining what has been done and finding the way forward. Because wicked problems have no stopping rule, their resolutions are good or bad instead of true or false, and the number of possible resolutions is infinite, evaluation is a means of deciding when the design is good enough, not a way of determining that an ultimate resolution has been found. Because there is no definite way to test wicked problems evaluation will never provide a definite answer, but rather stronger or weaker evidence that the assessed solution will meet its intended purpose when implemented.

2.5 EVALUATION IN DESIGN

The focus on evaluation in this thesis require an understanding of the concept of evaluation, specifically of evaluation in design practice and the development process. The dictionary definition of 'evaluate' is to determine the value or condition

of something, usually by careful study (Britannica Academic, 2018). According to the Institute of Electrical and Electronics Engineers [IEEE] (1999), if there are no established acceptance criteria with which to compare data, there can be no evaluation, merely a measurement. Acceptance criteria, in turn, can be formal, specific criteria related to the measurement such as operator diagnosis within a specific time limit, or informal, such as the evaluator's opinion regarding the acceptability of the performance (IEEE, 1999).

In the context of understanding the concept of design, Lawson and Dorst (2009) describe evaluation as one of five design activities that correspond to groups of skills that designers have. For example, designers must choose between generated alternatives, and must also know when to stop generating alternatives. The special evaluative skill designers must have is to make judgements between alternatives along many dimensions that cannot be reduced to a common metric (Lawson and Dorst, 2009). Because of the nature of the problems designers are used to dealing with (wicked problems) designers focus on the resolution, not the problem. Designers tend to create resolutions and evaluate them to make gradual improvements (Lawson and Dorst, 2009). If exhaustive analysis of the problem is not possible, evaluating generated alternatives to see how they may be improved is the key to be able to move forward.

Scriven (1967), when discussing evaluation of educational instruments, distinguished between two types of roles for evaluation: *formative* and *summative*. Formative evaluation, as described by Scriven (1967), is an assessment performed with the purpose of improving that which is being evaluated. Summative evaluation is described as assessment of the final product of a process. Noyes (2004) described formative evaluation methods as those more appropriate for use during the development of a product, and summative methods as those for application with the finished product. Noyes (2004, p. 60) also used the following analogy to describe the differences: "*when the cook tastes the soup, this is formative evaluation, but when the guest tastes the soup, this is summative evaluation*". In the context of usability testing, Nielsen (1993) described formative evaluation as an assessment done in order to improve design as part of an iterative design process. Summative evaluation is described as aiming to assess the overall quality of a design. The definitions by Nielsen (1993) are used in this thesis.

A different categorisation of evaluations can be made by distinguishing between verification and validation. In an ergonomics standard for the ergonomic design of control centres (ISO, 2006, p. 1) the evaluation process is defined as the "*combined effort of all verification and validation (V&V) activities in a project using selected methods and the recording of the results*". Verification is defined as "*confirmation, through the provision of objective evidence, that specified requirements have been fulfilled*" (ibid, 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). The focus in these definitions is the confirmation of fulfilled

requirements (verification) and needs (validation), indicating that verification and validation are summative evaluation activities. This view is in line with how verification and validation are described in NUREG-0711 (US NRC, 2012), a document often used as support when planning human factors activities in the nuclear power domain.

A slightly different view of verification and validation can be found within the field of systems engineering. Here, the verification and validation processes are described as closely related to quality management (INCOSE & Wiley, 2015), which indicate a summative focus. However, the importance of identifying what needs to be changed in the design if requirements and needs are not met is also emphasised, which indicate formative use of the verification and validation results as well, even if the activities primarily are performed for summative purposes.

2.6 THE DEVELOPMENT PROCESS

The evaluation activity is one of many activities performed to develop new products (or, as in this thesis, control room systems), and in order to understand the evaluation activity it helps to understand its relation to other activities performed in a development process. This section describes this relation using two different ways of modelling the development process: one based on how designers work and the other on what is produced in different parts of the process. The relation between the phases of a development process and design decisions made is explained, as well as the role of requirements in development work and their relation to evaluation. Lastly, the section describes human factors activities in the control room development process.

2.6.1 Models of the development process

The first way of modelling the development process presented here highlights the role of the evaluation activity in design work. This first way of modelling the development process focuses on how designers work. A conventional way to model the process is to divide it into the activities of defining the problem, analysing the problem, formulating requirements, generating solutions, choosing between solutions using the requirements, and implementing the chosen solution – with iteration between activities if needed (Lawson and Dorst, 2009). Lawson and Dorst (2009, p. 33) call this “*the conventional analysis synthesis evaluation model of designing*”. A similar process, defined for human-centred design activities, is described in ISO (2010). This process also contains ‘analysis–synthesis–evaluation’-steps with iterative loops but is more tailored to human factors-related design activities (Figure 2). In a messy reality, however, the actual work process of the designer (or the human factors specialist) may not always be as straight-forward as these processes make it seem. Even so, this way of modelling the process helps in understanding evaluation since it highlights the evaluation activity’s fundamental role in design work – to determine value in order to decide how to proceed.

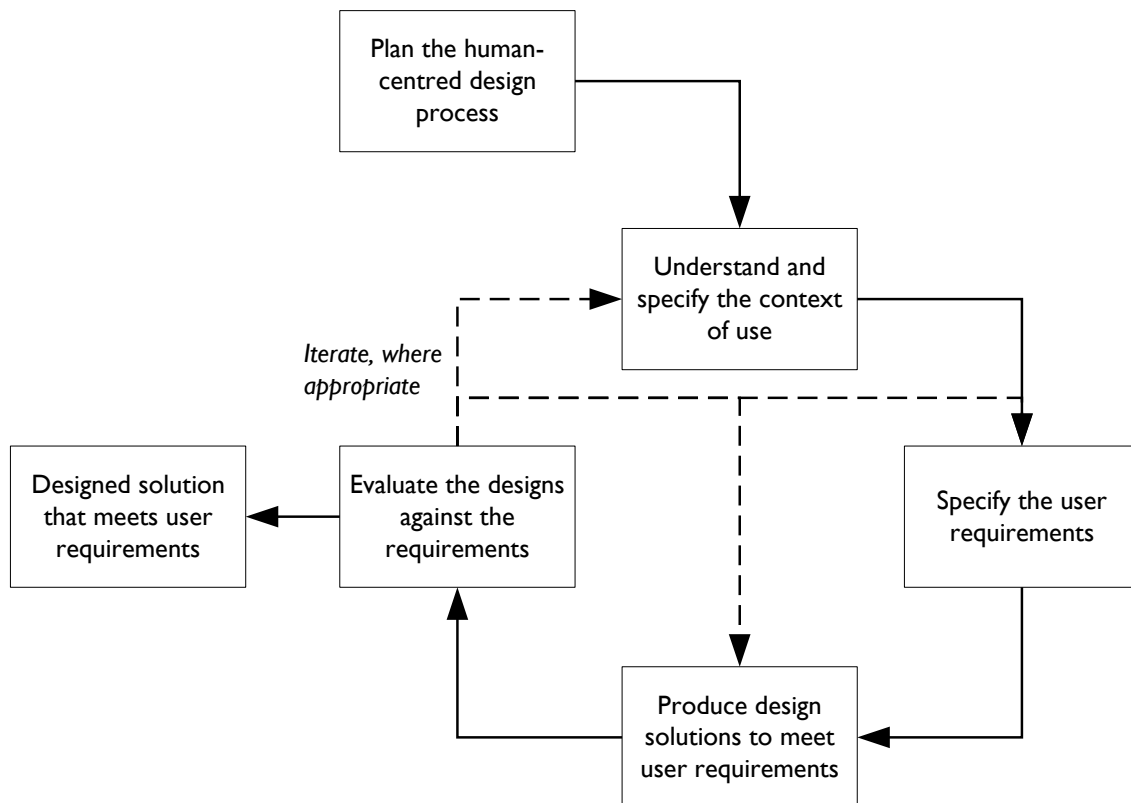


Figure 2: Human-centred design activities. Adapted from ISO 9241-201:2010 (ISO, 2010).

However, the process presented above is less useful when trying to understand evaluation in relation to the progression of design work and the work being done in the development project as a whole. The other way of modelling the development process presented here is based on what must be produced in different parts of the process (Ulrich and Eppinger, 2003; Lawson, 2006). Modelling the development process this way has benefits from the perspective of steering development work. Ulrich and Eppinger (2003, p. 14) describe the development process as “*the sequence of steps or activities which an enterprise employs to conceive, design, and commercialize a product*”. In this definition a development process concerns the steps the enterprise undertakes, not the activities of the individual designers. Ulrich and Eppinger (2003) stated that the benefits of using a well-defined development process are that it supports quality assurance, coordination, planning, management, and improvement. The two ways of modelling presented here do not contradict each other. The iterative loops of analysis-synthesis-evaluation in the process based on the designers’ work are part of the phases in a process based on what must be produced (Bligård et al., 2016). The analysis-synthesis-evaluation loop may be executed several times within each phase, thus being repeated several times during the course of the development process. The implication for this thesis is that evaluation is a recurring activity that may be undertaken several times as the development process progresses over time.

Though evaluation is a recurring activity it cannot be undertaken the same way each time. Aspects such as purpose, stakeholders, and available resources will differ between phases. Presenting the content of different phases of a development process based on what is produced helps understand how the circumstances for the evaluation activity shift between phases. There are numerous suggestions for how development processes for organisations (based on what must be produced) are, or should be, structured. They differ, among other things, in how much of the product life cycle they cover and where the lines between different phases are drawn. Some processes end after the design is finished and others include production. For example, the product development process suggested by Ulrich and Eppinger (2003) has six phases: planning, conceptual design, system-level design, detail design, testing and refinement, and production ramp-up (a description of each phase is included in Table 1). A plant, such as a nuclear power plant, is normally not viewed as a product. That does not mean that developing and modifying it 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 suggested that an established and documented modification process ensures consistency, repeatability, and traceability. Hale et al. (2007), in a special issue of the academic journal *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 (a description of each phase is included in Table 1). The beginning of this process is similar to the one proposed by Ulrich and Eppinger (2003), but the final phases differ. 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 correspondence between the two processes is illustrated in Table 1.

The phases in development processes such as those presented in Table 1 differ in the sense that the specificity in the design decisions made during the process gradually increases. 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. Constraints are applied by making design decisions (Bligård et al., 2016). A design decision is when a design variable is given a specific value, for example when the design variable ‘colour’ is given the value ‘red’ or a specific colour code. Constraining the possible values of design variables through design decisions limits underlying and dependent design variables, both in terms of which variables are relevant for consideration as well as possible values (Bligård et al., 2016). This makes the design variables considered increasingly specific, creating a natural order in the development process: to gradually move from making general design decisions to specific ones. Papin (2002, p. 2) described this process for the nuclear power domain, calling it “*the addition of successive ‘layers’ to the initial choice of the reactor technology*”. Phases

Table 1: Correspondence between phases in the development processes proposed by Ulrich and Eppinger (2003) and Hale et al. (2007).

Ulrich and Eppinger (2003)	Hale et al. (2007)
Planning Create project mission statement specifying target market, business goals, key assumptions, and constraints.	Business development Clarify the business case for pursuing an opportunity to develop a new technical system.
	Feasibility study Clarify the technical feasibility of the project and the possibilities of meeting the profitability requirements.
Conceptual design Deliver a design concept that describes the form, function, and features of the product.	Conceptual design Develop concept alternatives by selecting and arranging building blocks, select the best solution with respect to project objectives.
System-level design Define the product architecture, decomposing the product into subsystems and components.	Basic design Optimise basic design, define detailed design requirements and mature design to reduce cost, schedule, and quality uncertainties.
Detail design Specify the product in detail, for example the geometry, materials, and tolerances.	Detailed design Meet design requirements.
Testing and refinement Where preproduction versions of the product are constructed and evaluated to finalise the design.	
Production ramp-up Ramp-up trains the work force and works out remaining problems in the production process.	
	Fabrication, Installation, Start-up, Commissioning Realisation of design, front-end engineering, final checking and test before hand-over to customer.

in development processes modelled on what must be produced in different parts of the process can thus be differentiated based on the specificity of design decisions considered in that phase. A phase establishing a more overall design solution precedes a phase where a more detailed design is developed. In the processes described above (and compared in Table 1) this primarily applies to the phases of planning (and equivalent phases), conceptual design, system-level design/basic design, and detail/detailed design.

As stated above, this way of modelling supports steering development work and the actual development work being done by designers may not follow this process strictly. A designer might have to consider details to be able to make more general design decisions, thus having to ponder design decisions that ‘belong’ to another phase than the one with which the project is currently involved. However,

achieving coherent systematic development work (and avoiding infinite loops), requires design decisions being definitively finalised in a top-down manner (Bligård et al., 2016).

Requirements play an important role in development work, and this role has implications for evaluation. Requirements are precise descriptions of *what* a product (or, as in this thesis, a control room system) has to do, but not *how* this is to be achieved (Ulrich and Eppinger, 2003). Requirements cannot be exhaustively specified in the beginning of the design process, but must rather evolve in parallel with the design, or at least be revisited and adjusted during the course of the development process (Ulrich and Eppinger, 2003; Pew and Mavor, 2007; Berlin et al., 2017; Braha and Reich, 2033). Requirements relate to evaluation in the sense that they provide knowledge of what it is important to assess in a specific design, as well as acceptance criteria specifying when the design can be deemed to be good enough. The process of human-centred design activities described in ISO (2010) illustrates this relationship (Figure 2). In particular, the activity of verification is explicitly defined as assessment of the fulfilment of requirements (see for example the definition of verification in ISO, 2006).

2.6.2 Human factors activities in control room system development

This thesis is concerned with the human factors aspects of control room system evaluation. Human factors evaluation is seldom done in a vacuum. Rather it utilises the output, material, and experience from activities undertaken earlier in the development process, and evaluation output, material, and experience can be used as input to other activities. Understanding human factors work in the development process is therefore a way to understand the circumstances of human factors evaluation. 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. One standard recommending a process for how human factors aspects are to be included in control room system design is “*ISO 11064-1:2000 Ergonomic design of control centres – Part 1: Principles for the design of control centres*” (ISO, 2000). This standard presents a framework for a human factors design process for control room systems consisting of the following phases: clarification; analysis and definition; conceptual design; detailed design; and operational feedback (a description of each phase is included in Table 2). ISO 11064-1:2000 emphasises the iterative nature of the process. The review guide NUREG-0711 (US NRC, 2012), which is used as a process description by many actors in the nuclear power plant domain, does not have phases but stipulates a number of elements: planning and analysis; design; verification and validation, and implementation and operation (a description of each element is included in Table 2).

Table 2: Correspondence between phases in ISO 11065-1:2000 (ISO, 2000) and NUREG-0711 (US NRC, 2012). The phases of the process presented by Hale et al. (2007) are added for comparison.

Hale et al. (2007)	ISO 11064-1:2000	NUREG-0711
<p>Business development Clarify the business case for pursuing an opportunity to develop a new technical system.</p>	<p>Clarification Clarify the purpose, context, resources and constraints of the project, take into account existing situations which could be used as a reference.</p>	<p>Planning and analysis Develop a plan for the human factors work. Perform analyses such as operational experience review; functional requirements analysis and function allocation; task analysis; analysis of staffing and qualifications; and analysis of important human actions.</p>
<p>Feasibility study Clarify the technical feasibility of the project and the possibilities of meeting the profitability requirements.</p>	<p>Analysis and definition Analyse the functional and performance requirements to create a preliminary functions allocation and job design.</p>	
<p>Conceptual design Develop concept alternatives by selecting and arranging building blocks, select the best solution with respect to project objectives.</p>	<p>Conceptual design Develop initial room layout, furnishing designs, displays and controls, and communications interfaces necessary to satisfy the needs identified in the Analysis and definition phase.</p>	<p>Design Translate functional- and task-requirements into design requirements. Identify and select candidate designs, define the detailed design, and perform tests and evaluations. Verification and Validation Comprehensively determine that the final human factors design conforms to accepted design principles, and enables personnel to successfully and safely perform their tasks to achieve operational goals.</p>
<p>Basic design Optimise basic design, define detailed design requirements and mature design to reduce cost, schedule, and quality uncertainties.</p>		
<p>Detailed design Meet design requirements.</p>	<p>Detailed design <i>[validation included here]</i> Develop the detailed design specifications necessary for construction and/or procurement; the control room system's content, operational interfaces and environmental facilities.</p>	
<p>Fabrication, installation, start-up, commissioning Realisation of design, front-end engineering, final checking and test before hand-over to customer.</p>	<p>Operational feedback Conduct a post commissioning review to identify successes and shortcomings in the design in order to positively influence subsequent designs.</p>	<p>Implementation and Operation Ensure that the as-built design conforms to the verified and validated design. Consider the effect implementation of the design has on personnel performance. Ensure that the conclusions drawn from the validation remain valid over time. Ensure that no significant safety degradation occurs because of any changes made in the plant.</p>

ISO 11064-1:2000 and NUREG-0711 highlight the work of primary concern for the human factors specialist in a control room system development project. The overall correspondence between them is shown in Table 2. One aspect more pronounced in human factors processes for control room systems than in more generic processes is the development of procedures and training. The emphasis on procedures and training for nuclear power plants is evident in NUREG-0711, where procedure development and training programme development are equal parts of the design element together with human-system interface design. In the ISO 11064-1:2000 process the development of training regimes and the like is included in the detail design phase. The operation of control room systems is often very dependent on both procedures and training of personnel. While not unimportant for other products, procedures are not necessarily a requirement for use to be possible. 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. The implication of this for evaluation is that it emphasises the importance of assessing the control room *system*, not only the operator interfaces and other parts of the physical control room. Other elements, such as procedures and training, should be included in assessment.

An aspect that is more emphasised in the last element of NUREG-0711 than in the processes presented in Table 1 is the work done in relation to implementation of the design. A control room system is often in operation 24/7, which means that modifications to an existing control room system must be done while (at least part of) the plant is in operation or must be monitored in some way. Because of this, human factors issues during implementation must be considered in addition to the human factors issues of the implemented design in operation. For evaluation, this could mean including the tasks undertaken in the control room system during implementation of a new design, and not only focusing on how tasks can be performed after the new design has been implemented.

Yet another difference between processes for development of control room systems and other products (Table 1) is the emphasis in the former 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 room system during its lifespan. This is work that would normally fall outside the scope of a development project, but which is relevant to the plant owner who is responsible for operation of the plant. Thus evaluation in this part of the control room system life cycle also is important to ensure that the system continues to fulfil its intended purpose.

2.7 HUMAN FACTORS METHODS

Methods are important means for executing of human factors activities in the development process. According to the dictionary definition, a method is “*a particular procedure for accomplishing or approaching something, especially a*

systematic or established one” (Oxford Dictionaries, 2018). Human factors methods, specifically, are described by Andersson et al. (2011) as means for human factors specialists to achieve their ends, comparable to how a product is used by a user to execute a task. In line with the definition of human factors by IEA (2018) human factors methods in this thesis are defined as methods that are concerned with interactions between humans and other parts of a system.

There are many ways to categorise methods. Stanton et al. (2005) divided human factors methods into categories of data collection; analysis and assessment of different constructs (such as tasks, human error and workload); as well as design. Another example is Wilson (2005), who used the following categories to categorise human factors methods: general methods (that may be used within any of the other categories); collection of information about people, analysis and design; evaluation of human-machine system performance; evaluation of demands on people; and management and implementation of ergonomics. One theme reoccurring in both Stanton et al. (2005) and Wilson (2005) is to categorise methods according to the design activity they support, such as data collection, analysis, synthesis, and evaluation.

The category of human factors methods of interest in this thesis is evaluation methods. Drawing on the definitions of evaluation and method presented previously in the thesis, an evaluation method is here defined as a procedure through which the condition or value of something is determined. To distinguish an evaluation method from a data collection method, the method must include some sort of comparison with acceptance criteria as well, formal or informal (Electric Power Research Institute, 2005).

