NKS-332
ISBN 978-87-7893-413-0

Measuring Procedure Competence. Final Report from the NKS-R(14)112/13

Eitrheim, M. H. R.[1], Holmgren, L.[1], Savioja, P.[2], Hildebrandt, M.[1]

[1]IFE, Norway

[2]VTT, Finland

March 2015

Abstract

This report describes the development of an expert rating tool for measuring procedure competence. The Procedure Competence (ProCom) tool is organized around four overall competences: procedure planning, procedure execution, utilization of backgrounds, and adaptability. These are operationalized through behavioural markers, i.e., observable, predefined performance indicators. The ProCom tool was tested in the Halden Man-Machine Laboratory and in a Swedish training simulator. The test applications indicate that the ProCom tool provides an extensive approach for observing the use of emergency operating procedures. The ProCom tool is also supported by findings from a Finnish empirical study of emergency procedure usage. As procedure competence is closely related to other basic technical and teamwork skills, future applications of the ProCom tool should consider these aspects by expanding the tool or using complementary measures. Recent eye tracking technologies also seem promising for easily combining expert ratings with detailed performance data on how operators scan information in the control room.

Key words

Emergency operating procedures, competence measurement,
expert rating, eye tracking

NKS-332
ISBN 978-87-7893-413-0

Electronic report, March 2015
NKS Secretariat
P.O. Box 49
DK - 4000 Roskilde, Denmark
Phone +45 4677 4041
www.nks.org
e-mail nks@nks.org

IFE/HR/F – 2015/1623

Measuring Procedure Competence
Final Report from the NKS-R(14)112/13

Maren H. Rø Eitrheim[1]
Lars Holmgren[1]
Paula Savioja[2]
Michael Hildebrandt[1]

[1]IFE, Norway
[2]VTT, Finland

Institute for Energy Technology

Address	KJELLER NO-2027 Kjeller	HALDEN NO-1751 Halden	Availability
Telephone	+47 63 80 60 00	+47 69 21 22 00	
Telefax	+47 63 81 63 56	+47 69 21 22 01	

Report number	Date
IFE/HR/F-2015/1623	2014-12-31

Report title	Revision number
Measuring Procedure Competence	

Client	Number of pages
NKS	58

Client reference	Number of issues
NKS_R_2014_112	

Summary	Distribution
This report describes the development of an expert rating tool for measuring procedure competence. The Procedure Competence (ProCom) tool is organized around four overall competences: procedure planning, procedure execution, utilization of backgrounds, and adaptability. These are operationalized through behavioural markers, i.e., observable, predefined performance indicators. The ProCom tool was tested in the Halden Man-Machine Laboratory and in a Swedish training simulator. The test applications indicate that the ProCom tool provides an extensive approach for observing the use of emergency operating procedures. The ProCom tool is also supported by findings from a Finnish empirical study of emergency procedure usage. As procedure competence is closely related to other basic technical and teamwork skills, future applications of the ProCom tool should consider these aspects by expanding the tool or using complementary measures. Recent eye tracking technologies also seem promising for easily combining expert ratings with detailed performance data on how operators scan information in the control room.	Library, NKS

Keywords: Emergency operating procedures, competence measurement, expert rating, eye tracking

	Name	Date	Signature
Prepared by	Eitrheim, M. H. R., Holmgren, L., Savioja, P., & Hildebrandt, M.	2015-15-01	
Reviewed by	Sashidharan Komandur, AaUC	2015-01-29	
Approved by		2015-01-01	

Contents

1 Preface

1.1 Interfaces to other research projects

The NKS project Measuring Procedure Competence utilizes experiences from two areas of work in the international research program at IFE Halden (Halden Reactor Project, HRP): (1) Training of Control Room Operators: Assessment and Improvement, and (2) Resilient Procedure Use.

A Team Self-Assessment Tool (TESA) has been developed to allow control room crews assess their own level of competences (Holmgren & Skjerve, 2014). The TESA tool includes basic technical competences and teamwork competences during normal operation, emergencies and outage. A subset of the basic technical competences during emergencies was selected as a starting point for the development of the Procedure Competence (ProCom) tool in this NKS project.

Resilient Procedure Use is a study of staffing and support tools for knowledge-based operator actions in complex scenarios (Eitrheim et al., 2013; Hildebrandt et al., 2015). The simulator study in HAlden Man-Machine LABoratoy (HAMMLAB) provides rich examples of how emergency operator procedures are applied in complex and unexpected situations. We're utilizing simulator runs from this study to test and refine the procedure competence tool. Snapshots from the simulator runs are also used to illustrate the procedure competences. The experiences from using eye tracking in the study of procedure use are summarised in a separate chapter and illustrated in a video.

Within the Finnish National Research Programs on Nuclear Safety (SAFIR), VTT has performed detailed analyses of NPP operating crew behaviour in simulated accident scenarios. Their experiences from identifying and characterising variations in operating practices are summarized and compared to the suggested ProCom tool.

1.2 Outline of the report

The report provides a short introduction to the purpose of emergency operating procedures in nuclear power plants and the suggested approach for measuring operators' procedure competence. The development of the ProCom tool is described, followed by experiences from applying the tool at KSU Ringhals in Sweden and at the HAMMLAB at IFE, Norway. The chapters 4 and 5 summarize the adjustments that were made in accordance with these experiences and the results from the workshop with practitioners. The report also includes examples of the suggested behavioural markers for measuring procedure competence (chapter 6) and a comparison with the results from a study by VTT (chapter 7). Finally, experiences with using eye tracking for studying procedure use is summarized (chapter 8) and illustrated in a video.

The ProCom tool is included in Appendix2. Appendix 3 provides a short handbook intended for instructors and other users of the ProCom tool and eye tracking methods.

2 Introduction

2.1 The purpose of NPP emergency operating procedures

Nuclear power plant (NPP) emergency operating procedures (EOPs) are pre-designed plans to cope with a wide range of demanding and hazardous situations. The purpose is to increase operator performance by standardizing how work is conducted and reduce task complexity. The procedures are expected to lessen the cognitive burden on the operators by preventing them from solely relying on their memory and experience.

According to Roth et al. (1994) there are three prevailing views on the scope of the EOPs and the how they should be handled by the operators, known as procedure adherence:

1. The EOPs provide detailed guidance for all events, and the operators are required to follow the procedures by letter. Autonomous operator actions are only allowed when explicitly requested.
2. The EOPs provide detailed guidance to avoid or minimize core damage, but do not propose optimised responses for all situations. Autonomous operator actions are not required, or even desired, to minimize the risk of core damage.
3. The EOPs provide a systematic approach to prevent core damage, but operators' knowledge is needed in situations not fully covered by the procedures.

Whereas the first and second view advocate for strict procedural adherence, the third view acknowledges the need for autonomous cognitive activities by the operators, such as building a situation understanding and suggesting alternative response strategies in situations that are not fully covered by the procedure.

While executing pre-planned response actions specified by the procedure, the operator needs to assess their appropriateness and effects, taking a bird's eye perspective of the progress. Whenever a gap is detected between the planned response (e.g., presumptions in the procedure) and the current situation, the operator needs to decide whether the course of actions is still valid or should be altered. Any initiative and activities outside the procedure could be seen as temporary, as the operator acts autonomously until a procedure is again found (Dien, 1998). The phases of procedure following are illustrated in Figure 1 below.

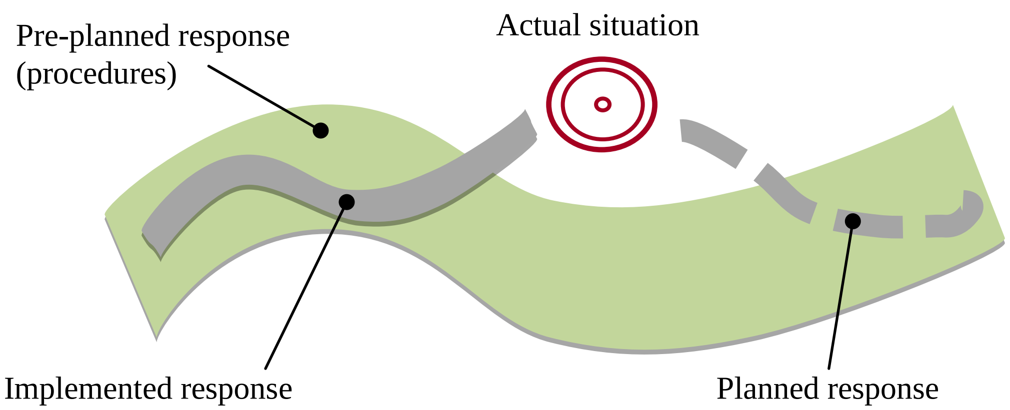

Figure 1 illustrates a temporary drift from the procedure path until a new procedure is found.