Functions a human factors method must fulfil can be found in the design domain. As argued by Dul et al. (2012), being design driven is a fundamental characteristic of human factors. Principal features of design methods should thus apply to human factors methods used in the development process. Cross (2008) defined design methods as any procedures, techniques, aids, or ‘tools’ that support designing, and stated that the main intention of design methods is to attempt to bring rational procedures into the development process. Design methods have two principal features, they: 1) *formalise* certain procedures of design, and 2) *externalise* design thinking (Cross, 2008). The purpose of formalisation is to minimise oversights, to avoid important factors being overlooked, and to widen the search for appropriate resolutions to design problems. The purpose of externalisation is to bring the thoughts on design out of the mind of the designer and out “into the world” – for example in verbal form or on paper. This process facilitates teamwork and the resolution of complex problems. Cross (2008) also stated that putting systematic work onto paper is a way of freeing the designer’s mind to pursue intuitive and innovative thinking. These features of a method further nuance the image of what an evaluation method must be able to do.

RESEARCH APPROACH

This chapter presents the author's background, her research interests and the philosophical worldview that influenced her work on this thesis. The chapter also describes and motivates the research design utilised.

3.1 RESEARCH INTERESTS AND WORLDVIEW

The author's educational background is in industrial design engineering. The design focus in this education taught how to use an iterative development process to handle wicked problems, how to work with things that do not yet exist, to be concerned with how things ought to be. The author's education also formed a view of the world where impact is achieved through the use of artefacts, and thus the use, and consequently user needs, must be a focus in development. The author have worked as a practitioner within the field of human factors engineering for over ten years, primarily with control room system development. The 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. The author's practical experience highlighted the importance of methods that are usable in practice. Methods are means to meet an end, bringing value only when they are used. These experiences have to a large extent shaped the scope and focus of this thesis.

The philosophical worldview that influenced the work in this thesis is pragmatism, where the concern is with what works to solve the problem at hand, using suitable assumptions and methods regardless of philosophical underpinnings (cf. Creswell, 2014). The fact that design affects human behaviour is a founding assumption within the field of human factors. Meister (1991) calls this a behavioural-physical transformation. The assumption that environmental aspects (such as design) affect human behaviour presupposes the belief in an objective world, that realities exist outside the mind (cf. the ontological view of realism, Crotty, 1998). If belief in an objective reality is a necessity for design of control room systems to be worthwhile, then it is also a necessity for the evaluation of that design. Evaluating a design is all about predicting its potential impact. Exploring and developing the practice of evaluation thus requires consideration of causal effects on an objective reality (cf. postpositivism as it is described by Creswell, 2014). Causal effects on an objective reality are however not enough to assess a control room system. Understanding that the meaning human beings construct as they engage with this reality (cf. constructivism as it is described by Creswell, 2014) is also needed, since this meaning will affect their behaviour

and well-being, which in turn will affect the performance of the control room system as a whole. Pragmatism allows using both approaches and allows them to complement each other since it focuses on the problem that is to be solved. The research methods compatible with a pragmatist approach are all methods that study what needs to be learned to solve the problem (Creswell, 2014). In this thesis, the two approaches are combined in the sense that while investigation of the meaning constructed by agents related to the object of study constitutes the main source of data, the data is interpreted vis-a-vis the belief in causal effects in an objective reality. In addition, the analysis of interpreted meaning has occasionally been complemented with more objectively observed data.

The purpose of this thesis was to increase understanding of human factors evaluation so as to be able to advance evaluation practices. This makes the work presented here an example of a type of design research called *research for design* (Frayling, 1993; Zimmerman et al., 2010). Research for design focuses on improving design practice, and the outcome often includes frameworks, philosophies, design recommendations, design methods, and design implications – known as *theory for design* by Zimmerman et al. (2010). In the present thesis, the implications of undertaking research for design are that quality criteria is derived from what is usable in design practice, not from how accurately reality is modelled. This is very much in line with the pragmatist approach.

Another category of design research is *research through design* (Frayling, 1993; Zimmerman et al., 2010). This category is defined based on the method used and not, as in the case of research for design, based on the purpose of the research. Research through design is about iteratively designing artefacts as a way of investigating what a potential future might be, it is about using making as a method of inquiry to address wicked problems (Zimmerman et al., 2007; Zimmerman et al., 2010). Advantages of research through design include a closer connection to the context of use. Through this connection important factors that are difficult to replicate in an experimental setting may be considered. Methods are means to an end and must be designed for a desired impact much like an artefact. Thus, research through design was a useful approach for the part of the work that explored methods used in the evaluation activity (part of RQ2).

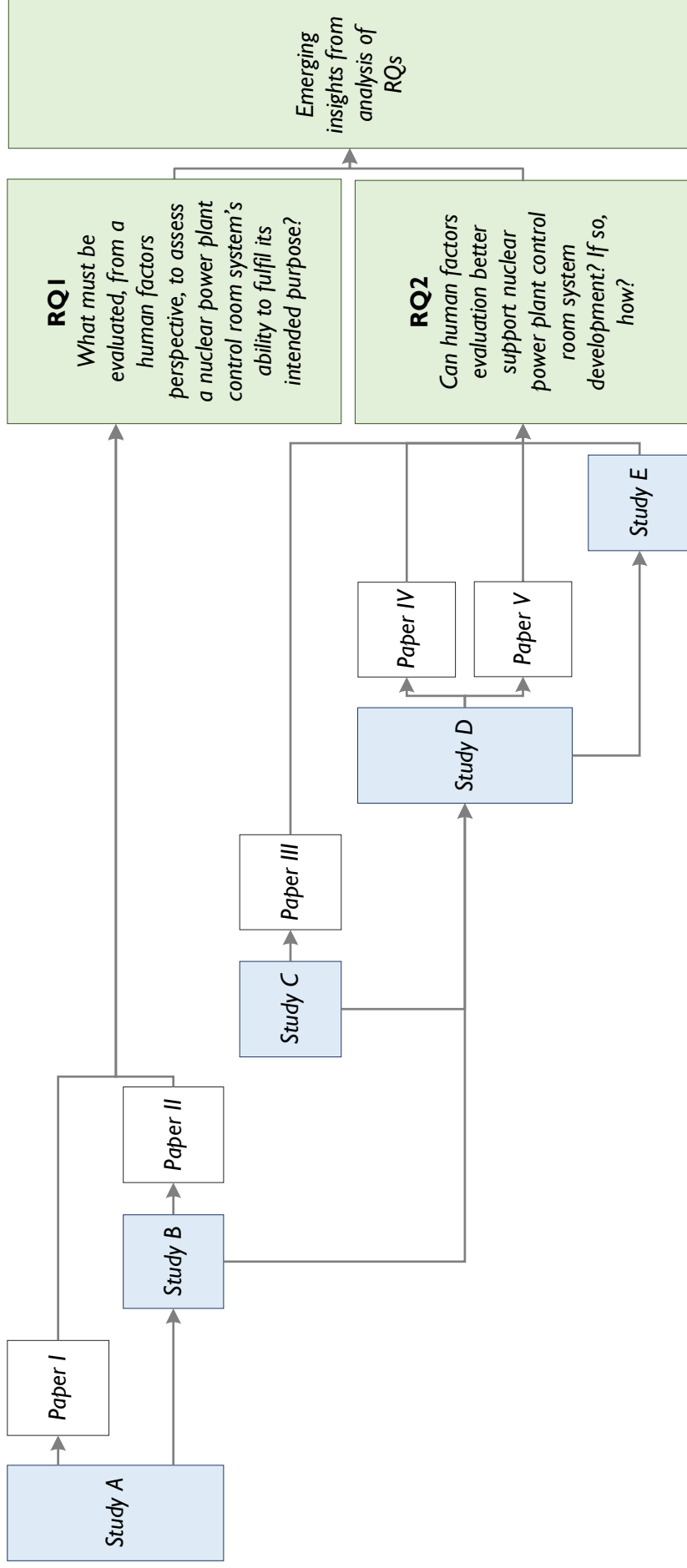


Figure 3: Connections between research studies, papers, and research questions.

3.2 RESEARCH DESIGN

The object of study in this thesis, the evaluation activity, was considered from two different angles: how the design of the control room system and its impact are regarded by the evaluation activity (RQ1); and the relation between the evaluation activity and the development process as a whole (RQ2). Figure 3 illustrates the connections between research studies, papers, and research questions in this thesis. As was described in Section 3.1, the purpose of the work presented in this thesis means that it falls under the category of research for design, which determines which phenomena are of interest and how findings are valued. In addition, part of the exploration of the second research question used a ‘research through design’-approach.

Planning an evaluation or developing an evaluation method requires deciding *how* to study the object to be evaluated; in other words, how the evaluation activity should regard the object of study. Evaluation is about taking measures and comparing these to acceptance criteria. To be able to measure you must know *what* to measure. The starting point for this thesis was thus to consider the measures needed to assess the control room system (RQ1). Identification of the aspects that contribute to the control room system’s ability to support safe operation using an empirical experimental approach may be possible in theory, but is difficult to achieve in practice. In theory variables could be changed and the corresponding effect on safe operation and operator well-being could be monitored, but the complexity of the control room system and its environment made this approach difficult. Instead, an empirical qualitative approach (an interview study) was undertaken to utilise the knowledge of subject matter experts (Study A). The resulting empirical data was complemented with a literature study of what other researchers regarded as important to assess in control room systems, and compared with the results of a thematic analysis of measures targeted by existing evaluation methods (Study B).

The second research question related to the relation between the evaluation activity and the development process. Study C examined when and how evaluation is performed in nuclear power plant development projects. Here, a literature study of empirical nuclear power plant control room evaluations was deemed to be an efficient and comprehensive way to gain understanding of practice within the domain.

The next part of the work in this thesis was concerned with the object of study itself, the evaluation activity, and the methods needed to execute it. This was also part of the exploration of the second research question, and was studied through a ‘research through design’-approach. In Study D (case studies), a qualitative empirical approach was undertaken. This approach was chosen because of a need to study the evaluation activity and its methods in their context. Following the case studies, the evaluation methods tested were modified. Due to the lack of access to suitable modernisation project in industry another qualitative empirical

approach (focus groups, Study E) was chosen to assess the modified method combination, and its generalisability to other domains. In this way, the knowledge of subject matter experts could be utilised in an efficient and feasible manner. The iterative development process undertaken in Study D and Study E (identifying requirements for suitable evaluation methods, selecting and modifying methods, testing the methods, modifying them again, and assessing them again) was used as a way to gain knowledge of the evaluation activity; as a way to perform research through design.

Lastly, the findings from all the studies were analysed with the aim of answering the research questions posed. Apart from answering the research questions, further insights emerged from the analysis which were compiled into a number of perspectives to serve as decision-making support in evaluation planning and method development.

RESULTS – SUMMARY OF STUDIES

This chapter contains a summary of the studies resulting in the appended papers, together with a presentation of the purpose, method, and key findings of Study E.

4.1 STUDY A (PAPER I)

The first research question concerned how the evaluation activity should regard the control room system design and its impact. Study A was the first step towards answering that research question and sought a foundation for evaluation measures for the primary rationale for nuclear power plant control room system evaluation – safe operation.

4.1.1 Purpose

The purpose of Study A 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.

4.1.2 Method

Study A 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). The characteristics of the interviewees can be found in Paper I. The semi-structured interviews contained both broader and more detailed questions. The more detailed questions utilised different viewing angles of the investigated topic to trigger the interviewees' thoughts in order to obtain more extensive answers. These angles were: task, functional, and structural point of view, as well as the necessary properties of the structural elements. The interview data was transcribed in full. The qualitative material from the interviews was analysed using thematic analysis.

4.1.3 Key findings

The paper concluded that aspects contributing to safe operation can be sorted into the following themes: situations, functions, tasks, characteristics, and structural elements. *Situations* describe states of and/or events in the surrounding environment that the control room system must be able to handle. *Functions* are the abilities the control room system must have, and *tasks* are what operators or technical systems in the control room system must be able to perform. *Structural elements* are the

entities that constitute the control room system, and the *characteristics* of the structural elements establish conditions for the design of artefacts as well as the behaviours and abilities of personnel. However, the formulation of the purpose in Paper I and the way the themes were defined are not in agreement with the definition of the control room system used in the present thesis. The contents of the situations theme and a sub-theme of the structural elements theme, ‘process and instrumentation and control (I&C) systems’, affect the control room system, but are a part of its environment, not a part of the system itself. A better formulation of the purpose in Paper I would be that it sought to identify aspects affecting the control room system’s ability to support safe operation from a human factors perspective.

One prerequisite for safe operation is controlled performance, and the contents of the identified themes provide examples of what is required to achieve this in the context of the nuclear power plant control room system. Together the themes can serve as a basis for defining evaluation measures.

4.2 STUDY B (PAPER II)

The findings of Study A were used as input for Study B, which continued to explore the first research question – what must be evaluated to assess the control room system’s ability to support its intended purpose.

4.2.1 Purpose

The purpose of this study was to identify categories of measures that can guide the choice of evaluation methods for assessing nuclear power plant control room systems. ‘Category of measures’ is a term used here to denote a group of measures that target the same quality of the system to be measured.

4.2.2 Method

The first step of the study consisted of identifying categories of measures. Measures targeted by existing human factors evaluation methods were compiled and analysed. For each method, measures collected by the method were noted. A thematic analysis was performed to identify themes within the compiled measures. The concluding list of themes was a set of categories of evaluation measures describing the different kinds of measures that are targeted by existing evaluation methods.

In the second step of the study, the goal was to identify these categories’ relevance in evaluation of nuclear power plant control room systems. “Relevance” here meant that a category of measures covered measures deemed important for evaluating nuclear power plant control room systems. This was done through comparison with literature detailing necessary aspects of well-functioning nuclear power plant control room work and literature proposing or utilising measures for evaluation

of nuclear power plant control rooms (Paper I was part of this body of literature). The assumption made in this analysis was that correspondence between the identified categories of measures and the literature would indicate the categories of measures that are relevant for control room evaluation.

4.2.3 Key findings

The study concluded that measures targeted by human factors evaluation methods can be grouped into six categories:

- *System performance*: Measures of the overall outcome of the functioning of the system as a whole. In the case of control room systems, this could be noting the value of crucial plant parameters such as tank levels and temperatures, when a shift team handles a scenario in a simulator. The fact that crucial plant parameters are kept within the limits for which the plant has been designed is reliant both on the functioning of technical systems (e.g. automatic functions) and the way the operators operate the plant.
- *Task performance*: Measures of how users perform tasks, such as the number and nature of errors in use, or time to complete tasks. This category also includes qualitative assessments of the users' way of working.
- *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.
- *Identification of design discrepancies*: Focus on the design of the control room system, identification of parts of the design that may induce errors in use or in other ways hinder the effect the design is meant to achieve. Typically this is done through a comparison with what experience has shown to be the best way to do it (i.e. design guidelines).

The definition of user experience however, needs to be further defined over and above what was done in Paper II to more specifically denote the measures in that category. There are many definition of user experience. One example is the definition in ISO (2010, p. 3), which states that user experience is a “*person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service*”. This definition does not sufficiently describe the user experience measures in the thematic analysis in Study B. Subjective opinions of perceptions and responses could be measures within the task performance, teamwork, and use of resources categories as well, for example how a user rates different dimensions of mental workload. User experience is a concept traditionally more often utilised for consumer market products, but Savioja et al. (2014) explored the concept of user experience within complex systems in the nuclear power domain. They introduced

the concept of user experience as “*an indicator of the users’ subjective feeling of the appropriateness of the proposed tool for the activity*” (ibid, p. 429). This definition is useful to denote the nature of the user experience category in Paper II. This definition allows user experience measures to be distinguished from measures of the users’ subjective opinions on task performance, teamwork, and use of resources.

Methods for collecting data from all six categories of measures are needed to fully assess a nuclear power plant control room system. When planning evaluation of such a system, the six categories of measures can guide the choice of human factors evaluation methods. In practice, the categories can be used to ensure that methods chosen for an evaluation cover important aspects of the control room’s contribution to safe operation. For example, the categories can be used to map the targeted measures in a planned evaluation and highlight gaps. When gaps are known, additional methods can be added to the evaluation to provide a more comprehensive assessment. By using the categories of measures presented in this paper control room evaluations during a development process can be more consciously planned.

4.3 STUDY C (PAPER III)

The second research question in this thesis considered the relation between the evaluation activity and the development process. Study C was undertaken to investigate current evaluation practices in the nuclear power domain, more specifically when and how evaluation is done in development projects. The concept of levels of design decision specificity (Bligård et al., 2016) was used to be able to compare the timing of assessments between projects. According to this concept, design variables considered in a development process are gradually more and more specific, creating a natural order to gradually move from making general design decisions (higher design level) to specific ones (lower design level).

4.3.1 Purpose

The purpose of this study was to compare utilised approaches to evaluate control room systems in the nuclear power industry and to explore how they relate to design decisions at different levels of specificity. The assumption behind this purpose was that identified gaps shown by a comparison and mapping of evaluation approaches to design decision levels should indicate needs for further development of evaluation approaches.

4.3.2 Method

This study was a review of academic literature on the subject. More specifically, a comparison of utilised approaches to evaluate control room systems in the nuclear power industry and an exploration of how they relate to design decisions at different levels of specificity. The review was executed in two steps.

In the first step, approaches utilised to evaluate control room systems in the nuclear power industry were sought by searching in a scientific database. From the search results, papers concerning evaluations proposed or performed in the industry were identified based on the contents of titles and abstracts. The review of the selected papers focused on determining if the proposed or performed evaluation activities were formative or summative, and on comparing the methodology used, especially the system representation used for the assessment.

The second step of the study explored how the identified evaluation approaches related to different levels of design decision specificity. For each reviewed paper, the evaluated design was mapped to corresponding design levels.

4.3.3 Key findings

The study concluded that formative evaluation approaches for more general design decisions (higher design level) are less common and not described in as much detail as summative evaluations including more specific design decisions (lower design level) as well. This gap has to some extent been addressed by academia, but guidance can be further detailed and improved, for example by further investigating evaluation approaches utilising system representations available in earlier project phases (when more general design decisions are normally made). Much can be gained from assessing control room system design decisions at higher levels, since this means design concepts can be evaluated earlier in the development process, making changes easier and cheaper to implement. The paper points to a need to further develop methodologies and methods suitable for formative evaluation of design decisions at higher levels, and to assess their applicability for control room system evaluation.

4.4 STUDY D (PAPERS IV AND V)

Study D continued the exploration of the second research question and investigated the means needed for the evaluation activity to play its part in the development process, the evaluation methods. Study D focused particularly on use in practice and the empirical data was used for two papers, Papers IV and V.

4.4.1 Purpose

Paper IV

The purpose of this paper was to seek understanding of the practical use of human factors evaluation methods for formative assessment of higher-level design decisions (early evaluation) in control room systems. The empirical data was used as a foundation for guidelines for human factors methods in early evaluation.

Paper V

The purpose of the study presented in this paper was to test the feasibility of methods for early formative evaluation of nuclear power plant control room systems, in other words assessment of higher-level design decisions.

4.4.2 Method

Case studies as a research method allow exploration of a method's advantages and disadvantages when used in the chosen context, and were thus deemed suitable for assessing method use in practice. Three control room modification projects at a Swedish nuclear power plant were used as cases for this study. They were all in stages in the development process where changes to the design were still possible, and formative evaluation was therefore worthwhile.

Two evaluation methods were tested in the three cases, a scenario-based talkthrough and heuristic evaluation. The scenario-based talkthrough was a method where the proposed design was assessed by letting users go through a number of scenarios using a representation of the design. In the heuristic evaluation method, a small set of evaluators assessed how well the proposed design complied with design guidelines. Both methods fulfilled the identified prerequisites for early formative evaluation of control room systems, and were modified to suit the specific context even better.

Execution of evaluation workshops

In each case, two assessment workshops were held, the first using the heuristic evaluation method and the second using the scenario-based talkthrough method. Participants were the projects' human factors specialists, persons with operational knowledge but not actively working as operators (heuristic evaluation), operators (scenario-based talkthrough), designers (in some of the scenario-based talkthrough workshops), and project leaders (in some of the scenario-based talkthrough workshops). The project's human factors specialist was the moderator during the workshops and the researcher merely observed. In the workshops, the designs to be evaluated were represented using 2D drawings on paper.

Data collection for research study

Discrepancies and new requirements identified during the workshops were compiled through note-taking and video analysis. After each workshop, the participants were asked for their opinions of the method, either through semi-structured interviews or written questions sent via email. The participants were also asked to rate their responses in questionnaires.

Data analysis – Paper IV

In order to analyse the data for Paper IV, the interview data and comments written in the questionnaires were coded into three broad categories. These categories were 1) positive and 2) negative statements regarding the methods and their use, as well

as 3) concrete suggestions for method improvement. Each coded statement was abstracted and synthesised into guidelines for method design. Similar guidelines were sorted and consolidated and if needed new descriptions were formulated to better denote the content of the new consolidated guideline.

Data analysis – Paper V

For Paper V, the two types of data collected, relating to the result of the evaluation workshops (discrepancies and new requirements) and relating to the participants' experiences of the methods (expressed in interviews and questionnaires), were first compiled and analysed separately. For the first type, the number of identified discrepancies and new requirements was counted, and items identified in both the heuristic evaluation and the scenario-based walkthrough were noted for each workshop. Items were also sorted into categories depending on content and categorised according to the level of design decision specificity. For the second type of data, the abstracted and synthesised statements from the data analysis in Paper IV were utilised, and structured to assess the desired goals of the method combination and the effects of the modifications made to the methods. In addition, the data was analysed to assess the usefulness of the method combination. The two types of data were then analysed together in relation to the goals, effects of modifications, and usefulness.

4.4.3 Key findings

Paper IV and Paper V were based on the same study, but analysis of the data was undertaken with different purposes in mind – resulting in two different kinds of outcome.

Paper IV

The participants' experiences from the three cases resulted in a list of guidelines for the development of evaluation methods suitable for assessing more general design decisions (higher design level) in control room systems in practice. 18 guidelines were identified, divided into three groups:

- Guidelines regarding the method's ability to provide support for *defining differences in circumstances*, *adapting* the method accordingly, *focusing* the evaluation effort on relevant aspects and help in *balancing* them against each other.
- Guidelines regarding execution of the evaluation workshops.
- Guidelines regarding the communicative purpose of the evaluation activity.

Comparing the guidelines identified in the present paper with more general guidelines for human factors and product development methods identified by other researchers showed both similarities and differences. One of the guidelines from literature stated that methods should “*require data that can feasibly be gathered*”. This was not a guideline that emerged from the analysis in Study D,

but it is an important one that should be added to the set of guidelines identified. The compiled results can be used to further develop methods suited for early formative evaluation of control room systems in practice.

Paper V

Analysis of the data from Study D for Paper V showed that the heuristic evaluation and scenario-based talkthrough methods can be used for early formative evaluation of nuclear power plant control room systems, and were found to be useful in practice. Combining the methods makes it possible to take advantage of the strengths of both methods. A combination of the two methods also allows a way to trade-off between efficiency and thoroughness in the evaluation by combining a search for typical design problems using guidelines with a use-focused approach to identify and locate problems not explicitly sought. The method combination could be further improved by 1) providing better support for adapting implementation of the methods to the development project in question and the control room system to be evaluated; and 2) by providing better support for practical execution of the evaluation activity. The feasibility of the preparations needed prior to the evaluation workshop also needs to be further tested.

4.5 STUDY E

Study E also addressed the second research question and constituted a second iterative loop in the development of the method combination tested in Study D.

4.5.1 Purpose

The method combination tested in Study D was modified to address identified weaknesses. A focus group study was executed with the purpose of assessing the usefulness and generalisability of the modified version of the method combination.

4.5.2 Modification of the method combination

The combination of heuristic evaluation and scenario-based talkthrough was modified based on the findings from Study D. To provide better support for adapting the method combination, the method description was supplemented with a common set of preparatory actions to tailor the methods to the development project and the design concept in question. More specifically, actions to define the purpose of the evaluation and the level of specificity of the design decisions to be assessed. The set of preparatory actions also stipulated actions to make the evaluation of large design proposals more manageable, intended to improve the practical execution of the workshops. The practical execution of the evaluation activities was further detailed by adding that the purpose of the evaluation workshop and the roles of all participants should be made clear to everyone at the start of the workshop. The method description also provided templates for documentation and support for the moderator during the workshop. The

categories for sorting identified discrepancies according to severity were updated. In order to further support adapting the scenario-based walkthrough, the method description was supplemented with a guide for scenario development as well as for formulating discussion questions.

4.5.3 Method

The method combination was presented and discussed in three focus groups. The participants in the first focus group were human factors specialists who had used the previous version of the method combination in the case studies in Study D. The participants in the second and third focus groups were human factors specialists with experience of control room development from other domains than nuclear power (maritime, train dispatch, process industry, telecom network operations centres, command centres, control centre for particle accelerators and detectors). Each focus group had three participants and was moderated by the researcher. In total, nine persons participated in the focus groups.

The participants were sent a description of the method combination to read before the focus group. Each focus group was 2-2.5 hours long and was audio- and video-recorded. The focus groups were structured and started with questions about formative evaluation in general and in the participants' domains in particular. The researcher presented the modified version of the method combination and the participants were given the opportunity to ask questions if anything was unclear. The participants were asked to individually write down what they liked best and least with the method combination. Their answers were then presented to and discussed in the group. They were asked if they believed the combination would be usable in their domain and if they thought it needed to be modified in any way. The focus group ended with a discussion about specific details of the method combination that had been questioned during analysis of the data from Study D.

The contents of the audio recordings were summarised per question for each participant. The summary of each focus group was sent to the groups' participants to give them the opportunity to read and comment on the content. In the analysis of the data, positive comments about the method combination were compiled and synthesised, as well as suggestions for improvement (based on negative comments). Answers regarding the participants' view of the usefulness of the method combination (for example if they would use it in the future and if they saw barriers hindering its use) were also compiled and synthesised.

4.5.4 Key findings

The participants in the focus groups were overall positive towards the method combination. The nuclear power human factors specialists (who had used the previous version of the method combination) were all willing to use the methods combination in future projects. The human factors specialists from other domains believed that the method combination would be possible to use in their domains

as well, but pointed out barriers that could make use of the method combination difficult. These barriers were: getting access to users, lack of human factors maturity in the organisation, and problems with obtaining a suitable system representation. These barriers related more to the organisational context of method use than the design of the method combination itself.

When asked what they liked most about the method combination, six of the participants mentioned how the methods complement one other, that they cover both detailed parts and the whole. Five participants noted that the detailed structure of the method combination was a positive aspect that would support execution, for example having all templates and guides compiled together. Other positive comments by individual participants were that it is beneficial to be made to define and communicate the purpose of the evaluation, that the probe questions were good for assessing the solution as a whole, that the methods capture positive aspects in the design as well as discrepancies, that workshop participants are asked to write down their comments so as to not disturb the scenario talkthrough, and that the methods can be used together as well as separately. All participants were also in favour of the templates given for documentation and moderator support.

The participants' negative comments regarding the method combination gave much input regarding how the methods could be further improved. More specifically, their comments indicate that the method combination should be improved in the following ways:

- In one of the focus groups, there was a discussion as to whether it would be better to perform the scenario-based talkthrough prior to heuristic evaluation. The scenario-based talkthrough is more suitable for assessing the new design as a whole, and knowledge from this workshop could then be used to focus the scope of the heuristic evaluation, thus making it more efficient. The method description should discuss the importance of an iterative evaluation process and allow flexibility in the order in which the methods are used in a project.
- Identifying discrepancies and finding solutions to those discrepancies should be done separately in the workshops, since starting to discuss solutions could hamper the identification of discrepancies. This could be addressed by encouraging the participants to write down their suggestions for resolutions during the identification of discrepancies and new requirements instead of vocalising them directly. Solutions could then be discussed in the group after discrepancies have been identified.
- Prioritising of identified discrepancies should be based on the consequences of not resolving them; discussion on the effort required to do so is better left to other forums in a project. The method combination should also handle the aspect that several discrepancies that individually have smaller consequences could have large consequences when combined.

- The template for moderator support should be made such that there is greater clarity about what needs to be adapted for it to suit a specific evaluation workshop.
- Positive aspects should be identified in the heuristic evaluation as well (not only in the scenario-based talkthrough) since this lessens the risk of them being altered or removed later in the project. Being asked to express positive aspects could also make the participants more comfortable in expressing negative aspects (giving a feeling of balance), as well increasing user acceptance of the new solution. In both workshops, this should be done continuously but also with a summarising discussion at the end.
- The heuristic evaluation needs to be even better adapted to design concepts that are large in scope. One suggested solution from the focus groups was to supply a checklist of type scenarios or operational modes to guide a more structured assessment of guidelines.
- Decision makers (for example project leaders) should be included in the heuristic evaluation as well. This would allow them to better understand the reasoning behind identified discrepancies and new requirements.
- More support for scenario creation could be given. The version of the method combination presented to the participants from the nuclear power domain contained a reference to a list of guidelines for scenario content in NUREG-0711 (US NRC, 2012), but support for other domains too would be beneficial.
- The written description of the method combination needs to be improved and made clearer.

The method combination was modified in line with the input from the focus groups. The only exception was that support for defining scenario content was only given for the nuclear power domain. The modified version of the method combination is presented in Appendix A.

ANALYSIS

This chapter presents the aggregated analysis of the findings from the presented studies to answer the research questions. It also presents the additional insights that emerged from this analysis in the shape of five perspectives that can guide evaluation planning and method development.

5.1 RQ1 - WHAT TO EVALUATE?

The first research question this thesis sought to answer was:

RQ1: What must be evaluated, from a human factors perspective, to assess a nuclear power plant control room system's ability to fulfil its intended purpose?

This research question was explored from an evaluation planning and method development point of view, so the focus was not to create a model of how the control room system works, but rather to investigate the 'probes' needed to satisfactorily determine if the proposed design is likely to have the intended effect. Control room systems and development projects are very diverse, and specifying specific evaluation measures that are valid for all variations is simply not possible (cf. Baber, 2005, for a similar argumentation on the topic of measuring usability in human-computer interaction). Specific measures must be operationalised from the intended purpose of each unique control room system and development project, based on knowledge of the control room system and how it works. The approach taken in this thesis was to explore characteristics of measures that could be used to guide the choice of measures and evaluation methods.

Since safety is the primary rationale for evaluation of a safety-critical system, Study A was an interview study to investigate those aspects of the nuclear power plant control room system that contribute to safe operation. The identified themes were intended as a foundation for evaluation measures. However, the issue of measures for control room system evaluation has been addressed by other researchers too. In Study B (Paper II), measures targeted by existing human factors evaluation methods were grouped into categories (system performance, task performance, use of resources, user experience, and identification of design discrepancies). These categories of measures were compared with the themes from Study A, as well as with the control room system models and evaluation frameworks of US NRC (2012), Braarud and Rø Eitrheim (2013), and Savioja (2014). Study A (Paper I) and Braarud and Rø Eitrheim (2013) both explored what contributes to safe operation of a nuclear power plant from a control room system perspective,

but did not propose any specific measures. Comparison with the categories of measures showed that the themes from Study A and the model by Braarud and Rø Eitrheim (2013) corresponded with all categories. US NRC (2012) and Savioja (2014) both proposed specific measures, but the former did not include measures of teamwork and the latter did not include measures in the identification of design discrepancies category. The categories of measures identified in Study B (Paper II) were also compared with measures utilised in empirical control room system evaluations reported in scientific literature. This mapping with the categories was diverse, some evaluations used measures from all categories and some only used measures in one category. Of the seven reviewed empirical evaluations in Study B, measures in the task performance category were included in six. Measures in the teamwork, use of resources, and identification of design discrepancies categories were included in five of the reviewed evaluations. User experience measures were used in four evaluations, and system performance measures in three evaluations.

All the identified categories of measures were found to be relevant for nuclear power plant control room system evaluation. They complement each other and should all be included in evaluation during the course of the development process. They are complementary in two ways: 1) in terms of the characteristics of the control room system they target, and 2) in terms of their sensitivity in identifying potential discrepancies (design variables that are determined in inadequate ways) and new requirements (important design variables that are not yet determined) in the design.

With regard to the first way, the categories are complementary because the groups of characteristics they target are each important for a well-functioning control room system. For example, a control room system should support both teamwork and a satisfactory mental workload for individual operators. Satisfactory performance is important, but so too is the subjective experience of the users. The exception to this is the identification of design discrepancies category. This category of measures targets the design itself (in contrast with the others that target impact of the design), and problems identified using measures in this category are problems because they affect the impact. A design that does not conform to design guidelines is problematic if it (potentially) leads to undesirable consequences; otherwise the mismatch is not a discrepancy.

The second way in which the categories are complementary relates to their difference in sensitivity when identifying discrepancies and new requirements. System performance measures are closely related to the system goal and thus a clear warning signal if they are unsatisfactory, but they are not very sensitive. Task performance, teamwork, and use of resources measures are more sensitive. For example, even if a competent shift team manages to maintain satisfactory system performance the team's members might experience high mental workloads that could indicate problems in the design, which in turn would decrease performance

for a less competent shift team. Measures in the user experience category can be even more sensitive, as argued by Savioja et al. (2014). The advantage of the identification of design discrepancies category is that it provides more input into how the design should be changed.

When planning an evaluation or developing an evaluation method, the categories can be used to guide the choice of measures. Specific evaluation measures must be operationalised for the specific control room system to be evaluated. However, since the categories represent groups of important characteristics they can help this operationalisation by steering towards a more diverse set of measures. Using the categories can highlight whether a set of measures is skewed in the characteristics they target. The categories can be used in the same way to select a set of measures that differ in sensitivity to increase the likelihood of identifying important discrepancies or new requirements. Also, by guiding the choice of measures, the categories guide the choice of evaluation methods.

5.2 RQ2 - HOW TO SUPPORT CONTROL ROOM SYSTEM DEVELOPMENT?

The second research question this thesis sought to answer was:

RQ2: Can human factors evaluation better support nuclear power plant control room system development? If so, how?

5.2.1 Identifying a gap

The first step taken in exploring this research question was to study current evaluation practices in the nuclear power domain. Study C (Paper III) explored utilised evaluation approaches in the nuclear power domain and compared the level of design decision specificity evaluated. The conclusion of this study was that formative evaluation approaches for assessing more general design decisions are less common and not described in as much detail as summative evaluations where more specific design decisions are evaluated as well. Formative evaluation of higher-level design decisions lessen the probability of late and expensive changes, as well as of producing control room designs that do not fulfil the intended purpose in an optimal way. Further detailing and improving of formative evaluation approaches for higher-level design decisions could thus allow human factors evaluation to better support control room system development.

5.2.2 Understanding the purpose of formative evaluation

In order to understand how formative evaluation may be developed to better support control room system development, the purpose of formative evaluation needs to be better understood. The primary purpose of a formative evaluation is to provide input to design. The nature of the input may however be further nuanced. One positive aspect of the method combination noted by the participants in the case studies (Study D, Paper V) was that the methods identified *both* discrepancies

and new requirements (or information that could be used to formulate new requirements). Acknowledging that input to design can be in the form of both identified discrepancies and new requirements forces an active decision on what input to design a specific evaluation method should deliver, so method development or modification can be more precise. Distinguishing between these two types of input is also useful in planning the execution of human factors activities in a development project. For example, knowing that a formative evaluation can serve as a data collection activity and identify new requirements will highlight its usefulness in earlier stages in a development project.

The fact that an evaluation may serve multiple purposes and should be able to be adapted to suit those purposes was one of the method guidelines for use in practice identified in Paper IV. Acknowledging complementary purposes of formative evaluation strengthens the argument for why the activity should be included in a development project. In Studies D and E, the interviews with the participants indicated another purpose than provision of input to design that the evaluation activity needed to play. The evaluation served a communicative purpose as well. If users are included in the formative evaluation activity, this can aid user acceptance of the new design. If project members are included in the evaluation, such as project leaders or designers, information on the use of a new design and the way this affects the design's ability to fulfil its purpose can be transferred more efficiently.

Another complementary purpose a formative evaluation activity may serve that was not expressed in Studies D and E, but which has been identified by other researchers, is supporting summative evaluation. When describing their stepwise approach to integrated system validation, Laarni et al. (2014) pointed out that test activities earlier in the development process can be used to guide integrated system validation efforts. The example they give is that more emphasis and effort can be placed on testing those design solutions shown to be less mature in earlier tests. Using formative evaluation methods like the scenario-based walkthrough could guide integrated system validation in a similar way by providing knowledge of a suitable selection of scenarios.

By acknowledging that a formative evaluation can identify design discrepancies and new requirements, serve a communicative purpose and support summative evaluation the evaluation activity can be better tailored to these purposes. By tailoring the evaluation activity, resources spent on formative evaluation can be more efficiently utilised. For example, executing an evaluation activity earlier during a project can mean that new requirements are identified while there is still time to adhere to them. Another example is sufficiently documenting an evaluation activity that will later serve as input to summative evaluation so it can be used when reporting to an external reviewer (such as a governmental authority).

5.2.3 Usefulness in practice

In order to be able to impact the actual design of control room systems, evaluation methods must also be useful in practice. Paper IV, using the interview data from Study D, identified guidelines for making human factors evaluation methods useful in practice. This set of guidelines can be used to steer evaluation method development to make methods more useful, thus having a greater chance of making an impact on control room development.

One subset of the guidelines presented in Paper IV relates to the adaptability of evaluation methods. This was identified in the analysis of the data from Study D as well as in one of the sets of guidelines found in literature, Andersson and Osvalder (2015), who stated that methods must be tweakable to fit the working context. Shorrock and Williams (2016) stated that even though work situations are complex, if methods are to be usable and useful it is important for practitioners that methods are only as complicated as necessary with regard to purpose. At the root of this need for adaptability of methods lies the trade-off between efficiency and thoroughness that is needed when evaluating a complex system when resources for doing so are finite (cf. Hollnagel, 2009). Developing or modifying a method that is useful in practice requires making this trade-off in a way that is satisfactory for the specific case.

5.2.4 A method combination for early formative evaluation

Addressing the gap identified in Study C (Paper III) highlights demands on evaluation methods. Assessing higher-level design decisions when they are taken, typically in the earlier phases of a development project, requires methods that are not dependent on high-fidelity system representations. This limits the range of appropriate methods. Methods that are dependent on use of the system representation being very similar to use of the finished system are not suitable for evaluation of higher-level design decisions. This limitation in available methods also limits which of the categories of measures from Study B (Paper II) it is possible to target. No methods were found that could adequately target measures in the categories of system performance and use of resources in early evaluation. The prerequisites for, and selection of, methods suitable for addressing the gap identified in Study C (Paper III) are further described in Paper V. The chosen method combination, heuristic evaluation and a scenario-based walkthrough, was found usable in case studies and focus groups (Studies D and E). The two methods complemented each other, both through the categories of measures covered and through the type of search for problems undertaken. The method combination merged two types of searches. The first type was a search for known typical design problems using guidelines (heuristic evaluation). The second type was a use-focused approach (scenario-based walkthrough) that identifies a mismatch between desired impact and predicted impact and searches for probable reasons for this mismatch in the proposed design. Paper V proposed that combining these

two types of searches allows a trade-off between thoroughness and efficiency when executing an evaluation. Section 5.3.3 elaborates on this topic. The resulting description of the method combination and the procedure for its use (Appendix A) provides human factors specialists with practical guidance for formative control room system assessment.

5.2.5 Concluding remarks on the analysis of RQ2

To conclude, the answer to the first half of the second research question is that nuclear power plant control room system development can be improved through better human factors evaluation practices. Study C (Paper III) indicated that there is room for improvement of formative evaluation practices in the nuclear power domain. Advancing the practice of formative evaluation can increase the likelihood of producing control room system designs that fulfil their intended purpose in a more optimal way, and decrease the likelihood of having to make late and expensive changes in design.

To answer the second half of the second research question, the work presented in this thesis identified a number of ways in which advancement of formative evaluation practices can be brought about. Nuancing the purpose of formative evaluation highlights the advantages the activity brings to the development process. Nuancing the purpose also allows resources spent on the evaluation activity to be used in a more purposeful way. Developing methods that are useful in practice is another way to advance formative evaluation practices, and Paper IV presented a number of guidelines that may be used for this purpose in method development. Achieving a suitable trade-off between thoroughness and efficiency is especially important for methods to be useful in practice. A combination of methods, heuristic evaluation and scenario-based talkthrough, was shown to be useful in practice for nuclear power plant control room system evaluation when tested in industry cases. The description of the method combination presented in Appendix A can be used as a concrete guide for human factors specialists.