Thus, following an emergency operating procedure involves multiple competencies: knowledge and skills necessary to accomplish the pre-planned response, attentional efforts in reading and executing the sequence of actions, monitoring of the procedure effectiveness and decisions on the future path of actions in the current or other procedures.

In the field of human performance, much effort has been made to improve operators' execution reliability during emergencies, including means for reducing errors and ensuring procedural adherence. As indicated above, the operators are expected to simultaneously perform prescribed, detailed sequences of actions and maintain a global overview of the situation. This could be a subtle balance: On one hand, the operators may not have a complete understanding of the situation and at worst fail to perform fundamental actions. Strict procedural adherence would overcome such weaknesses. On the other hand, if solely focused on the task at hand, the operator will probably be vulnerable for surprises. A certain distance to the ongoing work is needed in order to detect errors in the procedure execution and divergence between the procedure and the actual situation. The cognitive challenge of changing an established strategy in light of new information has been widely discussed in terms of fixation, garden paths, and related concepts (Woods, 1984; Roth, Woods & Pople, 1992). The process of changing strategy is especially important for nuclear process control where the direction of work can be highly constrained once a procedure has been chosen.

Thus, the duality between strict procedure adherence and independent assessment of the procedure applicability should be reflected when evaluating the operators' procedure competence. Below we will provide a general definition of procedure competence and how it can be measured through observation of simulated control room work.

2.2 Procedure competence

In this project we look at how control room crews apply emergency operating procedures. By procedure competence, we mean their ability to combine procedure skills, knowledge and attitudes *in practice* to handle emergency situations in an effective and efficient manner, and according to specified plant standards (IAEA, 2006). These competencies may be developed through a combination of education, experience, and training.

Due to the limited scope of this project, and its practical, user centred approach, the suggested rating tool is restricted to cover the practical application, or use, of emergency operation procedures, and not related and possibly overlapping technical and teamwork skills necessary to handle emergency situations, e.g., to build a situation understanding through the use of redundant plant indications, understand the plant implications, communicate and coordinate activities within the team.

2.3 Measuring procedure competence

One way of measuring procedure competence is to ask an external observer to assess the actual performance of the crew, so-called expert rating. For the 'non-observable' aspects of performance, the observer needs to draw conclusions based on the verbal exchange and reflections in the crew, in addition to the actual activities performed. For example, the ability to choose an optimal procedure strategy might be observed through the actual procedure applied and discussions of its purpose and appropriateness in the crew. These are the behavioural markers of the procedure strategy competence of the crew, i.e., the concrete, observable behaviours that will be rated by the observer.

The main motivation for identifying behavioural markers is to improve the reliability in observing and evaluating procedure competence, i.e., promote stable, consistent and precise assessments across observers. The behavioural markers may also serve as guidance on what concrete competences and practices the operators should hold and maintain through simulator training.

2.4 Development of a tool for measuring procedure competence

The goal of the project is to develop a simple-to-use, reliable method for measuring procedure competence using behavioural markers. The scope and limitations of the suggested tool are summarized below:

- The tool is aimed at evaluating the use of paper-based emergency operating procedures (EOPs). However, a similar tool could be developed for other types of control room procedures.
- The tool is developed for pressurized water reactors (PWRs) and needs to be adapted if used in boiling waters reactors (BWRs)
- The current version is based on Westinghouse EOPs for a Swedish NPP. There may be plant specific differences that should be addressed before application in other plants.
- The tool assumes that the operators and the observer are familiar with the EOPs, related background information and the plant process.
- The tool is generic and aimed at covering a variety of events requiring the use of EOPs. The different phases of handling these events including search for the appropriate procedure, use of background information, procedure execution and monitoring of its effectiveness should be reflected in the design of scenarios. Scenario recommendations will be included as a part of the handbook.
- To keep the tool as simple and practically applicable as possible, the behavioural markers are stated on a crew level. Based on the observations made, proposals for improvements can be provided individually and for the whole team.
- Related operator skills and performance, such as basic technical competence and teamwork are not a part of the tool.
- The tool could be useful for a wide range of applications, including training, Human Reliability Assessment (HRA), development of operator support tools, revision of procedures, and integrated system validation (ISV).

3 Development of the Procedure Competence (ProCom) tool

The chapters 3, 4 and 5 describe the development of the Procedure Competence tool, which included the following activities:

- The criteria and selection of procedure competences defined in a Team Self-Assessment Tool (TESA).
- Identification of behavioural markers related to procedure competences
- Language refinements
- Layout of the tool, scoring and user instructions
- Test application in HAMMLAB and at KSU, Sweden
- Improvement of the tool based on the test applications
- Examples illustrating the behavioural markers

3.1 Defining procedure competences and identifying their behavioral markers

The Team Self-Assessment Tool (TESA) includes basic technical competences and teamwork competences during normal operation, emergencies and outage (Holmgren & Skjerve, 2014). A subset of categories suggested as basic technical competences during emergencies was selected as a starting point for development of the Procedure Competence (ProCom) tool. The main criterion for selecting a category was the actual handling of procedures when performing the proposed activity. The original TESA categories and descriptions are included in Appendix 1.

Among the ten categories within emergency operation, four categories were included for procedure competence evaluation: procedure strategies; use/knowledge of procedure backgrounds; effective procedure abilities; and resilient procedure use. Table 1 below shows the four TESA categories of EOP handling used for development of the procedure competence tool.

Table 1: The four categories of EOP handling from the TESA used as a starting point for developing the procedure competence rating tool.

Category	**Details**
Procedure strategies	− Strategies are familiar by the crew − Optimal Procedure strategy is reflected
Use/knowledge of procedure backgrounds	− Important Step Basis used and known − Backgrounds used at decision making − Knowledge parts are known and used − Key Decision Points are familiar
Effective procedure abilities.	− Procedure entering with main goals is announced and understood by the crew − Procedure Use is thoughtful − Plant monitoring and control is effective − Procedure Following not literal and get stuck − Briefing requested if not Understood − Questioning attitude is kept
Resilient procedure use	− Actions and Plant Responses followed up − Preparedness for the unexpected − Alternative Strategies developed if necessary. − Safe handling of the unexpected

For each of the four categories above, we identified overt behaviour that could indicate procedure competence. The process of identifying and describing these behavioural markers was accomplished through informal interviews and discussions between a nuclear process expert and a human factors researcher. Typically we raised questions such as “How do we know that the crew is familiar with the procedure strategy?” and “How could we observe that the crew keeps a questioning attitude?” During this process, we also rephrased and added new aspects within the four categories of competence.

The tool was organized in a hierarchy of three levels: competences, rating factors and behavioural markers. “Competences” correspond to the four categories of EOP handling in Table 1 above. Within each competence, we defined more detailed rating factors. For example, the competence “Procedure execution” (replacing “Effective procedure abilities” in Table 1 above) covers rating factors such as ensuring sufficient progress, applying fold out pages and transferring to other procedures. Each rating factor was then broken down to one or more behavioural markers such as discussing procedure strategy in the crew, announce changes of criteria in fold out pages and explain transfer to a new procedure in own words The wording was kept as short and concise as possible, stating the target of the observations early in the utterances to promote readability. The utterances were consequently phrased in a positive wording that indirectly covers the erroneous behaviour. For example, to execute minor steps timely and safely implies that the operators ensure sufficient progress without compromising execution accuracy, such as missing steps or performing steps in the wrong order. Finally, we revised the order of procedure competencies, rating factors and behavioural markers with the purpose of following a typical sequence or time line during the handling of an emergency event. The original TESA categories, the revised competences for the ProCom tool, ratings factors and behavioural markers are listed in Table 2 below.

Table 2 shows the revised procedure competences, rating factors and behavioural markers in the first version (v1.0) of the ProCom tool. The original TESA categories are included in the first column to the left.