5.3 PERSPECTIVES TO GUIDE EVALUATION

Through the research questions the evaluation activity and the methods used were considered from two different angles: 1) how the control room system design and its anticipated performance are regarded by the evaluation activity (RQ1); and 2) the relation between the evaluation activity and the development process as a whole (RQ2). The following section presents insights that emerged when the findings from the studies were analysed to answer the research questions. The knowledge gained is compiled and presented here in the form of five perspectives to consider in evaluation planning and method development: 1) the purpose of the evaluation activity, 2) the object to be evaluated, 3) the tactic used in the evaluation activity, 4) the evaluation procedure, and 5) the use of the evaluation method. These perspectives provide decision support when planning evaluations

or developing evaluation methods by highlighting decisions that must be made. The perspectives are based on knowledge gained from studies focused on formative evaluation. While primarily tailored to, and discussed in relation to, formative evaluation, the perspectives can provide guidance for summative evaluation as well.

5.3.1 The purpose of the evaluation activity

Choosing *which* object to study, for example a control room system, should be obvious when planning an evaluation. However, deciding *why* the evaluation is to take place may not be consciously defined. The first perspective relates to the purpose of the evaluation activity, the reasons why the activity is undertaken (cf. the roles by Scriven, 1967). This will impose requirements on the evaluation method used as well as on its implementation. For example, for an evaluation with a formative purpose the method will need to be able to provide knowledge about how the design can be further improved, not only whether or not it conforms to acceptance criteria. A summative evaluation might have stricter requirements on the objectivity of the persons involved if the results are to be used as proof of quality to an external part (such as a governmental authority). Acknowledging a nuanced image of the nature of the design input of formative evaluation (identified discrepancies and new requirements) as well as the complementary purposes of formative evaluation (communicative and supporting summative evaluation) highlights when in a development project the activity may be of use as well as its benefits.

5.3.2 The object to be evaluated

The second perspective is the object to be evaluated. Not to define *which* object to evaluate, but to decide: 1) the measures needed to assess the design and 2) the design level to be evaluated. In an evaluation, the object to be evaluated is viewed and assessed through the use of evaluation measures. The evaluation measures are the probes that gather the data needed to answer the question the method is meant to answer. By operationalising the question suitable measures can be found. When suitable measures have been found a method for targeting those measures can be selected (or developed). For nuclear power plant control room systems, the categories of measures presented in Paper II may guide the choice of measures. In addition to the categories of measures presented in this thesis, there are of course other dimensions to consider to arrive at a sufficiently diverse set of measures. To consider whether, for example, the data collected need to be subjective or objective (IEEE, 1999), or quantitative or qualitative (Kovesdi et al., 2018), is also important.

Another aspect of the object to be evaluated that needs to be considered is the level of design decision specificity to be assessed. The system representation used in an evaluation should be matched to the level of design decision specificity to

be evaluated. When using a system representation that is too detailed in relation to the design decisions to be evaluated, there is a risk that the assessment may focus on details that are not of interest at that moment in time. Using a system representation that has too little detail in relation to the design decisions to be evaluated provides no support for the assessment since the aspects to be evaluated are simply not presented. Knowledge of the design levels that it is crucial to assess in different phases of a project can thus guide evaluation planning in terms of resources. Knowing the relevant level of design decision specificity decreases the risk of creating a system representation that is overly detailed (and often more expensive).

Design decision specificity will also restrict method choice (or method design) through the fidelity of the system representation. There are many different dimensions of fidelity to consider when developing system representations for evaluation: breadth of features, degree of functionality, similarity of interaction, and aesthetic refinement (Virzi et al., 1996). A system representation that compromises one or more of these dimensions in a way that is obvious to the user is considered low-fidelity according to Virzi et al. (1996). A system representation of higher-level design decisions will be of lower fidelity, and vice versa. The fidelity of the system representation will affect use of the system representation. Many aspects of the use of a low-fidelity system representation will be less similar to the use of the implemented system than a high-fidelity system representation. For example, the time it takes to simulate a task in a paper mock-up will not be an accurate representation of the time it would take to perform the same task in the implemented system. System representations of low fidelity are therefore less suitable for evaluation methods that rely on simulated use closely resembling actual use, such as usability tests. This restriction makes it necessary to consider design decision specificity in evaluation planning and method development.

5.3.3 The tactic used in the evaluation activity

The third perspective relates to the tactic used in the evaluation activity. As discussed in Paper V, an evaluation activity may 1) seek the existence and location of known typical design problems (unknown knowns) and 2) seek to identify and locate unknown problems (unknown unknowns). The word 'problem' here denotes both discrepancies in the proposed design and new requirements (or knowledge that can be transformed into new requirements). The first type of tactic utilises prior knowledge in the form of design guidelines. Design guidelines are knowledge about successful design resolutions presented as design advice (either about how to do something or how not to do something). An evaluation that uses this tactic focuses on the design itself. However, an evaluation using guidelines will most likely not identify problems not covered by the design guidelines. An evaluation that focuses on the impact of the system to be evaluated may, however, identify problems not explicitly sought. This evaluation uses the other type of tactic, seeking to identify and locate unknown problems. Impacts of a design are

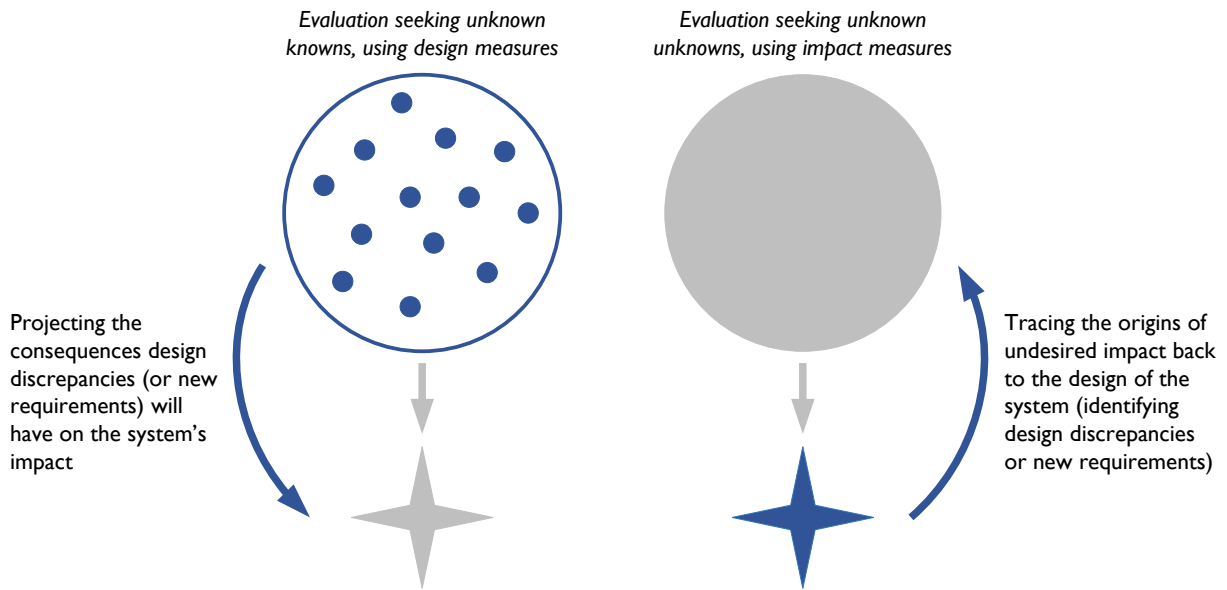


Figure 4: Illustration of the two types of evaluation tactic. The circle represent the system that is evaluated, and the "star" represents the impact of that system. Blue colour illustrates the primary focus in each type of evaluation, either on design or impact.

defined broadly here, and can be measured through any kind of output from the system. The categories of measures from Study B (Paper II) can be used here. Measures in the categories of system performance, task performance, teamwork, use of resources, and user experience can be used to measure impact, and will here be called impact measures. Measures belonging to the identification of design discrepancies category focus on the design and are here called design measures. Figure 4 illustrates these two types of tactics in evaluation.

When the purpose of the evaluation activity is to improve the design of the system (formative evaluation), projecting consequences and tracing origins is important in order to be able to decide which design changes are needed. When a design discrepancy has been identified through an evaluation using design measures (seeking unknown knowns), the consequences this discrepancy may have on the system's impact are projected to decide if a change in the system's design is needed. This is necessary because while design guidelines are based on knowledge about how good design resolutions should be undertaken, they must always be reviewed in the context of the system to be evaluated. Design guidelines can also be conflicting for a specific system, and decisions must relate to which guidelines to follow in each specific case. To decide what needs to be changed in an evaluation using impact measures (seeking unknown unknowns), possible causes of undesired impact are traced back to the proposed design. Figure 4 illustrates this tracing and projecting in the two types of evaluation tactics. An evaluation using design measures assesses discrete parts of the system (represented by blue dots in Figure 4). When the origin of undesired impact is traced back to the design of the system in an evaluation using impact measures, these may be sought in the design as a whole.

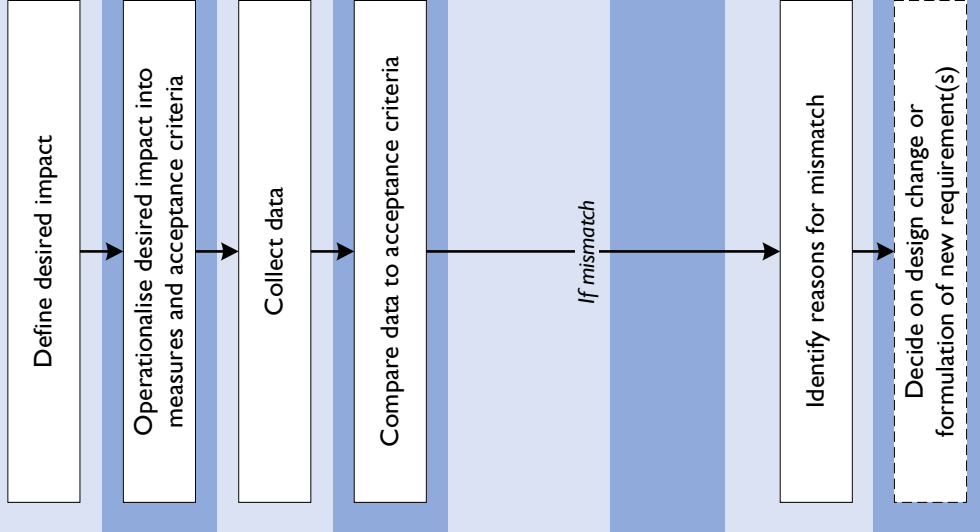
In theory, an ideal evaluation should identify all problems in a design. An evaluation with the goal of identifying and locating all unknown unknowns has to be very thorough, since the design as a whole must be taken into account. For a large and/or complex socio-technical system, this is an activity that will be very resource intensive (for instance assessing all possible use scenarios). In practice, this approach is not fully feasible. One way to make the evaluation more efficient is to utilise prior knowledge (design guidelines) and seek known unknowns. However, with this approach thoroughness is sacrificed since problems not covered by the design guidelines will most likely not be identified. With a combination of methods that targets both known unknowns and unknown unknowns a suitable trade-off between efficiency and thoroughness in the evaluation approach may be found.

5.3.4 The evaluation procedure

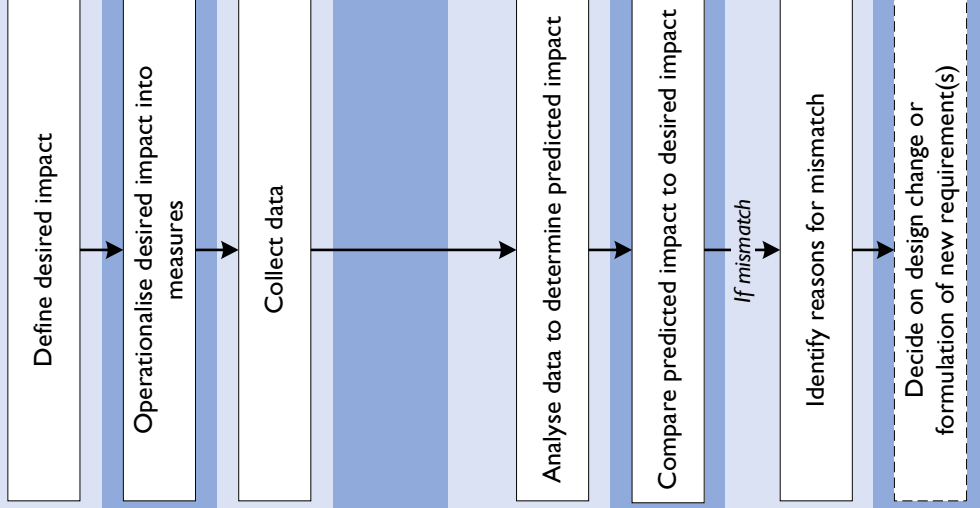
Knowing the tactic to be used in an evaluation, the evaluation procedure can be further detailed, which is the fourth perspective to take into account in evaluation planning or method development. There are several steps involved in performing an evaluation as part of a development process (Figure 5). Traditionally, the starting point of an evaluation is described as defining the desired impact of the design to be evaluated and operationalising that impact into measures and acceptance criteria (Type A in Figure 5). However, as stated in IEEE (1999), acceptance criteria can be formal or informal. If acceptance criteria are informal, no explicit breakdown of the desired impact into acceptance criteria is undertaken (Type B in Figure 5). The above is valid if the measures taken are impact measures. However, an evaluation may also use design measures (Type C in Figure 5). In this kind of evaluation the selection of guidelines for the evaluation can be viewed as corresponding to operationalisation of the desired impact into measures and guidelines. The definition of the desired impact is also often implicit.

In an evaluation using impact measures (Types A and B) the next step is to collect data. In an evaluation that uses design measures (Type C) this step is interlinked with the step that compares the design to the guidelines (data is “collected” by reviewing the design and making the comparison). A core step in evaluation is the comparison between the desired impact and the predicted impact of the design that is being evaluated. In an evaluation with formal acceptance criteria (Type A) the collected data is compared with the acceptance criteria. In an evaluation without formal acceptance criteria (Type B), the collected data must be analysed to determine the predicted impact of the design that is being evaluated before the comparison can be undertaken. If a mismatch between guidelines and the design is identified in an evaluation using design measures (Type C), the predicted impact of that mismatch must be determined to allow comparison with the desired impact.

Type A
Evaluation using impact measures **with** formal acceptance criteria



Type B
Evaluation using impact measures **without** formal acceptance criteria



Type C
Evaluation using design measures

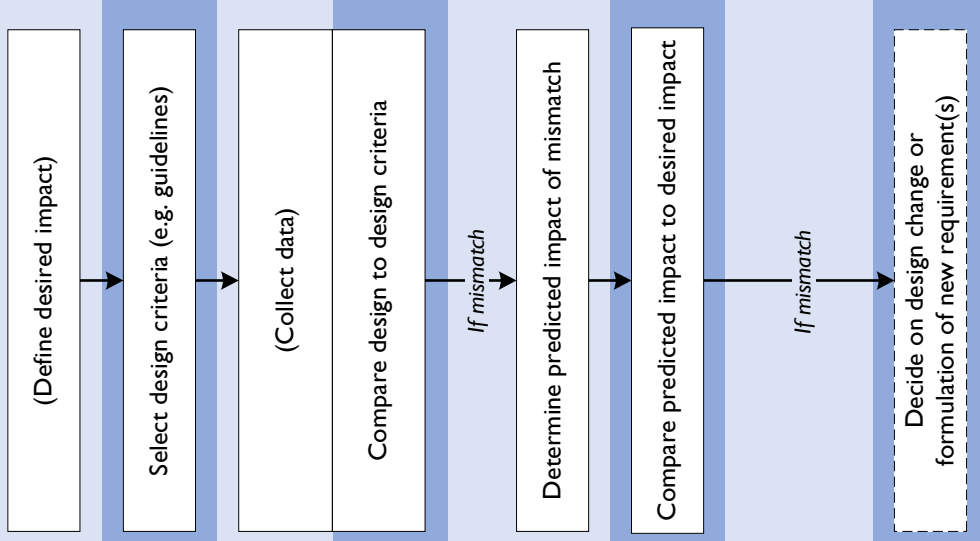


Figure 5: Different types of evaluation procedures – **Type A** with output measures with formal acceptance criteria; **Type B** with output measures without formal acceptance criteria; and **Type C** with design measures.

If there is a mismatch between the predicted impact and the desired impact in an evaluation using impact measures (Types A and B), the reasons for this mismatch may be sought. If the evaluation only has the purpose of determining if a design is good enough (summative evaluation), this step may not be necessary. Typically, however, information on *why* the design is not good enough is necessary in a development project. In an evaluation using design measures (Type C) this knowledge is a logical fallout of the comparison with guidelines. Deciding on design change or formulating a new requirement is strictly speaking not a part of evaluation, but an evaluation may guide the way the proposed design can be changed to have the desired impact, for example by supporting prioritisation of identified discrepancies and new requirements, or by identifying solutions.

An evaluation method may acknowledge and/or support all these steps, or some of them, and to varying degrees. In evaluation planning, choosing a specific method means that certain steps are determined by the method (such as the data to be collected), and the way of executing other steps is not, thus requiring conscious decisions for that specific evaluation activity. When developing a method, taking into account the steps in the evaluation procedure will highlight the actions the method should support.

The heuristic evaluation in the method combination tested in this thesis is a typical example of an evaluation using design measures (Type C). However, the scenario-based walkthrough, can be both an evaluation using impact measures without formal acceptance criteria (Type B) and an evaluation using design measures (Type C). If, during a scenario walkthrough, a potential use error is identified, this is an example of an impact measure. If, for example during a discussion of the probe questions, a mismatch between the design and an ideal design is identified, the evaluation uses design measures. The method combination in this thesis acknowledges all the steps presented above, but does not support all of them to the same extent. Data collection is supported (through a detailed description of workshop execution), but definition of desired impact and selection of guidelines is merely acknowledged. In the scenario-based walkthrough, operationalising desired impact into measures is to some extent given by the method (that use of the new design should be talked through), but the specific scenarios to be used must be specified for each design to be evaluated. Other steps are not explicit in the method description, but will be undertaken as a logical consequence of prescribed actions. One such example is analysis of data to determine predicted impact and comparison with desired impact. In a scenario-based walkthrough, the prescribed action to note discrepancies and new requirements forces the participants to analyse whether the design will support a use that will have the desired impact. Decisions on design change are supported by the method combination through the stipulation that suggestions for solutions should be noted if they are given, and through prioritisation of the identified discrepancies.

5.3.5 The use of the evaluation method

The fifth perspective that should be considered is the use of the evaluation method. A method can target the right measures, and have high validity and reliability, but if it is not properly adapted to the use, user and context of use there is a risk that it may be used in the wrong way or not at all. Only an evaluation method that is actually used will have an opportunity to impact the design, and a method that is used in the wrong way might not have the desired impact. A use-centred approach should be utilised when developing methods (Andersson et al., 2011). It is vital to take into account the use, user, and context of use of the specific method that is developed or modified so as to guide method design. In addition, in order to support design prior knowledge of good method design should also be utilised, such as guidelines. For example, Norell (1992), Andersson and Osvalder (2015), and Shorrocks and Williams (2016) have all written about characteristics of useful methods. Paper IV identified a number of guidelines for developing methods for early formative evaluation of control room systems. These guidelines point out the importance of a method being able to provide support for defining differences in circumstances, adapting the method accordingly, focusing the evaluation effort on relevant aspects and helping in balancing them against each other. The guidelines also took into consideration how methods should be designed with regard to execution of the evaluation workshop and the communicative purpose of the evaluation activity.

5.3.6 Using the perspectives

Using the perspectives will make decisions that are necessary in evaluation planning and method development more visible and thus enable more conscious choices. Using the perspectives will also deepen understanding of the evaluation activity, thus making it easier to argue for the benefits of a chosen evaluation plan or method design to different stakeholders.

Consciously reflecting on the purpose of an evaluation will make it possible to more closely tailor the evaluation activity to that purpose, for example when in the development process it should take place and the participants who should be included. More specifically, the nuances of the purpose of formative evaluation presented in this thesis can be used to highlight the advantages of performing formative evaluation for different stakeholders.