TESA	ProCom v1.0	ProCom Rating factor v1.0	ProCom Behavioural marker v1.0
Procedure strategies	Procedure planning	Choose optimal procedure strategy	The applicability of the procedure strategy is discussed
		Understand procedure purpose	Procedure goal and main actions explained in own words
Effective procedure abilities	Procedure execution	Ensure sufficient progress	Major action steps discussed and given high attention
			Minor steps timely and safely executed
			Briefings performed if steps are unclear
			Look ahead and prepare subsequent procedure steps
		Apply fold out pages	Changes of criteria in fold out pages timely announced
		Monitor Critical Safety Functions (CSF)	Changes of CSF timely alerted and actions started. If red condition: immediately.
		Correct transfer to other procedures	Transfer to a new procedure is explained
Use/ knowledge of procedure backgrounds	Utilization of backgrounds	Apply additional background information	Relevant background information explained
			If background applicable, steps are executed accordingly
		Handle Key Utility Decision Points (KUDP)	KUDP used for decision making when needed

TESA	ProCom v1.0	ProCom Rating factor v1.0	ProCom Behavioural marker v1.0
Resilient procedure use	Adaptability	Flexible execution of procedure	If step(s) not applicable, alternative actions performed
		Evaluate procedure effectiveness	Discuss if current procedure strategy brings plant to safe state
		Anticipate possible outcomes	Discuss and predict long term consequences
		Independent thinking	Compare different procedure strategies and main actions
			Agree on final goal and start actions to achieve safe state
		Monitor goal achievement progress	Discuss strategy progress and effectiveness

Procedure planning concerns choosing an appropriate procedure, and understanding its purpose and main goals. As described in the introduction, the nuclear process control planning and overall goals are strongly guided by the operating procedures. Accordingly, the operators should verify whether a procedure provides the most optimal strategy for a given situation. In complex and novel situations, the operators may need to consider multiple procedures to be applied in a given situation, their premises and overall goals, and choose the most optimal strategy provided by one procedure or a set of procedures. To achieve safe and efficient execution of the strategy chosen, the evaluation of procedure planning also suggests that the operators express the goals and actions in their own words before entering the procedure. This could enhance the understanding of the most important procedure steps to be fulfilled and why.

Procedure execution is the realization of the procedure planning in terms of ensuring performance progress, monitoring of criteria in fold-out pages and monitoring of the critical safety functions. The behavioural markers describe the optimal and expected way of executing a procedure. Thus, insufficient or erroneous actions, such as missing or misinterpreting procedure steps, are reflected indirectly through low rating scores on the behavioural markers. In accordance with the procedure plan, the operators should look ahead and prepare procedure steps whenever suitable, give major steps high attention and execute minor steps timely and safely. When steps are unclear, the operators are expected to perform briefings to clarify the step purpose, applicability and implications.

Each emergency operating procedure has its own background information document that includes information about analyses that were realized to develop the strategy of the procedure, information about the physics of the accident the procedure is supposed to deal with, and a detailed explanation of each procedure step. The operators are supposed to be familiar with the background information, explain the procedure steps and execute them according to the background information when relevant. In simulator training, the operators may also make use of the key utility decision points (KUDP) to support their decision making. The KUDPs indicate conditions in which the utility must determine an appropriate course of action.

The fourth area of competence, adaptability, concerns evaluation of the procedure effectiveness, predictions of long term outcomes, and flexible execution of procedure steps and strategies in situations where the procedure or parts of it cannot be applied as suggested. If minor deviations are detected, these may be handled without affecting the overall progression, for example by performing alternative actions while maintaining the overall goals of the procedure. If the strategy of the procedure need to be changed or cannot be accomplished, the operators need to define an adjusted strategy that meets the new situational demands. In such cases, the operators are expected to discuss whether the current strategy brings the plant to a safe state, compare different procedure strategies and main actions, and agree on a final goal. Depending on the length of the scenario, the operators may start actions in the current or other appropriate procedures to achieve a safe state.

3.2 Scoring of the behavioural markers

The behavioural markers suggested in Table 2 are evaluated for the whole crew. The observer is asked to rate their performance on a 5-point scale. For example, to what extent did the crew discuss if the procedure strategy brought the plant to a safe state? The observer is asked to rate this behaviour and indicate the level of competence observed from 1 (very weak) to 5 (very strong), see Figure 2 below. Many strengths with no or minor weaknesses that will impact safety of the reactor core and the public are considered to reflect strong or very strong performance. Several weaknesses that are considered relatively insignificant typically reflect satisfactory performance. Many and significant weaknesses are considered to reflect weak or very weak performance.

1	2	3	4	5
Very weak	Weak	Satisfactory	Strong	Very strong

Figure 2 shows the 5-point rating scale for the behavioural markers.

The ratings should reflect the overall impression of the behaviour during a scenario. Depending of the specific event and the actual progress in the scenario, the ratings could be based on only one instance or several instances during the time of observation. Some behavioural markers may not be applicable, e.g., the scenario doesn’t require any use of background information. If a behavioural marker is not applicable during the scenario, the observer is asked to indicate N/A in the scoring sheet. N/A means that the behavioural marker could not be evaluated and should be disregarded in the calculation of procedure competence scores.

For the competences on Procedure planning and Utilization of backgrounds, the competence grades are the average of the rating factor scores. For the other two procedure competences, Procedure execution and Adaptability, there are several behavioural markers. As some of the behavioural markers are expected to have stronger plant safety implications, the scoring is weighted to emphasize and reward the most important behaviours. For example, the notification of changes in critical safety functions (e.g., subcriticality, core cooling) is emphasized above the quality of executing minor procedure steps. The calculation of weighted scores is explained in Figure 3 below.

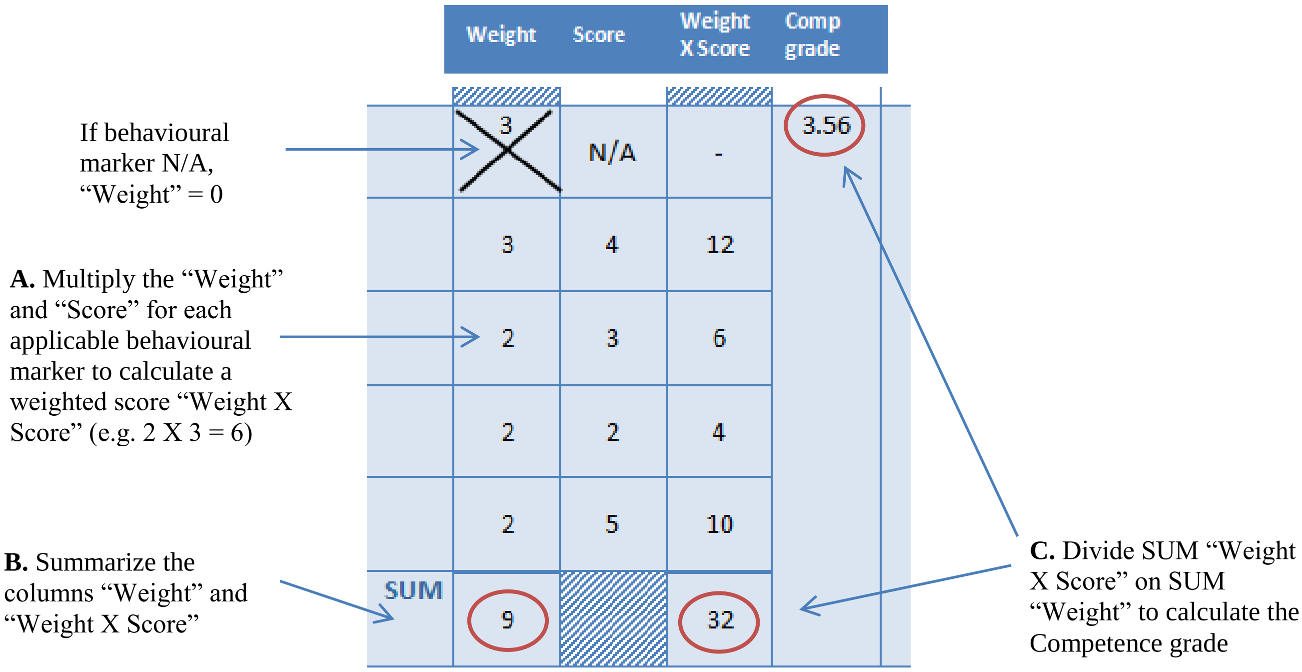

Figure 3 shows an example of how the weighted scores are calculated.

3.3 The tool layout

The four competences, the rating factors and the behavioural markers are organized into an A3 scoring and observation sheet. In addition to the numerical scoring, observations and proposals for improvements can be documented by making hand-written notes online. The tool could also be accessed as a spreadsheet on a computer for automatic calculation of competence grades and additional functionality such as logging of observation times (time stamps). The paper version of the procedure competence tool is included in Appendix 2 (ProCom v2.0, revised in line with application experiences described in chapter 4).