Considering the object perspective by using the categories of measures (and other dimensions, such as if the data collected should be subjective or objective) will improve the evaluation by guiding towards a set of diverse measures. Knowing the level of design decision specificity to be assessed in a specific evaluation will highlight restrictions in method choice, narrowing the scope of viable methods.

Knowing the design level of interest will also make it possible to tailor parts of the evaluation activity, such as the system representation, to ensure it delivers what is needed at that specific moment in the project timeline.

The tactic perspective helps in understanding two different starting points an evaluation activity may have – the object itself or its impact. The tactic perspective show how an evaluation starts by taking measures of either design or impact and traces/projects influence from one/on the other, as well as the advantages and disadvantages of both tactics. This thesis recommends a combination of both tactics, but considering the tactic perspective in evaluation planning or method development allows a conscious choice of what is lost if one of the two tactics is excluded.

The procedure perspective helps in understanding the steps that must be supported by an evaluation method or, in the case of evaluation planning, knowing which steps are not supported by the chosen method and must be more consciously planned.

The use perspective highlights the importance of tailoring the evaluation method to its use, users, and context of use. Like an artefact, a method will have an impact only through use, and successful evaluation planning and method development cannot disregard this perspective.

DISCUSSION

This chapter reflects on the research approach used in this thesis, relates the findings to previous research, and discusses the practical implications of the findings. The nuclear power plant control room system is a safety-critical socio-technical system with partly unpredictable behaviour, and this chapter also discusses how the findings relate to evaluation of such a system.

6.1 ADVANCING EVALUATION PRACTICES

The purpose of this thesis was to enhance the understanding of human factors evaluation of nuclear power plant control room systems in order to advance evaluation practices as part of the development process. The gap in today's evaluation practices in the nuclear power domain that steered large parts of the direction of this thesis was that evaluation is often executed when lower-level design decisions have been made, typically towards the end of the development process (Study C). The present thesis thus focused on evaluation of higher-level design decisions, preferably undertaken early in the development process. Furthermore, if assessment is executed early in the development process it is beneficial if it has a formative purpose, so this was an additional focal area.

An evaluation activity must be tailored to the control room system to be evaluated and to the development project it is part of. This tailoring is undertaken when planning the evaluation. To support this adaptation is one way to advance evaluation practices as part of the development process. Another way to improve evaluation is through the methods used, since methods that are better suited to their task have a better chance of resulting in the desired impact. This thesis identified five perspectives that can be used as decision support in evaluation planning and method development: 1) the purpose of the evaluation activity, 2) the object to be evaluated, 3) the tactic used in the evaluation activity, 4) the evaluation procedure, and 5) the use of the evaluation method. While primarily developed with formative evaluation in mind, the perspectives are helpful for evaluation practices in general. Within these perspectives, the categories of measures (Paper II) and guidelines for evaluation methods that are useful in practice (Paper IV) provide more specific tools for method choice and development. The description of the combination of the heuristic evaluation and scenario-based walkthrough methods (Appendix A) provides concrete guidance for human factors practitioners when performing formative control room system evaluation, in particular early in the development process.

6.2 REFLECTIONS ON THE RESEARCH APPROACH

The empirical studies presented in this thesis were executed within the Swedish nuclear power domain. Study A was an interview study with Swedish practitioners, but in Study B this was used together with other studies that balanced the geographical limitation. The mapping to categories of measures showed no difference between the themes from Study A and the other studies. Studies D and E were also limited to Sweden. The resulting guidelines from Study D were mapped to guidelines from other researchers, but only one of these sources was non-Swedish. A study of whether there is a difference in what makes evaluation methods useful in practice in other cultures would be an interesting topic for further research. The combination of the heuristic evaluation and scenario-based walkthrough methods was in this thesis only empirically tested in Sweden. However, nuclear power plant control room systems are very similar around the world, and the method combination has many similarities with evaluation approaches developed and utilised outside the Swedish nuclear power domain, such as CRIOP (Johnsen et al., 2011), developed for the Norwegian oil and gas industry, and verification and validation in NUREG -0711 (US NRC, 2012). It is thus deemed very unlikely that the usefulness of the method combination should be limited to the Swedish nuclear power domain.

As was pointed out in Paper III, the literature study was limited to evaluations reported in academic literature. It is thus possible that the analysis does not accurately show how evaluation is performed in industry; formative evaluation may be more prevalent and mature in industry than was shown in Study C. However, the findings in Study C correspond with the author's own experience of the nuclear power domain. In addition, the value of knowledge about how to execute formative evaluations presented in this thesis does not lie solely in its novelty. It also holds value in terms of being available for others to learn from. Shorrock and Williams (2016) emphasised that accessibility is an important factor if human factors methods are to have an actual impact in the world.

Case studies were chosen as a research method in Study D since it investigates a phenomenon in its real-life context (Yin, 2014), which was deemed the best way to study the usefulness of methods in practice. A necessary trade-off with this approach was that comparisons between evaluation workshops and cases were difficult since participants and design concepts differed between workshops and cases. Because of this, the analysis of causes behind findings could only be viewed as indicative. Initially, the ambition was that Study E should be another case study to test the modified version of the method combination. However, due to difficulties in finding suitable cases in industry within the necessary time schedule, a focus group study was chosen instead. The focus group included human factors specialists from both nuclear power and other domains. The results of Study E

showed that the specialists believed that the modified version of the method combination would be useful. However, this is something that should be tested further in studies where the method combination is used in practice.

The author of this thesis had several years of experience as working as a human factors specialist in the Swedish nuclear power domain before entering the world of academia. The population of human factors professionals in the Swedish nuclear power domain is limited, and most human factors specialists are known to each other. Some of the participants in the studies in this thesis were thus persons who were either former colleagues or persons the author had previously worked together with in development projects. This relationship may have had an impact on some participants' answers in interviews and questionnaires, mainly in Studies D and E where the questions related to materials produced by the author. However, no tendency toward more positive answers from these persons could be discerned compared with the responses by persons with whom the researcher had no previous relationship. The previous relationship between the author and some of the participants may also have had a positive impact in the sense that they were more relaxed and felt comfortable to speak freely in the interview situation.

The author's practical experience of the domain provided a preunderstanding of the topic that impacted the work presented in this thesis. A major benefit of being both a researcher and a human factors practitioner in the nuclear power domain was that it was easier to understand the subject matter experts who were interviewed, and it allowed the author to dig deeper into the explored topics. The author's connection to the domain also made it easier to gain access to a rather closed industry, for example by being allowed to observe evaluations of designs which for security reasons cannot be shown to anyone without the proper clearance. Preunderstanding of a subject may also include preconceived assumptions that are difficult to consciously question, which might leave possible paths of discovery unexplored. In conclusion, having a preunderstanding has advantages, but so has exploring a topic without having this preunderstanding. Both approaches are needed in research, and this thesis represents the former.

6.3 EVALUATION OF SYSTEMS WITH UNPREDICTABLE BEHAVIOUR

Evaluation of a proposed design is always about trying to predict future behaviour, no matter which system is evaluated. However, it may be more or less difficult to predict future behaviour for different systems. Evaluation of a nuclear power plant control room system must take into account three sources of unpredictability: the control room system itself, the environment in general and the control room system's ability to control a partly unpredictable system in particular. As was described in Section 1.1.1, the human agents in the system contribute largely, but

not solely, to the system's unpredictability. The inclusion of user representatives in evaluation (as is done in the scenario-based walkthrough) is one way to manage the issue of partly unpredictable behaviour in control room system evaluation. In an evaluation, the proposed design must be represented in some way. The technological subsystem is typically represented through drawings or mock-ups. The human agents of the personnel subsystem are more difficult to represent in such an artificial way. By including user representatives, the unpredictability aspect of human behaviour may be represented, and thus taken into account in the evaluation activity.

Heuristic evaluation can also to some extent evaluate a system with partly unpredictable behaviour. Design guidelines consist of existing knowledge about what constitutes good design formulated as design advice. Using design guidelines is a way to gain access to a larger knowledge base than can be covered through the experience of the participants in the evaluation. It would thus be possible to tailor the heuristic evaluation to assessment of a system with a behaviour that is partly unpredictable by using guidelines that support the design of such a system.

Another aspect of evaluation and unpredictable behaviour is raised by Savioja (2014), who argued that it is not always possible to understand in advance all relations between components of a system. Therefore, evaluation should be sufficiently broad in scope, to allow emergence of effects not envisioned in design. This will not make the evaluation able to identify all possible effects, but it will at least make it able to assess more of them. Making the scope sufficiently broad should be heeded in evaluation planning when selecting groups of users for participation, system representation, and scenarios.

A connection may also be made between unpredictability of a system and safety. As was described in Section 2.2, the need for the Safety-II perspective on safety stems from the realisation that the traditional approach to safety – trying to find causes behind failures, eliminating them or improving barriers – was not enough for systems whose behaviours are difficult to predict (Hollnagel, 2013). Patriarca et al. (2018, p. 79) describes resilience engineering as “*a paradigm for safety management that focuses on systems coping with complexity and balancing productivity with safety*”. Resilience engineering as a field is concerned with proactively managing risk for systems with inherent complex functioning and a corresponding need for variability in performance (Patriarca et al., 2018). To connect this to the content of this thesis, a parallel may be drawn to the concept of wicked problems (Rittel and Webber, 1973). One property of wicked problems is that not all repercussions of a proposed resolution to a wicked problem can be comprehensively evaluated. This is because resolutions to wicked problems, when implemented, generate consequences that span over time and space in ways that cannot be completely specified. Evaluation of a system with emergent properties faces the same problem. Design as a field has developed strategies to manage wicked problems. As described in Section 2.5, evaluation is an important part

of those strategies (Lawson and Dorst, 2009). For the same reasons a wicked problem cannot be comprehensively evaluated, all criteria that a resolution should fulfill cannot be comprehensively analysed before a resolution is suggested. When exhaustive analysis of a problem is not possible, one strategy for being able to move forward in the design process is to generate resolutions and assess how they can be improved. This evaluation may aim to be as comprehensive as possible with the available resources, but it will never identify *all* the ways the design can be improved. In design practice, this fact is accepted and iterative cycles of ‘analysis – synthesis – evaluation’ are used to make the proposed design as good as it can be under the given circumstances. Evaluation as part of an iterative design process (i.e. formative evaluation) is thus one way to tackle a system with emergent properties.

6.4 EVALUATION OF SAFETY-CRITICAL SYSTEMS

Evaluation helps guide the design of a system towards a resolution that has the desired impact. The implication of evaluation of a safety-critical system is that the consequences of failed or less successful assessment are more severe. Typically it means that undesired impact involves not only economical aspects, but also harm to humans and their environment. The advancement of evaluation practices should thus be more important for safety-critical systems.

The focus in this thesis has been evaluation of a safety-critical system – not how to assess safety. The support of safe operation is part of the purpose of the nuclear power plant control room system (the other part being to support operator well-being). An approach to evaluate a safety-critical system which through its performance can influence safety must thus allow assessment of the ability to support safe operation. The findings in this thesis presuppose the notion that it is not possible to prescribe a way of executing evaluation activities suitable for all control room systems, but rather that the evaluation activity must be adapted to each specific control room system and development project. For example, the precise measures needed to assess safe operation will be different for each control room system evaluation. Indications of what to measure may be given, for example in an industry standard, but the precise operationalisation must be done from case to case (cf. Baber, 2005, who uses similar argumentation for the measurement of usability in human-computer interaction). The categories of measures in Paper II provide this indication, but leave the detailed operationalisation to be done when planning a specific evaluation. Here it is worth noting that the indication of what to measure provided by the categories of measures presented in Paper II does not specifically target the ability to support safe operation. The categories of measures presented in Paper II were found relevant for assessing control room systems. This is not the same as saying that the categories assess the ability to support safe operation. Study A explored what people involved in the operation and design of nuclear power plant control room systems believed contribute to safe operation. The findings from this study were used, together with findings from other sources,

to identify categories of measures relevant for assessing control room systems. The models, frameworks, and empirical studies used in the comparison related to the control room system's ability to fulfil its intended purpose. Safe operation is part of this purpose, but not all the sources used in the comparison specifically took into account the ability to support safe operation.

Another aspect to consider when assessing safety-critical systems is the safety perspective used, Safety-I or Safety-II (see Section 2.2). The findings of this thesis provide support when adapting the evaluation activity to a specific control room system, but they also allow flexibility regarding the perspective of safety used when defining safe operation for that system. Using the example of defining measures again – if a safety-II perspective of safety is used when defining safe operation, specific measures can be operationalised in line with this perspective. The categories of measures in Paper II do not hinder this approach.

6.5 EVALUATION OF SOCIO-TECHNICAL SYSTEMS

Consideration of the control room system as a socio-technical system means acknowledging that the parts interact with each other and that changes to one element will cause ripple effects in the system as a whole (Hendrick and Kleiner, 2001). Thus, evaluation of a socio-technical system must take into account this interaction between parts and the system as a whole. In the method combination tested in this thesis, the focus on use in the scenario-based talkthrough is one way to assess the system as a whole. An impact measure such as task performance is affected by the design of the system as a whole – for example by the design of operator interfaces, control room layout, design of procedures, training, and work routines. Talking through how a scenario would be handled with a proposed design is a way to assess if work as done in the implemented system is likely to be similar to work as imagined. Potential mismatches may be traced back to any part of the control room system. Depending on the guidelines used, a heuristic evaluation may take account of a dimension in the system as a whole. Guidelines may, of course, be very specific and only relate to an isolated part of the system – but they may also state a principle that should be adhered to throughout the entire system. A guideline can also supply knowledge about how to handle known interactions in a beneficial way. However, an evaluation using design measures cannot assess interactions between elements as well as an evaluation using impact measures, and heuristic evaluation should not be the sole method used for assessing a socio-technical system.

Formative evaluation, especially of higher-level design decisions, is a way to avoid sub-optimisation. In a project with the purpose of changing the technological subsystem, it might be tempting to keep the existing personnel subsystem or the work system design constant, and tailor the changes in the technological subsystem to suit them. For example, changing organisational aspects such as the responsibilities of different roles in the shift team requires major changes

to training, work routines, and documentation - changes the organisation might be reluctant to implement. To avoid sub-optimisation, however, keeping parts of a socio-technical system constant should be done with careful consideration. Through formative evaluation the risk of sub-optimisation can be decreased. If necessary changes are identified when there is still time within the project to undertake them, they are more likely to be accepted.

6.6 PRACTICAL IMPLICATIONS

For developers of human factors evaluation methods, the outcomes of this thesis highlight decisions that have to be made when designing a method. Methods that are consciously tailored and useful in practice will be more likely to provide the desired impact in a development project. For a human factors specialist in industry, a better understanding of the evaluation activity makes it easier to implement in a development process. For example, understanding the purpose of an evaluation activity means understanding the impact that is desired from an evaluation activity, which is information necessary to time its execution in relation to other activities in the development process. Understanding the purpose also gives a better understanding of the requirements on that activity, such as more rigorous documentation demands.

This thesis has specifically focused on early evaluation. Studies D and E showed how evaluation with low-fidelity system representations can be undertaken for nuclear power plant control room systems. Having a concrete description of a method adapted to the domain and type of evaluation object in question, such as the one in Appendix A, could lessen resource demands for planning and executing assessments and help overcome barriers for early evaluation. More efficient use of resources is also possible when complementary purposes of an evaluation activity are acknowledged, which could strengthen the cost-benefit argument for why an early formative evaluation should be included in a development project.

One aspect of nuclear power plant control rooms that restricts evaluation is access to users. The population of users of a specific control room system is typically small, and since the plant often has to be in operation 24/7, limited in how available it is for additional activities apart from operating the plant. Methods to be used in this context should thus rely as sparingly as possible on active operators. For example, the method combination described in Appendix A uses persons with operational knowledge, such as ex-operators, rather than active operators in the heuristic evaluation. Combining the heuristic evaluation with the scenario-based talkthrough is also a way of lessening the demand for an extensive scenario-based talkthrough, and thus the need for access to active operators.

Even though a nuclear power plant control room system is a special evaluation object, it is not unique. Many traits are shared with control rooms in other domains, and Study E showed that human factors specialists in other domains thought

that the method combination would be viable in their domains as well. Detailed guidance connected to the perspectives, such as the categories of measures from Study B (Paper II), are connected to the nuclear power plant control room system as an evaluation object. The perspectives overall, however, are not dependent on the traits of the evaluation object or the nuclear power domain. They could thus be used for human factors method development and evaluation planning in other domains and for other evaluation objects.

6.7 RELATION TO PROPOSED EVALUATION APPROACHES

This thesis has studied the usefulness of evaluation approaches in practice. It is thus necessary to consider the outcome of this thesis in relation to current evaluation practices.

6.7.1 NUREG-0711

How to perform human factors evaluation in Swedish nuclear power plants is not stipulated by the Swedish Radiation Safety Authority, but in the general advice to SSMFS 2008:17 (Swedish Radiation Safety Authority, 2008b), referral is made to NUREG-0711 (US NRC, 2012) for examples of evaluation methodology. It therefore makes sense to compare the outcome of this thesis with the content of NUREG-0711.

The gap in today's evaluation practice that was identified in Study C (Paper III) indicated that formative evaluation is not as prevalent a practice as summative evaluation in today's nuclear power domain. At least not in terms of more publically documented activities that may be used to share experiences within the domain. The review criteria for so-called "HSI tests and evaluations" (formative evaluations) provided in NUREG-0711 contain aspects to consider when developing evaluation criteria, as well as documentation requirements. When taking the object-perspective in planning a control room evaluation, the aspects to consider when developing evaluation criteria presented in NUREG-0711 may be used to guide the choice of evaluation measures. If the combination of heuristic evaluation and scenario-based talkthrough is used as an evaluation approach, these aspects may to some degree be used to guide the choice of guidelines and scenarios. In addition, specifying the objective of evaluation activities is stipulated in this section of NUREG-0711, which agrees with the 'purpose of the evaluation activity'-perspective. In short, the review criteria regarding formative evaluation in NUREG-0711 may be fulfilled if the perspectives and method combination presented in this thesis are utilised for planning a control room system evaluation.

Apart from formative evaluation, NUREG-0711 also provides review criteria for summative evaluation. Review criteria for verification and validation, which from the description given is to be interpreted as summative evaluation, are elaborate in NUREG-0711. Much of the guidance given would be very useful when utilising

the method combination tested in Studies D and E, such as when sampling appropriate content for scenarios in the scenario-based walkthrough. Performing formative evaluations using the method combination would also provide synergy effects. Human factors guidelines and scenarios selected for the formative evaluation would be useful input to the summative evaluation as it is described in NUREG-0711. The review criteria of the verification and validation chapter of NUREG-0711 imply an elaborate methodology with little room for alternative paths. Thus the perspectives in this thesis are not very helpful when planning verification and validation according to NUREG-0711. The perspectives may, however, be used to better understand the methodology. Combining evaluation approaches that focus on impact measures with approaches that focus on design measures is customary in verification and validation, and NUREG-0711 may be used as a domain-specific example here. Two verification activities are described through the review criteria in NUREG-0711: task support verification and human factors engineering verification, as well as a validation activity, integrated system validation. The verification activities involve reviewing whether the design meets predefined criteria (derived from tasks identified in task analyses) and whether it fulfils human factors guidelines. These verification activities can be said to seek known unknowns using design measures (cf. the tactic perspective). The integrated system validation is meant to assess if the design supports safe operation of the plant through performance-based tests. Typically, this is achieved by letting operators handle scenarios in a full-scale simulator. Thus integrated system validation seeks to find unknown unknowns using impact measures (cf. the tactic perspective). To conclude, the findings of this thesis may support, and to some extent be supported by, formative evaluation according to NUREG-0711. For summative evaluation, the perspectives presented in this thesis may be used to better understand the methodology, but offer little additional support to the already elaborate review criteria.

6.7.2 Approaches for stepwise evaluation

As was discussed in Paper III, Laarni et al. (2014) presented a stepwise validation approach for nuclear power plant control rooms where sub-systems were validated successively before the final validation. Even though the paper focused on how this stepwise approach builds evidence for the final assessment of design acceptability, the possibility of improving the design based on information from the step-wise assessments was also acknowledged. Simulator testing was seen as a central task, but other methods was used as well, such as observation of training sessions, expert evaluation, interviews, HSI-oriented walkthroughs, questionnaires, and focus groups (Laarni et al., 2011). There are similarities between the method combination utilised in this thesis and parts of the approach presented by Laarni et al. (2011), in particular the reliance on expert (heuristic) evaluation, and walkthroughs/walkthroughs. Study D (Paper V) in this thesis showed that the method combination makes it possible to assess higher-level

design decisions before a simulator is available. Laarni et al. (2014) pointed out that this stepwise validation approach provides input for planning of final integrated system validation, such as highlighting parts of the design that should be especially emphasised. The method combination utilised in this thesis may be used in the same way.

Another evaluation approach for nuclear power plant control room evaluation described in Paper III is the one presented by Boring et al. (2015), where it is argued that a series of formative evaluations will provide more complete evidence of the safety of a new control room system than a single summative evaluation does (see also Boring and Lau, 2017; Boring, 2017; Kovesdi et al., 2018). Their Guideline for Operational Nuclear Usability and Knowledge Elicitation, abbreviated GONUKE, is a graded approach to evaluation (Boring et al., 2015). GONUKE includes heuristic evaluation, usability testing, and operator feedback on design as formative evaluation approaches. The scenario-based talkthrough utilised in this thesis could be used as a method to perform formative evaluation before usability testing is suggested in the GONUKE approach. Such an activity would combine the user study and knowledge elicitation evaluation types in GONUKE. GONUKE acknowledges the fact that evaluations may serve different purposes as argued in the first perspective in this thesis. GONUKE also combines evaluations focusing on impact measures with evaluations focusing on design measures (cf. the tactic perspective).

6.7.3 The CRIOP methodology

As discussed in Paper V, CRIOP (Crisis intervention and operability analysis) is a verification and validation methodology from another domain, the oil and gas industry (Johnsen et al., 2011). A CRIOP analysis consists of two main phases: 1) a general analysis using checklists to verify that the control room system satisfies the stated requirements, and 2) a scenario analysis to verify that the control room system satisfies the implied needs. CRIOP thus also utilises a combination of an evaluation focusing on impact measures and an evaluation focusing on design measures, like the method combination tested in this thesis. The main difference between CRIOP and the method combination in the present thesis lies in how much specific direction is given by the method, particularly for the execution of the scenario-based talkthrough. CRIOP is an elaborate verification and validation methodology whose execution is specified in great detail. With its focus on identifying weak spots and iterative execution, a CRIOP analysis will serve formative as well as summative purposes. The method combination tested in Papers IV and V was aimed at formative evaluation only. Compared to CRIOP, the method combination in this thesis give less specific direction on execution, leaving more freedom to the person planning and managing the evaluation to make decisions as seen fit. The scenarios for the scenario-based talkthrough can be prepared prior to the workshop, making it possible to have a shorter evaluation workshop (in CRIOP, they are developed and documented during the workshop).

The scenario analysis in a CRIOP requires going through a list of questions for each event in a scenario, whereas the method combination in the present thesis only requires going through a list of discussion questions after all scenarios are finalised. Studies D and E can be said to have explored a less resource-demanding approach to formative evaluation than CRIOP (which aims at summative evaluation as well). The overall impression from the case studies in Study D was that the gain outweighed the costs. Moreover, some participants even questioned whether or not the approach could be made even more efficient. The participants in Study E did not address this topic at all.

6.7.4 Concluding remarks on the relation to proposed approaches

The above discussion show that the outcome of this thesis is compatible with, and has many similarities to, regulations and proposed evaluation approaches. The contribution of this thesis in relation to current regulations and practices is that it provides a deeper understanding of formative evaluation activity, and shows that evaluation of higher-level design decisions is possible using lower-fidelity system representations. This thesis not only proposes a combination of two different types of evaluation, but also describes why this is beneficial.

6.8 DEVELOPING HUMAN FACTORS AS A DESIGN DISCIPLINE

As argued in Chapter 3, with its aim to advance evaluation practices, the present thesis falls into the category of research for design (Frayling, 1993; Zimmerman et al., 2010). The evaluation activity, and evaluation methods, are meant to support the process of designing. The perspectives presented in this thesis can help frame the design of a method or the modification of a method as part of evaluation planning. They help by directing thought to specific issues and temporarily suspending others (cf. the description of framing in design by Lawson and Dorst, 2009). Improving evaluation planning or method development help improving the process of designing. Through its simultaneous focus on human factors issues, this thesis can thus be said to develop human factors as a design discipline.

The need to develop human factors as a design discipline is something that has been brought up by other researchers. As has been stated earlier in this thesis, human factors is design driven, and as a profession human factors applies theory, data and methods to design (Dul et al., 2012). The approach of merely applying human factors knowledge to design has however been questioned (Vicente et al., 1997; Norros, 2014; Kant, 2017; Norros and Savioja, 2018). Dul et al. (2012) concluded that human factors was underexploited, and Norros (2014) continued this discussion by arguing for the need to develop human factors as a design discipline. Norros (2014) discussed three methodical perspectives she believes could support the development of high-quality human factors, and the last of these is adopting design thinking in human factors. The argument is that

if human factors is expected to be design driven, as Dul et al. (2012) claims, then human factors needs to adopt an epistemology of a design discipline instead of the current epistemology of an applied scientific discipline.

Instead of holding basic science ideals with standardised and well-controlled forms of creating knowledge, Norros (2014) argued that human factors should actively develop formative (i.e. developmental) methods in analysis. The focus on formative evaluation in this thesis connects to this. Formative evaluation is a central activity in design work. When dealing with wicked problems that cannot be definitely defined and for which all possible resolutions cannot be exhaustively specified, evaluating a created resolution to learn what works and what does not becomes a viable way forward for the designer. This thesis has increased knowledge about how formative evaluation of control room systems can be done earlier in the development process. When evaluating what does not yet exist, the impact the design might have on the world must be projected. Earlier evaluation means a wider gap to bridge. Study D showed that it is possible to perform a formative evaluation of control room systems that is not dependent on users directly performing tasks, and to get results that are perceived as valuable. Using design guidelines in evaluation (as in a heuristic evaluation) means that existing knowledge can be used in assessing future impact and bridging the gap. However, one prerequisite for this is that the reasoning behind each specific guideline is known and preferably also documented. Norros (2014) continued to argue that that formative methods should not consider involvedness a threat to objectiveness, but rather an advantage in knowledge creation. The scenario-based talkthrough method utilised in this thesis involves users and use their knowledge about how they work in the existing system to imagine how they might be able to work in the future system. Involvedness is definitely considered an advantage rather than a threat to objectiveness here. The discussion on developing human factors as a design discipline is continued by Norros and Savioja (2018), where human factors as design thinking is presented as one of four principles that can lead human factors to be a knowledge constructing design practice rather than just an application of existing scientific knowledge. One claim is that through design thinking, human factors can be more effective by being oriented to “*contextual constructive solving of complex problems with the purpose of developing the usefulness of the artifactual environment*” (Norros and Savioja, 2018, pp, 170). The perspectives presented in this thesis are meant to make the evaluation activity more adapted to the reality of its practice, thus being both contextual and constructive.

The conclusion of the above reasoning is that the results of this thesis support the development of human factors as a design discipline foremost by its development of the practice of formative evaluation.

FURTHER WORK

With the exception of parts of Study E, the body of work presented in this thesis was undertaken within the Swedish nuclear power domain. A natural continuation of the work would thus include exploring the relevance of the findings in this thesis to other domains and other geographical locations. In particular, the usefulness of the perspectives in evaluation planning and method development must be tested in an industrial context. The method combination of heuristic evaluation and scenario-based walkthrough was the product of an iterative method development process, but this process needs to continue beyond the scope of this thesis. Primarily, the method combination should be improved in terms of support for choosing a suitable system representation, selecting guidelines, formulating discussion questions, and handling evaluation of design proposals that are large in scope. Like the perspectives, the method combination should also be further tested in industrial cases, especially in other domains than nuclear power, and continuously improved.

One facet of the evaluation activity that was not in focus in this thesis, but which would be interesting to explore further, is the impact of the type of system representation on evaluation outcome. Andersen and Broberg (2015) investigated the influence of simulation media on simulation outcome in the domain of hospital work systems. Replicating these studies in the nuclear power domain could further advance control room system evaluation practices. It would be especially interesting to explore the connection between type of system representation and level of design decision specificity. Identifying system representations that encourage assessment of higher-level design decisions could make early evaluations more effective and thus further improve the practice of formative human factors evaluation of control room systems.

CONCLUSIONS

In order to be able to advance evaluation practices as part of the development process, this thesis aimed to increase understanding of human factors evaluation of nuclear power plant control room systems. The first research question in response to this purpose concerned the aspects that are relevant to assess to be able to evaluate the control room system's ability to fulfil its intended purpose. A set of categories of measures was identified that can be used to guide the choice of measures for an evaluation: system performance, task performance, use of resources, user experience, and identification of design discrepancies.

The second research question focused on the relation between the evaluation activity and the development process as a whole. This research question asked if, and how, human factors evaluation can better support control room system development. The research efforts identified a gap in today's evaluation practice and consequently focused on formative evaluation of more general design decisions (higher design levels), preferably undertaken early in the development process. One way to improve formative evaluation is to nuance the purpose of the evaluation. This allows resources to be spent in a more purposeful way and highlights the advantages the evaluation activity brings to the development process, thus strengthening the argument for undertaking formative evaluation in development projects. Developing methods that are useful in practice is another way to advance formative evaluation practices, and this thesis presented a number of guidelines that may be used for this purpose in method development. A combination of two methods, heuristic evaluation and scenario-based walkthrough, was used to explore the evaluation activity in practice. This method combination was found to be useful for formative assessment of higher-level design decisions in nuclear power plant control room systems. The description of the method combination supplied in this thesis is a concrete guide for human factors practitioners undertaking control room system evaluations, especially early in the development process.

From the exploration of the research questions five perspectives emerged for use as decision support by a human factors specialist in evaluation planning and method development:

- *The purpose of the evaluation activity:* the reasons why the activity is undertaken. An evaluation may provide input to design (formative evaluation), or provide quality assurance (summative evaluation). Formative evaluation can identify discrepancies in the design, but may also identify new requirements (or knowledge that can be used to formulate new requirements). A formative

evaluation may also serve communicative purposes, or provide input to a summative evaluation. Actively defining the purposes of the evaluation activity allows for a better tailoring of the activity to this purpose.

- *The object to be evaluated:* two important views when considering the object to be evaluated are the measures needed to assess the design and the level of design decision specificity. The categories of measures presented in this thesis can be used to guide the choice of evaluation measures (and by association, evaluation method). The level of design decision specificity to be evaluated affects what is a suitable fidelity of the system representation, which in turn may limit method choice. Being aware of the level of design decision specificity to be evaluated also makes it possible to focus evaluation efforts where they are most useful.
- *The tactic used in the evaluation activity:* if the evaluation seeks the existence and location of known typical design problems (unknown knowns) or seeks to identify and locate unknown problems (unknown unknowns). The word 'problem' here denotes both discrepancies in the proposed design and new requirements (or knowledge that can be transformed into new requirements). An evaluation that seeks unknown knowns uses design measures (the focus is on the design itself), and an evaluation that seeks unknown unknowns uses impact measures (the focus is on the impact of the design). This thesis recommends a combination of both tactics, but considering the tactic perspective in evaluation planning or method development allows a conscious choice of what is lost if one of the two tactics is excluded.
- *The evaluation procedure:* execution of the evaluation activity can be divided into a number of steps. An evaluation method may acknowledge and/or support all these steps, or some of them, and to varying degrees. Choosing a specific method means that certain steps are determined by the method, limiting the choices that have to be made. When developing a method, taking account of the steps in the evaluation procedure will highlight the actions the method should support.
- *The use of the evaluation method:* the user, use, and context of use of the evaluation method should be taken into account to develop (or modify) a method that is useful in practice and bring about the desired impact. The guidelines for developing or modifying human factors evaluation methods that are useful in practice presented in this thesis support consideration of this perspective.

REFERENCES

- Andersen, S. N., Broberg, O. (2015) Participatory ergonomics simulation of hospital work systems: The influence of simulation media on simulation outcome. *Applied Ergonomics*, vol. 51, pp. 331-342.
- Andersson, J., Bligård, L.-O., Osvalder, A.-L., Rissanen, M. J., Tripathi, S. (2011) To Develop Viable Human Factors Engineering Methods for Improved Industrial Use. In *Design, User Experience, and Usability, Pt 1 I, Human Computer Interaction International, Orlando FL 2011, Lecture Notes in Computer Science*, vol 6769, eds. A. Marcus, pp. 355-362. Berlin: Springer.
- 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.
- Baber, C. (2005) Evaluation in human-computer interaction. In *Evaluation of human work (3rd ed.)*, eds. J. R. Wilson and N. Corlett. Boca Raton, FL: Taylor & Francis.
- Berlin, C., Bligård, L.-O., Simonsen, E. (2017) Using the ACD3-ladder to manage multi-phase requirements on end-user products. In *Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 4: Design Methods and Tools*; 21-25 August 2017, Vancouver. pp. 425-434.
- Bligård, L. O., Simonsen, E., Berlin, C. (2016) ACD³ - a new framework for activity-centered design. In *NordDesign 2016*; 10-12 August 2016, Trondheim.
- Boring, R., Lau, N. (2017) Measurement sufficiency versus completeness: Integrating safety cases into verification and validation in nuclear control room modernization. In *Proceedings of the AHFE 2016 International Conference on Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries*; 27-31 July 2017, Walt Disney World®, Florida. pp. 79-90.
- Boring, R.L. (2017) As low as reasonable assessment (ALARA): Applying discount usability to control room verification and validation. In *Risk, Reliability and Safety: Innovating Theory and Practice*, eds. L. Walls, M. Revie and T. Bedford, pp. 950-955. London: Taylor & Francis Group.
- Boring, R. L., Ulrich, T. A., Joe, J. C., Lewb, R. T. (2015) Guideline for Operational Nuclear Usability and Knowledge Elicitation (GONUKE). In *6th International Conference on Applied Human Factors and Ergonomics*; 26-30th of July 2015, Las Vegas, Nevada.
- Braarud, P. Ø., Rø Eitheim, 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).

- Braha, D., Reich, Y. (2033) Topological structures for modeling engineering design processes. *Research in Engineering Design*, vol. 14, pp. 185-199.
- Evaluate. (2018) In *Britannica Academic*. <http://academic.eb.com> (24 May 2018).
- Buchanan, R. (1992) Wicked Problems in Design Thinking. *Design Issues*, vol. 8, no. 2, pp. 5-21.
- Creswell, J. W. (2014) *Research Design : Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: Sage.
- Cross, N. (2008) *Engineering design methods: strategies for product design* (4th ed). Chichester: Wiley.
- Crotty, M. (1998) *The foundations of social research: Meaning and perspective in the research process*. London: Sage.
- Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W. S., Wilson, J. R., van der Doelen, B. (2012) A strategy for human factors/ergonomics: developing the discipline and profession. *Ergonomics*, vol. 55, no. 4, pp. 377-395.
- Electric Power Research Institute (2005) *Guidance for the Design and Use of Automation in Nuclear Power Plants*. Palo Alto: Electric Power Research Institute. (Report no: 1011851).
- Flach, J. M. (2012) Complexity: learning to muddle through. *Cognition, Technology & Work*, vol. 14, no. 3, pp. 187-197.
- Frayling, C. (1993) Research in Art and Design. *Royal College of Art Research Papers*, vol. 1, no. 1, pp. 1-5.
- Hale, A., Kirwan, B., Kjellén, U. (2007) Safe by design: where are we now? *Safety Science*, vol. 45, no. 1-2, pp. 305-327.
- Hendrick, H., Kleiner, B. (2001) *Macroergonomics: An Introduction to Work System Design*. Santa Monica, California: Human Factors & Ergonomics Society.
- Hollnagel, E. (2009) *The ETTO principle: Efficiency-Thoroughness Trade-Off. Why things that go right sometimes go wrong*. Farnham, UK: Ashgate.
- Hollnagel, E. (2011) 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.
- IAEA International Nuclear Safety Advisory Group [INSAG] (1999) *Basic Safety Principles for Nuclear Power Plants*. Vienna: International Atomic Energy Agency. (INSAG: 75-INSAG-3).
- INCOSE & Wiley (2015) *INCOSE Systems Engineering Handbook : A Guide for System Life Cycle Processes and Activities* (4th ed.). New York: Wiley.

- Institute of Electrical and Electronics Engineers [IEEE] (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 [IAEA] (2007) *IAEA Safety Glossary Terminology Used in Nuclear Safety and Radiation Protection — 2007 Edition*. Vienna: International Atomic Energy Agency.
- International Atomic Energy Agency [IAEA] (2016) *Safety of Nuclear Power Plants: Design*. Vienna: International Atomic Energy Agency. (Specific Safety Requirements: IAEA Safety Standards Series No. SSR-2/1 (Rev. 1)).
- International Electrotechnical Commission [IEC] (2009) *IEC 60964:2009 Nuclear power plants - Control rooms - Design*. Geneva: International Electrotechnical Commission.
- International Ergonomics Association [IEA]. (2018) *Definition and Domains of Ergonomics*. <http://www.iea.cc/whats/> (18 March 2018).
- International Standard Organisation [ISO] (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 [ISO] (2006) *ISO 11064-7:2006 Ergonomic design of control centres – Part 7: Principles for the evaluation of control centres*. Geneva: International Standard Organisation.
- International Standard Organisation [ISO] (2010) *ISO 9241-210:2010 Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. Geneva: International Standard Organisation.
- Johnsen, S. O., Bjørkli, C., Steiro, T., Fartum, H., Haukenes, H., Ramberg, J., Skriver, J. (2011) *CRIOP: A scenario method for Crisis Intervention and Operability analysis*. Trondheim: SINTEF.
- Kant, V. (2017) Muddling between science and engineering: an epistemic strategy for developing human factors and ergonomics as a hybrid discipline. *Theoretical Issues in Ergonomics Science*, vol., pp. 1-30.
- Kovesdi, C., Joe, J., Boring, R. (2018) A guide for selecting appropriate human factors methods and measures in control room modernization efforts in nuclear power plants. In *Joint Proceedings of the AHFE 2018 International Conference on Human Factors in Artificial Intelligence and Social Computing, Software and Systems Engineering, The Human Side of Service Engineering and Human Factors in Energy*; July 21-25 2018, Orlando, Florida.
- Laarni, J., Savioja, P., Karvonen, H., Norros, L. (2011) Pre-validation of nuclear power plant control room design. In *International Conference on Engineering Psychology and Cognitive Ergonomics (EPCE 2011)*; 9-14 July 2011, Orlando, Florida. pp. 404-413.

- Laarni, J., Savioja, P., Norros, L., Liinasuo, M., Karvonen, H., Wahlström, M., Salo, L. (2014) Conducting multistage HFE validations – constructing Systems Usability Case. In *Proceedings of the ISOFIC/ISSNP 2014*; 24–28 August 2014 2014, Jeju, Republic of Korea.
- Lawson, B. (2006) *How designers think*. Oxford: Architectural Press.
- Lawson, B., Dorst, K. (2009) *Design expertise*. Oxford: Elsevier.
- Meister, D. (1991) The epistemological basis of human factors research and practice. In *Proceedings of the Human Factors Society*; 1991, pp. 1275–1279.
- Nemeth, C. P., Herrera, I. (2015) Building change: Resilience Engineering after ten years. *Reliability Engineering & System Safety*, vol. 141, pp. 1–4.
- Nielsen, J. (1993) *Usability Engineering*. San Diego: Academic Press.
- Norell, M. (1992) *Stödmetoder och samverkan i produktutvecklingen*. The Royal Institute of Technology. (PhD Diss.).
- Norros, L. (2014) Developing human factors/ergonomics as a design discipline. *Applied Ergonomics*, vol. 45, no. 1, pp. 61–71.
- Norros, L., Savioja, P. (2018) Principles of Human Factors Engineering. In *Handbook of safety principles* [E-book], eds. N. Möller, S. O. Hansson, J.-E. Holmberg and C. Rollenhagen, pp. 164–195. Hoboken, NJ: John Wiley & Sons Inc.
- 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).
- Osvelder, 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.
- Method. (2018) In *Oxford Dictionaries*. <http://www.oed.com> (24 May 2018).
- 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.
- Patriarca, R., Bergström, J., Di Gravio, G., Costantino, F. (2018) Resilience engineering: Current status of the research and future challenges. *Safety Science*, vol. 102, pp. 79–100.
- Perrow, C. (1999) *Normal Accidents: Living with High-Risk Technologies*. Princeton, NJ: Princeton University Press.