4 Application of the tool in HAMMLAB and a Swedish training simulator

4.1 Experiences from internal tests by IFE process experts

The ProCom tool was applied and reviewed by two nuclear process experts at IFE. The process experts tested the tool both online during accident scenarios in HAMMLAB, and by looking at video recordings of previous accident scenarios in HAMMLAB. The accident scenarios were performed on the RInghals Pressurized water Simulator (RIPS) with participating crews from Sweden and the US (Eitrheim et al., 2013; Hildebrandt et al., 2015). In addition, one of the process experts and the human factors researcher looked at video recordings of accident scenarios to provide examples of the behavioural markers, see chapter 6. The main lessons learned from the internal test applications are summarized below:

- The use of EOPs is tightly coupled to observations and interpretations of process information to achieve an adequate situation understanding and mitigate deviations and disturbances. The operators need to have basic technical knowledge in order to understand the plant responses and apply the procedures in an appropriate way. The scope of the ProCom tool is limited and does not consider these performance aspects when evaluating procedure competence.
- It can also be difficult to separate procedure competence from teamwork aspects affecting the procedure handling such as leadership and communication in the crew. For example, one operator might suggest alternative actions to what are prescribed in the procedure, but the initiative is not followed up or discussed in the crew.
- The basis for evaluating the procedure competence is closely related to the scenario content and complexity. In the more straightforward or uncomplicated scenarios, the behavioural markers may only be observed once during the scenario, and the procedure handling requires less discussion in the crew. In more complex scenarios, the crew may need to change strategy and continuously discuss the applicability of steps and procedures repeatedly. In the latter case, the observer might get a more nuanced understanding of the crew’s competences. This may also be achieved when observing the crew across a number of scenarios.
- The application of the tool requires that both observers and operators have sufficient knowledge of the plant process and the procedures available. Some of the rating factors were not applicable in the HAMMLAB scenarios observed as the participants had limited knowledge about the plant process or the plant procedures.

4.2 Experiences from testing the ProCom tool at Ringhals KSU

The ProCom tool was applied by three instructors at Ringhals KSU during six weeks of simulator training. The experiences gained in this period were shared with IFE in a meeting in the spring and in a workshop at IFE in the autumn. The main lessons learned from the test applications at the training simulator are summarized below:

- Overall, the ProCom tool was well received by the instructors. The tool covers aspects of procedure competence that are not currently considered in the simulator training or the feedback to the operators. Thus, the tool provided new ideas and perspectives for the instructors both on what should be evaluated, and how this could be achieved in terms of concrete behaviours to be observed during simulated accident scenarios.
- The ProCom tool could be a good starting point for aligning the performance evaluation across instructors. Currently, the evaluation is highly subjective, and the documentation of

what performance aspects should be evaluated and the criteria for these is sparse. The ProCom tool encourages discussions between instructors, and could also be a tool for communicating expected skills and competences to the operators.

- The instructors are not used to numerical scoring for evaluating the crew performance. They felt uncomfortable with writing scores and calculating competence grades based on the performance observed, but agreed that this could be helpful to provide precise and specific feedback to the operators and a support for organizing future training in accordance with the improvement needs identified.
- Similar to what the IFE process experts experienced in HAMMLAB, the instructors pointed to the challenges with evaluating procedure competence in isolation, without considering teamwork aspects or other technical skills. The instructors also noticed that some of the rating factors were highly dependent on the scenario complexity.
- The instructors would like to have more space for making comments online. For some of the rating factors, they found it useful to make notes and suggest preliminary scores that could be adjusted later in the scenario. One instructor also emphasized the need for individual feedback to the operators, which would require more space for documenting improvement proposals.

4.3 The test scores collected at IFE and KSU Ringhals

The data material from testing the ProCom tool covers 15 ratings of crews in HAMMLAB and the training simulator at KSU Ringhals. Direct comparisons of ratings by multiple observers can only be made in two cases:

1. One process expert from IFE and the deputy unit manager applied the tool during a competence assessment scenario in the KSU Ringhals training simulator (2 ratings)
2. One process expert from IFE and two instructors from KSU Ringhals applied the tool for evaluating a US crew in HAMMLAB (3 ratings, see chapter 5).

The remaining 10 ratings were performed for different crews acting in different scenarios. Thus, the inter-rater reliability, i.e., the agreement between raters using the tool, cannot be fully evaluated based on the data we have so far.

4.3.1 Application of the ProCom tool in a competence assessment scenario at KSU Ringhals

One process expert from IFE and the deputy unit manager applied the tool during a competence assessment scenario in the KSU Ringhals training simulator. Procedure planning was only rated by one observer, and will not be discussed here. The procedure execution was rated as 3.06 (satisfactory) by one observer, and 3.46 (satisfactory – strong) by the other observer. The detailed scores for procedure execution are illustrated in Figure 4 below. The weighting of the behavioural markers are indicated in brackets below the horizontal axis. Both observers rated the behavioural markers on looking ahead and using fold out pages as satisfactory (3). Observer 1 (blue columns) rated the execution of both major and minor steps as better than Observer 2 (red columns), while Observer 2 rated the qualities of briefings, monitoring of critical safety functions and transferring to a new procedure as better than Observer 1. Observer 1 reported that he expected more discussions among the operators than observed for this crew. The handling of background information and key utility decision points was rated as weak (2) by both observers. Although variability in ratings of the behavioural markers for Adaptability (see Figure 5), the overall competence grades were similar for the two observers (2.16 and 2.21 – weak).

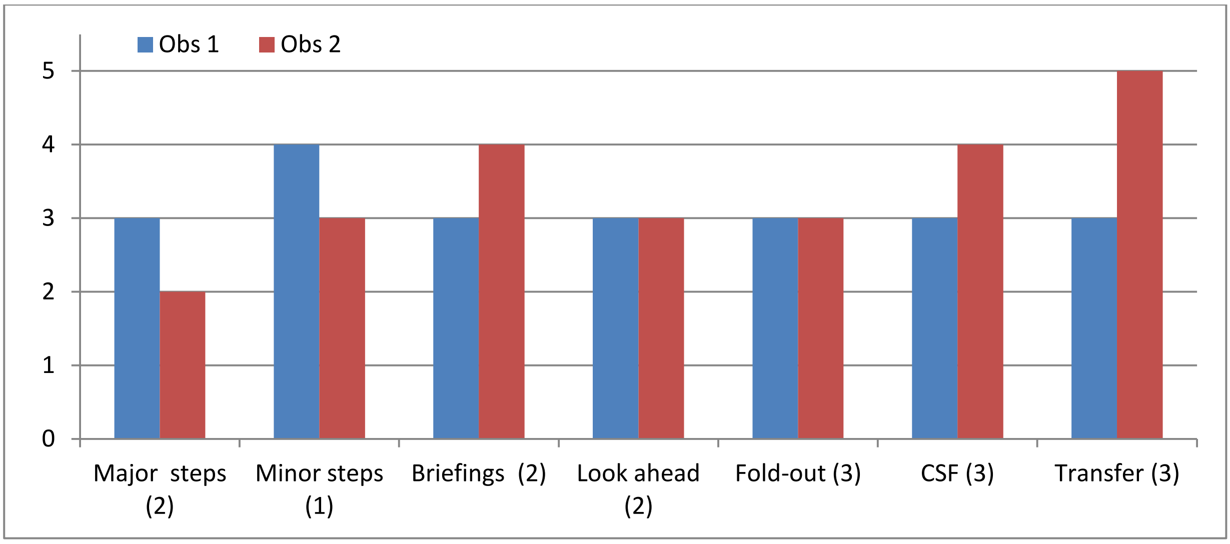

Figure 4 shows the ratings made by two observers (red and blue bars) for the behavioural markers related to procedure execution. The weighting of the behavioural markers are indicated in brackets below the horizontal axis.

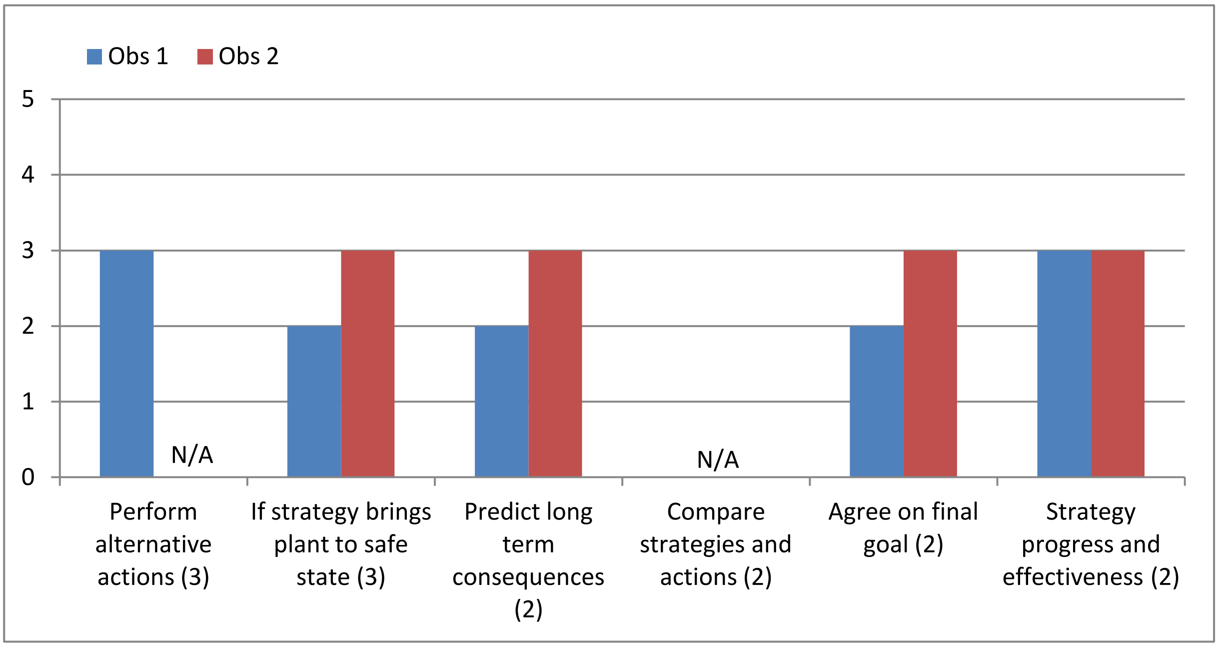

Figure 5 shows the ratings made by two observers (red and blue bars) for the behavioural markers related to adaptability. The weighting of the behavioural markers are indicated in brackets below the horizontal axis.

4.3.2 Application of the ProCom tool in the KSU training simulator and HAMMLAB, 10 different scenarios

The competence ratings described in this chapter were performed in HAMMLAB and the KSU Ringhals training simulator:

- Two process experts at IFE applied the ProcCom tool for observations of Swedish and US crews in HAMMLAB, both online and by use of video recordings. Some of the US crews had limited experience with the simulated plant process and the EOPs. For these crews, the ratings of procedure planning and application of background information were not applicable. The ratings were performed for different crews in different scenarios.

- Three instructors at KSU Ringhals applied the ProCom tool for observations of crews during simulator training. The ratings were performed for different crews in different training scenarios.

The sensitivity of the ProCom tool and the distribution of competence grades are illustrated in two figures: Procedure execution and Adaptability. For procedure execution the grades range from 2.3 (weak) to 4 (strong), see Figure 6 below. Four of ten crews were rated as having less than satisfactory competence (3) on procedure execution. Typically, these crews got lower scores on using fold out pages and monitoring critical safety functions.

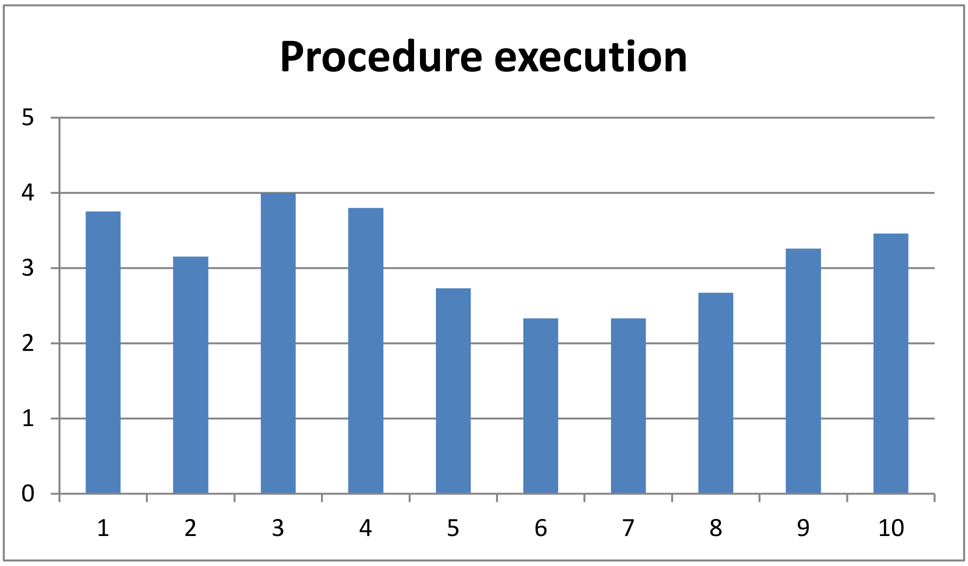

Figure 6 shows the procedure execution competence ratings for 10 crews in different scenarios.

Overall, the competence grades on Adaptability are lower than for the Procedure execution with a range from 1.8 (weak) to 3.1 (satisfactory), see Figure 7 below. Four of ten crews showed satisfactory competence on Adaptability. All crews showed satisfactory or strong competence for the performance of alternative actions when procedure steps were not applicable. However, evaluation of the procedure effectiveness, comparison of procedures and prediction of long term consequences were rated as weak for most of the crews.

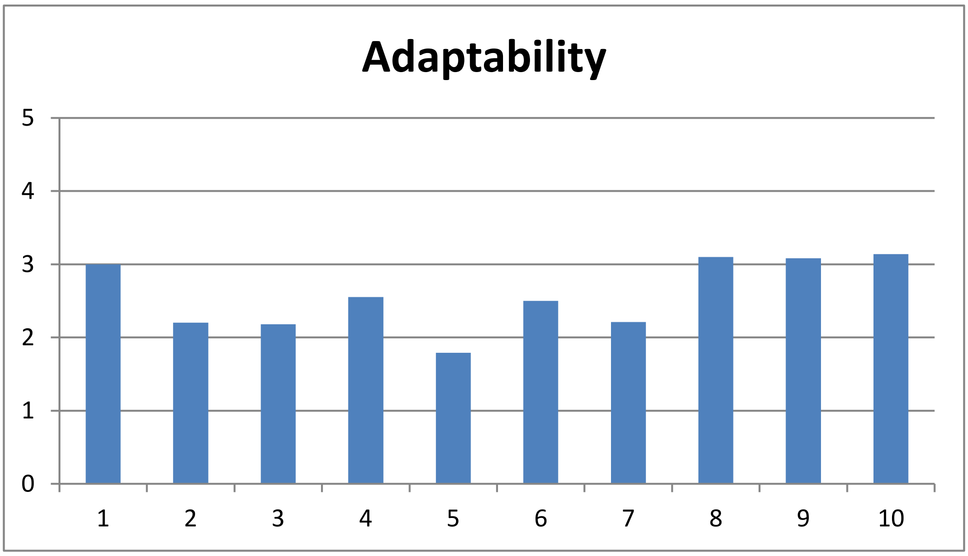

Figure 7 shows the competence ratings on adaptability for 10 crews in different scenarios.

5 Revision of the ProCom tool and results from workshop

Based on discussions with process experts at IFE and instructors at KSU Ringhals we made the following changes to the ProCom tool:

- The rating factor "Correct transfer to other procedures" was merged with the rating factor "Understand procedure purpose". The behavioural marker was rephrased in order to better cover both aspects: "Goal and main actions explained in own words when entering a new procedure".
- The rating factor "Monitor goal achievement progress" was removed, as this overlaps with the rating factors "Evaluate procedure effectiveness" and "Independent thinking". The behavioural marker "Agree on final goal and start actions to achieve safe state" was rephrased to "Agree on final goal and start actions in current or other appropriate procedure to achieve safe state".
- The scoring sheet was increased in size, from A4 to A3 format, including a new column for "Improvement proposals". The main purpose was to provide more space for observations, detailed feedback to the operators and identification of future training needs.

The new version of the the ProCom tool, v2.0, is included in Appendix 2. The new version of the ProCom tool was discussed and tested in a workshop with instructors from KSU Ringhals, a process expert and a human factors researcher at IFE. The instructors and the IFE process expert observed one crew in a HAMMLAB scenario that involved loss of coolant inside and outside containment. The same scenario is utilized to provide examples of behavioural markers below, see chapter 6.1 for a more detailed description of the scenario. The procedure competence was evaluated individually by use of video recordings. After the scenario was completed, the ratings and the reasoning behind them were discussed in plenary.

Procedure planning and utilization of background information were rated as satisfactory by all observers. For procedure execution, the competence grades ranged from 2.6 (weak – satisfactory) to 4.2 (strong). The detailed ratings of the behavioural markers for procedure execution are illustrated in Figure 8 below.