- Pew, R. W., Mavor, A. S. (2007) *Human-System Integration in the System Development Process: A New Look*. Washington, D.C.: National Academy of Sciences.
- Rittel, H. W. J., Webber, M. M. (1973) Dilemmas in a general theory of planning. *Policy Sciences*, vol. 4, no. 2, pp. 155-169.
- 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.
- Scriven, M. (1967) The methodology of evaluation. In *Perspectives on Curriculum Evaluation (AERA Monograph Series – Curriculum Evaluation)*, eds. R. Tyler, R. Gagne and M. Scriven. Chicago: Rand McNally and Co.
- Shorrock, S. T., Williams, C. A. (2016) Human factors and ergonomics methods in practice: three fundamental constraints. *Theoretical Issues in Ergonomics Science*, vol. 17, no. 5-6, pp. 468-482.
- Simon, H. A. (1996) *The Sciences of the Artificial* (3rd ed.). London: MIT Press.
- 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.
- SSMFS 2008:1. *The Swedish Radiation Safety Authority's Regulations and General Advice concerning Safety in Nuclear Facilities*. Stockholm: Swedish Radiation Safety Authority.
- SSMFS 2008:17. *The Swedish Radiation Safety Authority's Regulations and General Advice concerning the Design and Construction of Nuclear Power Reactors*. Stockholm: Swedish Radiation Safety Authority.
- Ullman, D. G. (1997) *The Mechanical Design Process*. New York: McGraw-Hill.
- Ulrich, K. T., Eppinger, S. D. (2003) *Product Design and Development* (3rd ed). New York: McGraw-Hill.
- United States Nuclear Regulatory Commission [US NRC] (2012) *NUREG-0711 Human Factors Engineering Program Review Model Revision 3*. Washington, DC: United States Nuclear Regulatory Commission.
- Vicente, K. J. (2004) *The Human Factor*. Toronto: Random House.
- Vicente, K. J., Burns, C. M., Pawlak, W. S. (1997) Muddling Through Wicked Design Problems. *Ergonomics in Design*, vol. 5, no. 1, pp. 25-30.
- Wilson, J., R. (2005) Methods in the understanding of human factors. In *Evaluation of human work* (3rd ed), eds. J. Wilson, R. and E. Corlett, N., pp. 1-31. [e-book].

- Virzi, R. A., Sokolov, J. L., Karis, D. (1996) Usability problem identification using both low- and high-fidelity prototypes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground*; 13–18 April 1996, Vancouver. pp. 236–243.
- Yin, R. K. (2014) *Case study research: design and methods* (5th ed). London: SAGE.
- Zimmerman, J., Forlizzi, J., Evenson, S. (2007) Research Through Design as a Method for Interaction Design Research in HCI. In *Proceedings of CHI*; April 28–May 3 2007, San Jose, California. pp. 493–502.
- Zimmerman, J., Stolterman, E., Forlizzi, J. (2010) An analysis and critique of Research through Design: towards a formalization of a research approach. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*; August 16–20 2010, Aarhus. pp. 310–319.

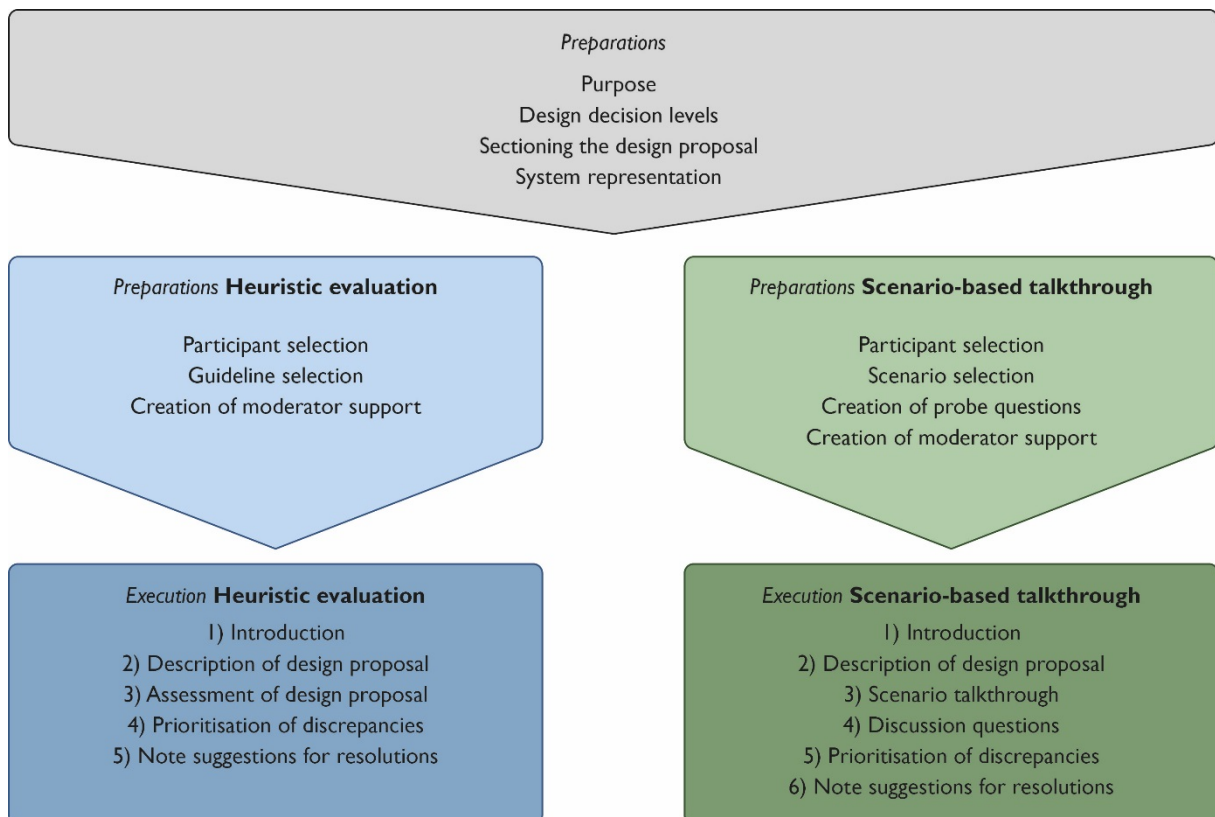
APPENDIX A

Heuristic evaluation & Scenario-based walkthrough

– Evaluation of control room systems to provide input to design

Heuristic evaluation & Scenario-based walkthrough

- *Evaluation of control room systems to provide input to design*



SUMMARY

This document describes a combination of two methods for evaluating control room systems with the primary purpose of providing input to design. The two methods are *heuristic evaluation* and *scenario-based walkthrough*.

CONTENTS

1	Introduction.....	1
1.1	Relation to the development process	2
2	Joint preparations.....	4
2.1	The purpose of the evaluation.....	4
2.2	Design decision levels to be evaluated	4
2.3	Sectioning the design proposal	5
2.4	System representation	5
2.5	Participants	5
3	Heuristic evaluation	6
3.1	Preparations.....	6
3.1.1	Selection of workshop participants.....	6
3.1.2	Selection of guidelines	6
3.1.3	Moderator support.....	6
3.2	Execution.....	7
3.2.1	Workshop procedure	7
4	Scenario-based walkthrough	9
4.1	Preparations.....	9
4.1.1	Selection of workshop participants.....	9
4.1.2	Selection of scenarios.....	10
4.1.3	Discussion questions.....	13
4.1.4	Moderator support.....	14
4.2	Execution.....	14
4.2.1	Procedure workshop.....	14
5	Templates	16

I INTRODUCTION

This document describes a combination of two methods for evaluation of control room systems with the primary purpose of providing input to design (so-called “*formative evaluation*”). The two methods are heuristic evaluation and scenario-based walkthrough. This text focuses on the practical execution of the evaluation.

This method combination requires that a human factors (HF) specialist is responsible for planning and execution, and this text is written for that group of personnel.

Chapter 2 describes the decisions and necessary preparations for both methods. Chapter 3 describes heuristic evaluation and Chapter 4 scenario-based walkthrough. Each method is divided into a preparation part and an execution part. Figure 1 provides an overview of the method combination.

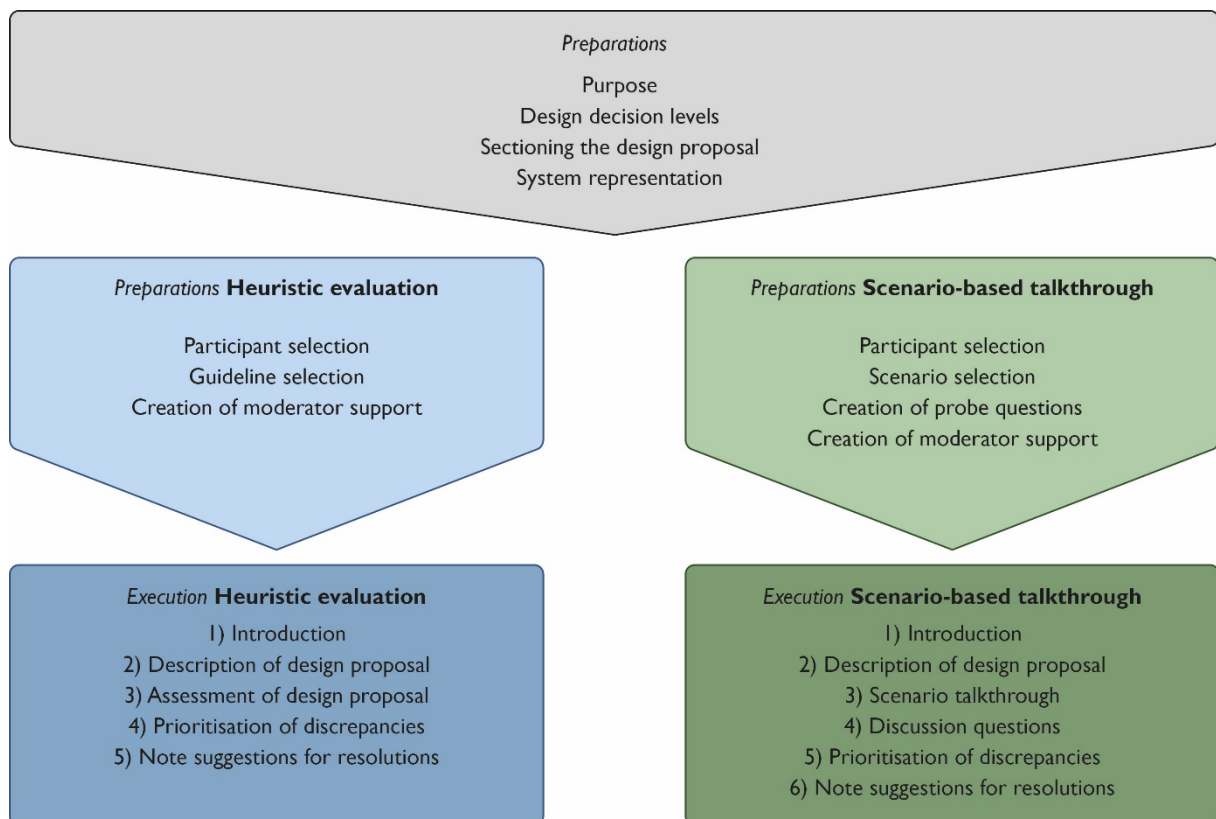


Figure 1: Overview of method combination.

1.1 Relation to the development process

An evaluation activity in a specific development project may serve multiple purposes. The method combination described here is tailored to provide input to improve a design, so-called *formative* evaluation. A formative evaluation can identify discrepancies in a proposed design, but may also identify new requirements (or information that can be used to formulate requirements), as well as good aspects of the design that should be kept when developing it further. A formative evaluation may also serve communicative purposes. Through the evaluation activity can different stakeholders meet, exchange information, and create a common view of the design proposal and how it should be developed further. This can facilitate user acceptance of a new design and give project members better knowledge of the future use of the design.

An evaluation may also have the purpose of assessing and documenting the quality of the design, so-called *summative* evaluation. Because of external demands (for instance from a governmental authority), summative evaluation is often required to be more systematic, transparent and documented than formative evaluation. Meeting these demands is easier if the evaluation activity was planned to fulfil a summative purpose from the onset. The method combination that is described in the present document is tailored for formative evaluation, but it also contains aspects that are useful to bear in mind if the evaluation activity aims at fulfilling summative purposes as well¹.

Evaluation should be undertaken repeatedly during the course of the development process, and not only towards the end. The method combination described here can be used early in the development process since neither heuristic evaluation nor scenario-based talkthrough is dependent on a system representation that can be used in a way that is very similar to real use. Therefore, they can for example be used before detailed operator interfaces are developed. During the course of the development process, the method combination presented here should be complemented with other evaluation methods to fully assess the control room system, for example with usability tests.

The methods combined here complement each other. Heuristic evaluation focuses on identifying and locating known typical design problems in the proposed design using design guidelines. Design guidelines are knowledge of successful design solutions presented as design advice (either about how to do something or how *not* to do something). However, an evaluation using guidelines will most likely not identify problems not covered by the design guidelines. The scenario-based talkthrough focuses on the impact of the proposed design, such as how users perform tasks, and the findings can be traced back to the design discrepancies that caused them. Through this approach the scenario-based talkthrough is able to identify and locate design discrepancies that are not typical and explicitly sought. However, assessment through scenarios is a resource-intensive way to evaluate, and complementing this approach with heuristic evaluation is a way of handling the trade-off between efficiency and thoroughness.

There is no strict order in which the two methods should be executed, in other words whether heuristic evaluation should be executed first or whether the scenario-based talkthrough should be undertaken first. There are advantages with both approaches. If the scenario-based talkthrough is executed first it can provide insights into weak parts of the proposed design that should be

¹ For a similar, but more elaborate, methodology than the one presented here, see Johnsen, S. O., Bjørkli, C., Steiro, T., Fartum, H., Haukenes, H., Ramberg, J., Skriver, J. (2011) CRIOP: A scenario method for Crisis Intervention and Operability analysis. Trondheim: SINTEF. This methodology may be used as further guidance when planning an evaluation activity meant to fulfil summative purposes as well as formative.

emphasised in the heuristic evaluation, and guidelines should be selected in line with this. Undertaking the heuristic evaluation first gives the participants a good opportunity to become more familiar with the proposed design. This is an advantage especially if the same persons with operational knowledge participate in both the heuristic evaluation and the scenario-based talkthrough, since less time is needed in the latter to explain the design.

2 JOINT PREPARATIONS

2.1 The purpose of the evaluation

Since an evaluation activity can serve multiple purposes (such as formative, summative, communicative) it is important to define its purpose(s) to be able to adapt the evaluation accordingly. For example, if the evaluation is required to fill a communicative purpose, it is important to include project leaders and designers. If the aim of the evaluation is to fulfil a summative purpose, however, it might be better not to include project leaders.

2.2 Design decision levels to be evaluated

Define the design level², that is to say the level of specificity of the design decisions to be evaluated. Compare the design concept to be evaluated with the examples in Table 1. If several design levels are applicable for the design concept in question, choose the lower one.

Table 1: Design levels and examples. "Effect" is regarded as the highest design level, and "Interaction" as the lowest.

<i>Design level</i>	<i>Example of design decision for control room systems</i>
Effect (The effect that the system* is intended to achieve in its context; impact goal)	<ul style="list-style-type: none"> • Desired impact, e.g. "Operating the plant without exposing personnel, the general public, or the environment to harmful levels of radiation from liquid or solid waste" • Affected user groups, e.g. operator in the waste management control room, maintenance personnel, cleaning personnel • Definition of overall functions, e.g. "Enable discharge of water with acceptable levels of radiation in the ocean", "Enable transportation of solid waste" • Definition of overall tasks, e.g. "Handle solid and liquid radioactive waste"
Usage (How the system is used by its users)	<ul style="list-style-type: none"> • Definition of affected process systems/functionality of new process systems in the plant. • Definition of tasks, e.g. "Separate radioactive particles from liquid waste", "Perform administrative tasks", "Handle disturbances" • Number of operators who will use the system • Location in the plant (e.g. building and room) – movement patterns and distances • Medium for operator interfaces (e.g. screen-based or analogue)
Architecture (The technical architecture of the system)	<ul style="list-style-type: none"> • Definition of more detailed tasks, e.g. "Filter water from tank A to tank B", "Test water in tank C" • Location of operator interfaces in control room – movement patterns and distances • Number of screens
Interaction (The interaction between the system and user/context in detail)	<ul style="list-style-type: none"> • Definition of even more detailed tasks, e.g. "Press button D", "Monitor value on gauge E" • Design of operator interfaces • Exact placement of components (xyz-position)

* E.g. a control room system that is to be developed or modified.

² For more information on design levels, see Bligård, L. O., Simonsen, E., Berlin, C. (2016) ACD³ - a new framework for activity-centered design. In *NordDesign 2016*; 10-12 August 2016, Trondheim, Norway.

2.3 Sectioning the design proposal

Some parts of the evaluation activity may be difficult if the design proposal to be evaluated is large in scope. It may, for example, be difficult to assess how well a large-scope design proposal fulfils a more general guideline in the heuristic evaluation or relates to a discussion question in the scenario-based talkthrough. In these cases, it may simplify the evaluation if the design proposal is sectioned into smaller parts (for instance different parts of the plant, subsystems, different parts of the control room) and the assessment is undertaken per part. It may for example be easier to assess whether the design is consistent if smaller parts are assessed one at a time. It is, however, important to not *only* consider the parts – an assessment of the design as a whole must also be performed to avoid sub-optimisation.

2.4 System representation

The design proposal needs to be visualised to be communicated efficiently to the participants in the evaluation. Visualising the design proposal makes it easier for the participants to imagine its use, and may for example consist of 2D drawings (of operator interfaces, rooms, or buildings), digital 3D models, scale models or full-scale models. The system representation should visualise the design decisions to be evaluated (see section 2.2).

A physical representation of the design proposal is preferable, since this allows more direct interaction with the system representation during the workshops. This will facilitate communication within the group. If the system representation is a top-view drawing in 2D or a scale model, each participant may be given a representation of themselves (such as LEGO™ figures) that they can move around to further ease communication within the group and to aid in visualising usage.

2.5 Participants

Participants in the evaluation workshops should be HF specialists, user representatives, as well as designers and project leaders from the development project in question. More support for selecting participants is provided in section 3.1.1 (heuristic evaluation) and in section 4.1.1 (scenario-based talkthrough).

3 HEURISTIC EVALUATION

Heuristic evaluation is a classical so-called usability inspection method³. In short, the method involves assessing how well a proposed design fulfils a number of design guidelines. This chapter describes the preparations that should be made ahead of the heuristic evaluation workshop, as well as the execution of the evaluation workshop.

3.1 Preparations

3.1.1 Selection of workshop participants

The following persons should participate in the evaluation workshop:

- *HF specialist(s)*. An HF specialist is the moderator and steers the workshop, but should also take notes to document the result of the workshop. The moderator is responsible for explaining guidelines and requirements if needed. If more than one HF specialist participates in the workshop, decide beforehand who should moderate and who should take notes. If the aim of the evaluation activity is to fulfil a summative purpose as well as a formative one, it is suitable if the HF specialist has not been involved in the creation of the design proposal to be evaluated.
- *Representatives of the users* (e.g. persons with operational experience). Persons familiar with the work to be done are necessary in the evaluation to be able to decide if the design proposal is in line with the guidelines. These persons do not have to work actively in the user role, but should have done so in the past. If the aim of the evaluation activity is to fulfil a summative purpose as well as a formative one, it is suitable if the user representatives have not been involved in the creation of the design proposal to be evaluated.

It can also be beneficial to include the following persons:

- *Designer*. If the other participants do not have sufficient knowledge of the technical aspects of the design proposal, a designer should be added as a participant. This person can then answer technical questions during the workshop.
- *Project leader*. It may be beneficial to include the project leader in the workshop, since this gives him or her a better understanding of HF-related requirements and guidelines and their relation to the design proposal.