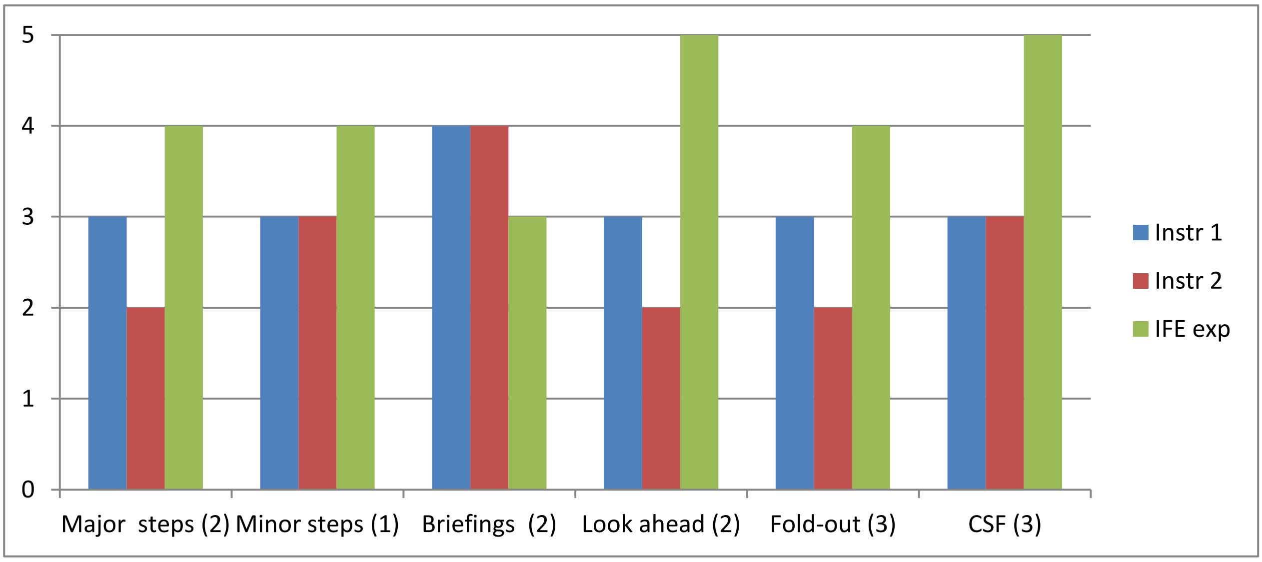

Figure 8 shows the ratings made by two instructors (red and blue bars) and the IFE process expert (green bars) for the behavioural markers related to procedure execution. The weighting of the behavioural markers are indicated in brackets below the horizontal axis.

Overall, the IFE process expert rated the behavioural markers with higher scores than the instructors from KSU Ringhals. One reason for this might be that the IFE process expert was involved in the design of the scenario, had observed the same crew and a number of other crews in the same scenario previously. The KSU Ringhals instructors observed the crew and the scenario for the first time. After observing the scenario, the IFE process expert commented that this was one of the best performing crews, possibly applying a normative approach to the competence evaluation, i.e., compared the performance of this crew with other crews' handling of the same scenario. Although the observers rated the competence based on concrete behavioural markers, they seemed to have different expectations of the extent to which procedure steps and related information should be discussed in this scenario, and whether the crew was looking ahead and prepared subsequent procedure steps in a sufficient way. Thus, the differences in ratings between the observers may be due to their knowledge and familiarity with the scenario, the experiences from observing other crews, the performance criteria and operationalization of these for each of the behavioural markers.

The competence grades for Adaptability ranged from 2.8 (weak – satisfactory) to 3.4 (satisfactory). The detailed ratings of the behavioural markers for Adaptability are illustrated in Figure 9 below. Overall, the ratings were more congruent across observers than for the Procedure execution.

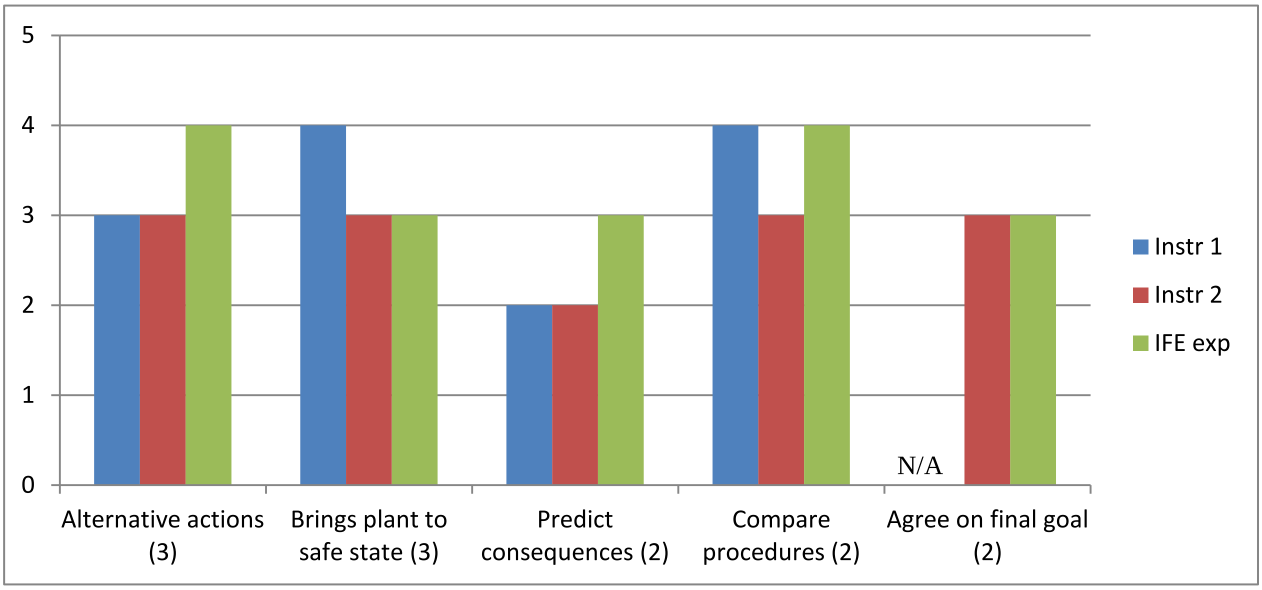

Figure 9 shows the ratings made by two instructors (red and blue bars) and the IFE process expert (green bars) for the behavioural markers related to adaptability. The weighting of the behavioural markers are indicated in brackets below the horizontal axis.

To summarize, the ProCom tool seems sensitive to performance variability, with a majority of the test scores in the range from weak (2) to strong (4) competences. The data material does not allow for any systematic inter-rater reliability investigations, but the experiences so far indicate that there may be some observer variability, especially when assessing the procedure execution. A preliminary recommendation is therefore to organize a pre-scenario briefing to discuss the criteria for evaluating the behavioural markers and any scenario-specific considerations to be made.

In the workshop with the instructors we also discussed that the utilization of background information could be difficult to observe, as the operators are supposed to know this by heart and may apply the information without stating it orally. The applicability of the behavioural markers suggested for utilization of background information will depend on the scenario content and the expectations communicated to the crews for explaining background knowledge. We also discussed whether the competence on background information is better evaluated through interviews with the crews after the scenario.

6 Illustration of the behavioural markers and scenario recommendations

To enhance the understanding of the behavioural markers related to procedure planning, execution, utilization of backgrounds and adaptability, this chapter describes the handling of two scenarios by two different crews in HAMMLAB (Eitrheim et al., 2013; Hildebrandt et al., 2015). The first scenario, Loss of coolant inside and outside containment, covers all behavioural markers except “Briefings performed if steps are unclear”. This behavioural marker is illustrated in the second scenario described, Multiple SG tube breaks. The purpose is not to discuss the quality and scoring of the behavioural markers, but to provide examples of the competences observed with use of the ProCom tool. The order of the behavioural markers reflects the actual sequence of events and progress made in the scenarios. Finally, scenario recommendations for future applications of the ProCom tool are provided in the end of this chapter.

6.1 Loss of coolant inside and outside containment

When the scenario starts, the plant is at 100 % power.

- One feed water pump on turbine 31 is not available due to preventive maintenance.
- Two dump test switches are forgotten in test position, meaning that steam dump is not available.
- One of the RHR/LHSI suction valves on train B is not completely closed and the second valve in series has an induced crack in the sluice disc.

Early in the scenario, one of the two operable feed water pumps on T31 will fail, and the turbine load will be reduced accordingly. The steam dump valves will not open due to the miss-positioned test switches on the steam dump control. At the end of the T31 load reduction, T32 trips on a generator protection signal. Since the dump control is blocked, the reactor coolant system (RCS) temperature and pressure increase rapidly, and the pressurizer and steam generator (SG) relief valves open. The reactor control rods start to decrease the reactor power.

A former induced crack in a RHR/LHSI suction valve develops to a partial internal break on the sluice valve disc. The pressure increases rapidly in train B, the safety valve 8708B opens and starts chattering, releasing hot RCS water to the pressurizer relief tank (PRT) inside the containment. The chattering safety valve causes heavy pressure spikes in the RHR system, in the auxiliary building. This causes a break on the pressure side of the RHR/LHSI-pumps (possible to isolate) and finally the safety valve 8708B is stuck 75 % open (impossible to isolate). Reactor trip and safety injection (SI) will be automatically released. The two breaks initiate alarms for fire and high radiation in the containment and in the auxiliary building.

6.1.1 Behavioural markers observed in the scenario

- *Look ahead and prepare subsequent steps*

The crew is working in E-0, step 5-6. Both turbines and the reactor have tripped. The crew has detected that automatic SI is released due to low RC pressure. While the shift supervisor (SS) is reading the procedure steps loud, the balance of plant operator (BOP) identifies high levels in all three SGs and that the AFW pumps will stop. Thus, the SS decides to manually reduce the auxiliary feed water (AFW) flow to avoid additional cool down. This is an example where the crew looks ahead and prepare subsequent steps, as manual reduction of AFW is introduced later in the procedure (step 14). By reducing the cool-down at this point, they may achieve a stable state earlier in time.

- *Changes of criteria in fold out pages timely announced.*

In step 13, the STA verifies and announces that the criteria for stopping the reactor coolant pumps (RCP) are fulfilled. Stopping the RCPs reduces the risk for damages to the pumps shaft seals.

- *Discuss and predict long term consequences*

In step 22, the crew is checking whether the SG tubes are intact. They have detected alarms on radioactivity and fire in the auxiliary building. Since RHR train B is pressurized and is indicating flow, they agree that they probably have a LOCA outside the containment. This is an example of a crew discussion where the current indications and step responses are summarized, predicting a possible outcome and consequences.

- *Major action steps are discussed and given high attention*

Monitoring of the SGs in step 22 is given high attention by the crew, and discussed thoroughly as explained above. The SS ask the STA to continue monitoring the SGs, while they proceed to the subsequent step in E-0. Sampling of SG may take 5-10 min before getting certain indications.

- *The applicability of the procedure strategy is discussed*
- *Compare different procedure strategies and main actions*

In step 23, the SS orders transfer to E-1, to reduce the break within the containment. Previously, they have also observed indications of a break outside the containment. Thus, they believe that there are breaks both outside and inside of the containment. The crew agrees to initiate isolation steps in ECA-1.2 in parallel with the use of E-1. They compare and discuss the applicability of both procedures in order to stop the release of radioactivity to the atmosphere as early as possible.

- *Discuss if current procedure strategy brings plant to safe state*

The STA anticipates the formal transfer point from E-1 to ECA-1.2, in which they have already started. The crew has not explicitly discussed the effect of applying E-1, but chosen to run the two procedures in parallel.

- *Minor steps timely and safely executed*

The crew communicates effectively all steps consecutively; they don't stop or delay on minor procedure steps. For example, in E-1, step 6, they're asked to check the pressurizer Pneumatic Operated Release Valves (PORVs) and the block valves. None of these checks are related to the main strategy of the procedure, but verifications needed.

- *Procedure goal and main actions explained in own words when entering a new procedure*
- *Changes of CSF timely alerted and actions started.*
- *Relevant background information explained*
- *If background applicable, steps are executed accordingly*
- *If step(s) not applicable, alternative actions performed*
- *Agree on final goal and start actions in current or other appropriate procedure to achieve safe state*

When transferring to a new procedure, the SS should be aware of and may ask the board operators to monitor the Critical Safety Functions (CSF). These should be verified every 15th minute in order to detect adjacent problems beyond the scope of the current procedure. The crew detects an orange indication on integrity immediately. The SS informs that they cannot progress in E-1, and need to transfer to FRP-1 to reduce possible strains to the reactor vessel. According to FR-P.1, cooling of RCS should be stopped and the pressure in the RCS should not be increased. Based on the background information, STA explains that they have a certain cool down of RCS and that the RCS pressure is increasing due to SI flow to the cold legs. Thus, the crew needs to monitor that the RCS conditions don't change beyond what is caused by the SI. They decide to resume to ECA-1.2/E-1 with the restrictions explained above in order to isolate the affected circuits. They also prepare the main steps in ES-1.2, which will be used to cool down to cold shutdown and lower the RC pressure (when feasible), in order to minimize or stop the leakage inside the containment.

- *Key Utility Decision points (KUDP) used for decision making when needed*

A KUDP was not presented in this scenario. These are mainly handled by the Technical Support (TS), but the crew may use a KUDP document to support their decision if the TS is not yet available (within the first two hours of an emergency situation). For example, the crew may get into a situation where they need to decide which SG should be used for cool down when multiple SGs are damaged. The KUDP document describes pros and cons when selecting a SG for cool down, and what should be considered to minimize any negative effects.

6.2 Multiple steam generator tube breaks

There are tube leaks in two steam generators, SG2 and SG3, coincident with stuck open safety valve in SG3. When the scenario starts, the plant is at 100 % power.

- A heavy northwest storm causes extreme high sea water level and much sea weed and grass entering the screen house threatening the main cooling water supply to the turbine condensers.
- The screen house is manned with extra resources.
- The operation manager has ordered to shut down the plant in a controlled manner starting with turbine T31.
- A steam generator (SG) sampling valve on SG3 is forgotten in closed position.

Due to a big amount of sea grass entering the screen house, the overpressure hatch opens and the main cooling water pumps strainers start to clog in 2 minutes. Accordingly, insufficient cooling water reaches the condensers which lead to increasing condensers pressures, load reductions, dump restrictions on high condenser pressure turbine trips and reactor trip. SG PORVs and safety valves open to relief the overpressure to the atmosphere, but one safety valve on SG3 remains stuck open after the pressure transient.

At the transient RCS and SG tube bundles are as well exposed for pressure transients causing tube leaks in SG2 and SG3 which implies decreasing RCS pressure giving automatic Safety Injection. SG 3 has now both an open safety valve and a quite big tube leak implying uncontrolled radiation releases to the public. Since the sampling valve in SG3 is closed, a radiation alarm from SG3 will not be received automatically.

6.2.1 Behavioural marker observed in the scenario

- *Briefings performed is steps are unclear*

In E-3 step 2 Identify Ruptured SG(s), the secondary operator does not understand the SG2 level behavior and announces that he is skeptical that SG2 has a tube leak. He repeats the announcement twice without response from his team. At the third time a briefing is held and it`s concluded that SG2 has a small tube leak.

In ECA-3.1 step 7 Check Ruptured SG(s) level: The reactor operator (RO) is uncertain how he should interpret the step. A kind “Briefing” is held where the question ideas are bandied without coming to any conclusion. The RO annoyed says “Then I continue to the next step.”

6.3 Scenario recommendations for future applications of the tool

As discussed earlier in the report, the applicability of the ProCom tool and the basis for evaluating procedure competence is to a large extent depending on the scenario design, which provides the conditions for observing the crews. For future applications of the ProCom tool, we propose the following recommendations when designing the scenarios and planning the evaluation:

- The ProCom tool considers competences for using emergency operating procedures. This can only be observed in accident scenarios with use of EOPs in major parts of the scenario.
- The scenario should require transfer between procedures. Preferably the operators would need to change procedure at different points in the scenario.
- The scenarios should include situations where the operators would benefit from preparing subsequent procedure steps.
- The scenario design should imply changes of criteria in fold out pages and changes of critical safety functions.
- The scenarios should include challenging situations in which the procedures would not be fully applicable
- The crews should be trained to perform frequent updates or briefings whenever transferring to new procedures, important information is received, or if the situation is unclear.
- Towards the end of the scenario, a discussion might be initiated to clarify what the crew expects regarding the long term consequences and what procedures and actions they suggest to mitigate these

7 Empirical findings concerning emergency operating procedure usage at a Finnish NPP

This chapter presents a study of EOP usage which has been conducted at a Finnish nuclear power plant (NPP). First, the empirical study is presented (Ch. 7.1) and next, some comparisons between the empirically identified habits of action contributing to the system level resilience of NPP operations and the ProCom tool are made (Ch 7.2).