3.1.2 Selection of guidelines

If suitable requirements and/or guidelines are already documented in the project, it is beneficial to use them in the heuristic evaluation. If no suitable requirements and guidelines exist in the project, they can be selected from standards and collections of guidelines. Exactly which guidelines are suitable depends on the design proposal to be evaluated, as well as which standards are applicable for the domain. For example, some guidelines may be impossible to use in a specific phase of a development project because the level of detail in the design proposal is not sufficient to determine if the guideline is fulfilled or not.

3.1.3 Moderator support

In order to support the moderator during the workshop and support documentation, a table with the selected requirements and guidelines should be created (see the template at the end of this document). As a minimum, it should have the following columns:

³ See e.g. Nielsen, J. (1994) Heuristic Evaluation. In Usability Inspection Methods, eds. J. Nielsen and R. L. Mack, pp. 25-62. New York: John Wiley & Sons.

- “Number” (numbering of requirements)
- “Guidelines/requirements” (description of the guideline/requirement)
- “Identified discrepancies in the design” (for evaluation findings)
- “New requirements” (for evaluation findings, new requirements/information that is relevant for the continuing development of the design)
- “Prioritisation” (how severe the discrepancy is in relation to other identified discrepancies)
- “Suggestion for resolution” (for evaluation findings, ideas on how to resolve the discrepancy)

If a **task analysis** has been previously developed for the control room in question, it can be brought to the evaluation workshop to provide an overview of the tasks that the design proposal must support. Scenarios developed for the scenario-based talkthrough can be used for this as well. This can make the assessment more systematic if the design proposal is large in scope (the fulfilment of a specific guideline can be analysed for the design as a whole, for part of the design, and for different tasks).

3.2 Execution

The procedure in this section is written for the moderator’s perspective.

3.2.1 Workshop procedure

- 1) **Introduction** to workshop
 - a. **Purpose of the workshop:** Explain the purpose of the evaluation activity to the participants.
 - b. **Desired result:** Explain to the participants that the aim is to identify discrepancies, identify new requirements to further detail the solution, identify positive aspects, and identify possible ways to resolve identified discrepancies.
 - c. **Roles and tasks during the workshop:** Explain to the participants that everyone has a responsibility to critically review the design proposal, and to review it against the requirements/guidelines as objectively as possible. As a moderator, it is important that you do not try to defend the design proposal, try instead to focus on the desired impact of the design and on whether the design will contribute to this. Encourage the participants to take individual notes to remember their thoughts when they cannot express them out loud, such as resolutions to discrepancies in the guideline review.
- 2) **Description of design proposal**
 - a. **Desired impact** of the design: Describe the impact the proposed design is meant to have on its environment.
 - b. **Design proposal:** Describe the design proposal.

- 3) Review of design proposal* against guidelines. *For each guideline:*
 - a. **Discuss how well the design meets the guideline.**
 - b. **Note** (done by the participant appointed to take notes):
 - i. **discrepancies in the design proposal**
 - ii. **new requirements/ideas that are relevant for the further development of the design** (but which are not direct discrepancies in the design)
 - iii. **positive aspects in the design proposal**

* If the design proposal is large in scope it might be more manageable to section the design proposal into smaller parts and assess each part individually against the guidelines. However, the design as a whole should also be reviewed to avoid sub-optimisation. A task analysis or scenarios may also be used to achieve a more systematic review.

- 4) (*When all guidelines are reviewed*) Ask if the participants have discovered **additional discrepancies/new requirements/positive aspects** that are not related to the guidelines. Note them.
- 5) (*When all guidelines are reviewed*) **Prioritise identified discrepancies.** This prioritisation should take into account how severe the consequences of identified discrepancies may be and how often these consequences may occur. The importance here is to facilitate a discussion about the severity of the discrepancies in relation to each other (e.g. “more serious than ...” / “less serious than ...”) to provide a basis for discussions about how to handle the discrepancies in other forums. No consideration should be given during the workshop as to how difficult it would be to resolve the discrepancies.
- 6) (*When all guidelines are reviewed*) Note **ways to resolve** identified discrepancies. This should be done separately when all guidelines are reviewed so as not to inhibit the identification of deficiencies. The identification of possible resolutions does not have to be comprehensive and deep, but obvious resolutions should be noted.

4 SCENARIO-BASED TALKTHROUGH

The scenario-based walkthrough presented here is inspired by other scenario-based evaluation methods⁴, but is tailored for control room evaluation. In short, the method helps users simulate or describe how they would have handled different usage scenarios using the design proposal. This chapter describes the preparations that should be made ahead of the scenario-based walkthrough workshop and the execution of the evaluation workshop.

4.1 Preparations

4.1.1 Selection of workshop participants

The following persons should participate in the evaluation workshop:

- *HF specialist*. An HF specialist is the moderator and steers the workshop, but should also take notes to document the result of the workshop. If more than one HF specialist participates in the workshop, decide beforehand who should moderate and who should take notes. If the purpose of the evaluation activity is to fulfil a summative purpose as well as a formative one, it is suitable if the HF specialist has not been involved in the creation of the design to be evaluated. It is important that the HF specialist does not influence the other participants' opinions of the design.
- *Representatives of the users* (e.g. operational personnel). Persons actively working in the user role in question, for example as control room operators. The role of the user representatives at the workshop is to simulate/describe the use of the design proposal and give their opinion of its imagined use. If the aim of the evaluation activity is to fulfil a summative purpose as well as a formative one, it is suitable if the user representatives have not been involved in the creation of the design to be evaluated.

It can also be beneficial to include the following persons:

- *Designer*. It can be beneficial to include designers at the workshop since this allows the users to get answers to any detailed technical questions they might have. It also gives the designers a better understanding of the use of the proposed design. It is important that the designers do not influence the other participants' opinions of the design.
- *Project leader*. It may be beneficial to include the project leader in the workshop, since this gives him or her a better understanding of the use of the design proposal. It is important that the project leader does not influence the other participants' opinions of the design.

⁴ Primarily *Group-based expert walkthrough*, see Folstad, A. (2007) Group-based Expert Walkthrough. In COST294-MAUSE 3rd - Review, Report and Refine Usability Evaluation Methods (R³UEMs), 5 March 2007, Athens. pp. 58-60., and *Participatory simulation*, see Andersen, S. N. (2016) Participatory simulation in hospital work system design. Technical University of Denmark. (Dissertation).

4.1.2 Selection of scenarios

A scenario is a description of a situation to be addressed using the design proposal. Scenarios can be of **three types**⁵:

- **A (“On the spot” scenarios):** Scenarios are developed during the evaluation workshop.
- **B (Case stories):** Scenarios are developed prior to the workshop and describe a situation to be handled by the users. As users describe or simulate how they would handle the situation with the help of the proposed design, the moderator can add new unexpected events in the situation that the users should handle.
- **C (Task sequences):** Scenarios are developed prior to the workshop and have a more script-like form where the tasks the users should execute using the proposed design are described in greater detail. As users describe or simulate how they would handle the situation with the help of the proposed design, the moderator can add new unexpected events in the situation that the users should handle.

See decision support in Tables 2 and 3 to choose **which of the scenario types are suitable** for the evaluation in question.

If scenario types B or C are found suitable, a decision must be made regarding the content of the scenarios. For nuclear power, section 11.4.1 (“Sampling of Operational Conditions”) in NUREG-0711⁶ provides good support in this regard.

⁵ From Andersen, S. N. (2016) Participatory simulation in hospital work system design. Technical University of Denmark. (Dissertation).

⁶ United States Nuclear Regulatory Commission (2012) NUREG-0711 Human Factors Engineering Program Review Model Revision 3. Washington, DC: United States Nuclear Regulatory Commission.

Table 2: Decision support for choice of scenario type⁷.

<i>Scenario type</i>	<i>Suitable when...</i>	<i>Preparations and execution</i>
A ("On the spot"-scenarios)	<ul style="list-style-type: none"> • ...documentation on the use of the proposed design is limited (e.g. no task analyses have been made) and data collection needs to be done. • ...if there are no available resources for developing scenarios ahead of the workshop. 	<p><i>Preparations:</i> This scenario type does not require any preparations prior to the workshop, as it is developed during the course of workshop.</p> <p><i>Execution:</i> With this scenario type, the workshop participants have a lot of influence on the content of the workshop and the aspects of the proposed design that are evaluated. However, there is a risk that the participants may not simulate/describe the tasks in the proposed design and only discuss the scenarios. Therefore, expectations of participants' roles and tasks during the workshop must be clearly explained, and during the course of the workshop the moderator must encourage behaviour in accordance with the roles as needed.</p>
B (Case stories)	<ul style="list-style-type: none"> • ...the imagined use of the proposed design is documented. • ...the available resources for developing scenarios before the workshop are limited. • ...there is a need for a discussion about new/alternative ways of working and strategies to handle different situations that is greater than the need to control the scenario talkthrough in detail (e.g. to evaluate specific tasks in detail). 	<p><i>Preparations:</i> The scenarios should include a description of the initial situation and what the users' goals are (end situation). Examples of information that may be included to describe the initial and end situations: mode of operation, value of relevant process parameters, ongoing work, where different actors are located (for example, where in the plant personnel are currently performing tasks), important events in the environment (such as the weather). To increase the level of realism in the scenario talkthrough unexpected events for the users to deal with can be prepared for the moderator to initiate when appropriate.</p> <p><i>Execution:</i> With this scenario type, there is a risk that the participants may not simulate/describe the tasks in the proposed design and only discuss the scenarios. Therefore, expectations of participants' roles and tasks during the workshop must be clearly explained, and during the course of the workshop the moderator must encourage behaviour in accordance with the roles as needed.</p>

⁷ Largely influenced by Andersen, S. N. (2016) Participatory simulation in hospital work system design. Technical University of Denmark. (Dissertation).

Table 3: Decision support for choice of scenario type, continued⁸.

Scenario type	Suitable when...	Preparations and execution										
C (Task sequences)	<ul style="list-style-type: none"> • ... the imagined use of the proposed design is documented. • ...adequate resources for developing scenarios prior to the workshop are available. • ...there is a need to control the scenario talkthrough during the workshop in detail (there is a need to assess specific tasks at a detailed level) that is greater than the need for a discussion about new/alternative ways of working and strategies for handling different situations. • ...there is a need to assess parallel task sequences (e.g. different users performing different tasks at the same time). • ...the users are not familiar with the tasks (e.g. if the proposed design entails new tasks). • ...the evaluation activity is designed to serve a summative purpose, since this scenario type allows transparency and documentation in detail regarding the content of the workshop. 	<p><i>Preparations:</i> The scenarios should include a description of the initial situation and what the users' goals are (end situation), as well as the tasks the users should perform in between. Examples of information that scenarios may include: mode of operation, value of relevant process parameters, ongoing work, where different actors are located (for example, where in the plant personnel are currently performing tasks), important events in the environment (such as the weather). The level of detail in the script for the user's simulation of tasks during the workshop should correspond to the design level (see examples in Table 4 below). To identify new requirements it is beneficial if the level of detail in tasks corresponds to the design level <i>below</i> the design level of the proposed design. If the ability to take specific decisions is to be evaluated, scenarios should be formulated so that users are forced to make decisions, a particular decision should not be provided by the scenario. To increase the realism in the scenario talkthrough, unexpected events for the users to deal with can be prepared for the moderator to initiate when appropriate.</p> <p><i>Table 4: Examples of level of detail for tasks in scenarios for different design levels.</i></p> <table border="1" data-bbox="1093 743 1939 962"> <thead> <tr> <th data-bbox="1093 743 1279 767">Design level</th> <th data-bbox="1290 743 1939 767">Example of level of detail for tasks in scenario</th> </tr> </thead> <tbody> <tr> <td data-bbox="1093 775 1279 831">Effect</td> <td data-bbox="1290 775 1939 831">Like in scenario type B (a more detailed scenario is not possible for this design level)</td> </tr> <tr> <td data-bbox="1093 839 1279 895">Usage</td> <td data-bbox="1290 839 1939 895">"Separate radioactive particles from liquid waste", "Execute administrative tasks", "Handle disturbances"</td> </tr> <tr> <td data-bbox="1093 903 1279 927">Architecture</td> <td data-bbox="1290 903 1939 927">"Filter water from tank A to tank B", "Test water in tank C"</td> </tr> <tr> <td data-bbox="1093 935 1279 962">Interaction</td> <td data-bbox="1290 935 1939 962">"Press button D", "Monitor value on gauge E"</td> </tr> </tbody> </table> <p><i>Execution:</i> With scenario type C (task sequences), it might be easy to focus on simulating/describing the detailed tasks and not as much on discussing the solution as a whole. The discussion questions in the scenario-based talkthrough workshop are meant to help focus the discussion on the design proposal as a whole.</p>	Design level	Example of level of detail for tasks in scenario	Effect	Like in scenario type B (a more detailed scenario is not possible for this design level)	Usage	"Separate radioactive particles from liquid waste", "Execute administrative tasks", "Handle disturbances"	Architecture	"Filter water from tank A to tank B", "Test water in tank C"	Interaction	"Press button D", "Monitor value on gauge E"
Design level	Example of level of detail for tasks in scenario											
Effect	Like in scenario type B (a more detailed scenario is not possible for this design level)											
Usage	"Separate radioactive particles from liquid waste", "Execute administrative tasks", "Handle disturbances"											
Architecture	"Filter water from tank A to tank B", "Test water in tank C"											
Interaction	"Press button D", "Monitor value on gauge E"											

⁸ Largely influenced by Andersen, S. N. (2016) Participatory simulation in hospital work system design. Technical University of Denmark. (Dissertation).

4.1.3 Discussion questions

The discussion questions are meant to initiate a discussion about important aspects of the design proposal that may not have been considered in the scenario walkthrough. This part is particularly important for scenario type C (task sequences), where it might be easy to focus on simulating/describing the detailed tasks and not as much on discussing the solution as a whole.

One category of aspects that may not be sufficiently considered in the scenario walkthrough is the users' experiences, perceptions and feelings associated with using the design and how appropriate the design is for its intended use. User experiences are an important source of information about the suitability of the design, and can be used to capture more subtle issues that were not expressed during the scenario walkthrough⁹.

Another category of aspects worth considering as a basis for discussion questions are principles for the design of the technical system (such as the operator interface). Such principles (design guidelines) are concrete design advice based on knowledge of human abilities and limitations. The use of design guidelines as the basis for discussion questions is a way to utilise existing knowledge in the evaluation activity. A decision on which guidelines to use as a basis for discussion questions should be coordinated with the guidelines used in the heuristic evaluation. For example, it may be appropriate to base the discussion questions on more general guidelines that are difficult to evaluate in the heuristic evaluation. However, some guidelines may be regarded as so important that they should be reviewed in both evaluation workshops.

Discussion questions should be formulated specifically to suit the design proposal to be evaluated. Some examples of discussion questions¹⁰:

- Does this design minimise the risk of persons being harmed or subjected to harmful substances? If not, why not?
- Does this design support an appropriate workload? (One that does not lead to overload, problems with vigilance/wakefulness and attention, or loss of skills.) If not, why not?
- Does this design consider the physical limitations and possibilities of you as a user? (Sight, hearing, space, reach, movement.) If not, why not?
- Is the design consistent? (E.g. considering the way tasks are to be executed and how objects are placed.) If not, why not?
- Does the design proposal allow flexibility in how tasks are executed? Can this flexibility lead to negative consequences? If not, why not?
- Does this design offer suitable functionality to monitor and control the process? Are you able to do what is needed to manage the process? If not, why not?
- Does this design allow you to easily and without effort monitor and control the process? If not, why not?
- Does this design allow you to work in accordance with procedures and regulations? If not, why not?

⁹ For a longer discussion on the benefits of employing user experience as an indicator for assessing complex socio-technical systems, such as control room systems, see 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.

¹⁰ Based on 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) and design guidelines from NUREG-0711 (see earlier footnote for full citation).

4.1.4 Moderator support

To facilitate the moderator's steering of the workshop, it may be beneficial to prepare a document that provides an overview of what to do during the workshop. Suitable support makes it easier for the moderator to assess how the workshop progresses (for example, how many scenarios have been reviewed) and how much remains to go through. Such support for the moderator should, for example, provide an overview of:

- All the steps in the workshop procedure
- Different parts of the design proposal
- All the scenarios
- Discussion questions

Gathering all this information onto a single sheet of paper provides a good overview. See the Template chapter at the end of this document for an example (grey fields indicate parts that must be specified for each evaluation).

4.2 Execution

The procedure in this section is written for the moderator's perspective.

4.2.1 Procedure workshop

- 1) **Introduction** workshop
 - a. **Purpose of the workshop:** Explain the purpose of the evaluation activity to the participants.
 - b. **Desired result:** Explain to the participants that the aim is to identify discrepancies, identify new requirements to further detail the solution, identify positive aspects, and identify possible ways to resolve identified discrepancies.
 - c. **Roles and tasks during the workshop:** Explain to the participants that everyone has a responsibility to critically review the design proposal, and to review it as objectively as possible. As a moderator (or designer or project leader), it is important that you do not try to defend the design proposal, try instead to focus on the desired impact of the design and on whether the design proposal will contribute to this. Remind the designer and the project leader that their primary task is to answer questions. Encourage the participants to take individual notes to remember their thoughts when they cannot be expressed out loud, such as resolutions to discrepancies in the scenario talkthrough.
- 2) **Description of design proposal**
 - a. **Desired impact** of the design: Describe the impact the proposed design is meant to have on its environment.
 - b. **Design proposal:** Describe the design proposal.
- 3) Scenario talkthrough. *For each scenario:*
 - a. **Describe the scenario** (user representatives should have the scenario description on paper as well)
 - b. **Ask the user representatives to handle the scenario using the system representation and think aloud what they do and why.** This is especially important for scenario types A and B ("On the spot" scenarios and case stories), where there is a risk that the participants may not simulate/describe the tasks in the proposed design and only discuss the scenarios. Keep the design level to be evaluated in mind and try to make the user representatives execute tasks at a level of detail relevant to the design decisions at this design level. To simulate tasks, it

may for example be helpful to move LEGO™ figures on a paper drawing, or to walk between paper drawings of operator interfaces set up in a room while interacting and communicating with other users as in the real situation.

As a moderator, it is beneficial to *show* the participants how you want them to simulate tasks, and make sure all user representatives get to voice their opinion.

- c. **Note discrepancies/new requirements/positive aspects on Post-its** – they are not to be spoken aloud during the actual scenario talkthrough! (This may inhibit the flow of the simulation of tasks.)
 - d. **Discuss, together in the group, identified discrepancies/new requirements/positive aspects connected to the scenario** (HF specialists go through their notes last to minimise influence on the other participants). Note according to step e below.
 - e. **Note** (done by the participant appointed to take notes):
 - i. **discrepancies in the design proposal**
 - ii. **new requirements/ideas that are relevant for the further development of the design** (but which are not direct discrepancies in the design)
 - iii. **positive aspects in the design proposal**
- 4) (*When all scenarios are reviewed*) **Discussion questions.** If a question has already been discussed sufficiently during the scenario talkthroughs there is no need to ask it again. If the design proposal is large in scope discussion questions may be asked for each part of the design to make the question more manageable.
- a. **Ask discussion questions** [according to the selection described in section 4.1.3]
 - b. **Note** (done by the participant appointed to take notes):
 - i. **discrepancies in the design proposal**
 - ii. **new requirements/ideas that are relevant for the further development of the design** (but which are not direct discrepancies in the design)
 - iii. **positive aspects in the design proposal**
- 5) (*When all scenarios are reviewed*) **Prioritise identified discrepancies.** This prioritisation should consider how severe the consequences of identified discrepancies may be and how often these consequences may occur. The importance here is to facilitate a discussion about the severity of the discrepancies in relation to each other (for example “more serious than ...” / “less serious than ...”) to provide a basis for discussions about how to handle the discrepancies in other forums. No consideration should be given during the workshop as to how difficult it would be to resolve the discrepancies.
- 6) (*When all scenarios are reviewed*) Note **ways to resolve** identified discrepancies. This should be done separately when all scenarios are reviewed so as not to inhibit the identification of deficiencies. The identification of possible resolutions does not have to be comprehensive and deep, but obvious resolutions should be noted.

Overview scenario-based talkthrough