7.1 Study of EOP usage in a Finnish NPP

In order to better understand how emergency operating procedures are used in complex work, a study at a NPP training simulator was conducted. The data was collected at a Finnish plant in 2008 and analysed from different perspectives. The analyses of procedure usage were conducted in 2011-2012 and the results of the study published in detail in 2014 (Savioja et al).

7.1.1 Methods: Data collection and analysis

7.1.1.1 Particularities of the plant

The particular nuclear power plant is of type pressurised water reactor consisting of two separate units. The plant originates from the late 1970's and produces currently close to 500MW electrical power in each unit. A normal control room operating crew consists of three operators: shift supervisor (SS), reactor operator (RO), and a turbine operator (TO). The responsibilities of the operators are divided so that RO takes care of the primary circuit: heat generation and cooling. TO's responsibility is the turbine operation and electricity generation. SS has a leading role in making crucial operative decisions and ensuring the duties of both RO and TO.

All twelve operating crews of the plant participated in the study which means that altogether 44 operators acted as users in the experiment. (In addition to the normal crew the trainees within the crews also took part in the exercises). Thus, the operating experience of the participants varied from 1 to 32 years of experience. There were 18 participants in the experience group 1 - 9 years, 13 participants in the experience group 10 - 19 years, and 13 participants in the experience group over 19 years.

The EOPs of the plant have been designed so that there are two different identification procedures: Incident identification (I0) and Accident identification (A0). Depending on particular automatic signals and alarms, the operators choose either of the above. In choosing which identification procedure to take into use, operators receive support from automation system; when the respective plant protection signal is launched a support display appears on the operating screen which commands to take either I0 or A0 into use. Also, the operators tend to know by heart the criteria (the automatic plant protection signals) which indicate which EOP should be taken into use. Each operator role has a designated flowchart type procedure which has been designed for the specific operator tasks. The identification EOPs of TO and RO prescribe actions related to respective sections of the power plant process. The identification EOP of the SS prescribes actions that further ensure the actions of TO and RO and bring thus redundancy to the activity of the crew.

7.1.1.2 Simulated accident scenario

As a scenario, a design basis accident LOCA (loss of coolant accident) was utilised. The specific LOCA was midsize, which at this plant means that reactor and turbine scrams were automatically released, containment isolation was completed, and safety injection water systems were initiated by the automation system. Also, diesel generators were started up in order to assure energy supply for safety systems. The operators' tasks in this type of scenario consist mainly of double-checking and assuring that all the automatic safety systems are functioning as required and of further identification of the situation e.g. locale of the leak. In this scenario, there was one additional simulated failure in the safety systems: A particular plant protection signal did not function correctly and thus containment isolation was not completed automatically.

7.1.1.3 Data collection

The operating activity of each crew in the simulated accident scenario was observed both online and via recordings. The recordings were in audio, video, and simulator log formats. Each operator carried a head mounted camera which enabled analysis of direction of gaze and communications. In addition, there were overview video cameras and audio recorders registering operating activity. The process events and all operations were recorded in simulator logs.

Later, selected parts of the activity were transcribed into spread sheets in which the courses of action for each crew were depicted on a detailed level. This description included process operations, verbal communications, movements (person's position in the control room), and directions of gaze (when distinguishable in the data) for each crew.

7.1.1.4 Data analysis

As a first step, the simulated accident was carefully analysed from the perspective of critical functions which are endangered in the situation and the required respective operator actions. This analysis produced a functional situation model (FSM) (Savioja et al 2012) of the accident situation. This model depicts the generic critical functions of nuclear power production in the light of this particular emergency situation. The model also has a dimension which depicts the main operations that the operators are supposed to conduct in the situation. In the model the operations are connected to the functions. Thus the model describes *the meaning* of each operation: The model makes explicit both what actions operators take in order to gain control of the process, and, for which operational purpose. The model is the reference for analysing how operating crews in their activity take into account important process information and connect the activity to the overall goals in the situation.

The data analysis was focused by selecting relevant episodes from the scenario. Identification of the relevant episodes commenced with the FSM. First, it was considered which parts of the chosen scenario were relevant from the point of view utilising procedures. Next, the videoed process control activity of one crew was carefully transcribed and analysed by two researchers (see Salo et al. (2009)). The findings concerning the operating practices of the one crew suggested that although the performance was impeccable in the sense that all proceduralized actions were carried out, it seemed that towards the later episodes the crew adopted an operating practice which was not very sensitive to the process information and was tuned more towards the procedure than actually the process. In the detailed analysis all the distinguishable decision making points were marked in a spread sheet containing the transcribed activity. The points were compared with the relevant parts of the FSM and as a result four *episodes* were identified as being crucial for successful accident management in the given scenario: 1) Initial detections and scram 2) Coping with the plant protection failure 3) Detection of new indications concerning the nature of situation 4) Diagnosis and choosing the event based procedure. Later the episodes 1 and 2 were combined and the analysis of the activity of all crews was conducted with 3 episodes.

The following data analyses consisted of two phases. First, the differences in the behaviour of the crews were identified with regard to critical process control tasks found relevant in the episodes, and secondly the differences were classified and graded. This process is reported in detail by Savioja et al. (2014).

7.1.2 Findings of the study

The main finding of the whole study is that within the closely defined envelope of EOP adherence, there still were differences within how the crews acted in the situation. These differences were identified within six different but equally important process control tasks: information usage, situation identification, dealing with automation, decision making, communication, and leadership.

7.1.2.1 Grading of the crew habits in different process control tasks

Within these six tasks, the crew behaviours were classified according to their *interpretative* power. The grading was conducted with three classes: interpretative, confirmative, and reactive (see more, in Norros, 2004). An interpretative habit is such that behaviours can be identified which point in direction of e.g.

- expressing interest on the present situation
- urging to own interpretation of situational demands
- questioning the observed phenomena
- building expectations of future events.

Nearly an opposite type of habit is a reactive habit. Reactive habit is such that it reflects:

- general passivity in actions
- lack of expectations concerning the situation
- no indications of own interpretations

Weakness of reactive habit is that it is able only to react to situations; there are no anticipatory aspects. Therefore the readiness to act, especially in an unanticipated situation can be presumed to be lower.

A confirmative habit is such that neither reactive nor interpretative characteristics can be identified; therefore it constitutes a third type. Confirmative habit can be described as

- taking the situation for granted
- acting in a pre-defined way
- and over emphasizing rules and procedures.

In a confirmative habit repetition is dominant whereas in the interpretative habit also the adaptive potential of a habit is represented. These pre-defined labels of habit types were used as the analysis frame, but the qualitative characteristics of each class in this particular accident scenario were grounded in the observational data.

7.1.2.2 Crews' habits in information usage

The crews' habits in information usage became an item of interest in two analysed episodes: episode 1 concerning initial detections and scram and episode 2 concerning detection of new indications about the nature of the process situation. In both these episodes it was crucial that the crew would take notice of the new information that was available in the control room concerning the process situation and act according the information.

There was variation in crew behaviours ranging from utilising only the alarm information to extensive and diverse process parameter observations and dialogue. Some crews gathered redundant and diverse information before conducting any process interventions. This was considered interpretative because it reflects an objective of validating the initial observations by gathering process information more profoundly. Also, these crews typically gathered different types and abstraction levels of information, e.g. alarms, display support system, process parameter values, trends, automation information, procedures. An important characteristic was also utilisation of both redundant and diverse information sources. Most importantly it was typical to the crews to jointly reflect on the acquired information i.e. construct own understanding of the situation and create assumptions concerning it. This reflection process is a knowledge creation process which increases the opportunity to spot any initially false basic assumptions. It also shows that the crew is tuned towards understanding the dynamic process phenomena as it is taking place. If the crew conducted information gathering this way, the habit of information usage was considered interpretative.

Confirmative habits in information usage were also identifiable in the data. Some crews seemed to conducts double-checking more as rule. While this habit is adequate as such, it does not indicate a deeper strive to understand the situation and being one with the real-time process. The problem with the confirmative habit in this case is that it may lead the operators to only consider such new information that validates the existing one, which is a well-known source of erroneous situation interpretation. The confirmative habit also utilises multiple sources of information but the observations are not as profound as with the interpretative habit because diversity of sources is not utilised to the same extent. Most importantly, there are no signs of joint knowledge creation process in the behaviour of the crew. As is characteristic for a confirmative habit, the rule-based approach generally produces a good result because in most cases the rule applies.

In analysing the data, an indicator pointing to reactive habit was utilising alarm information only. This was considered reactive as it reflects considering singular information reliable enough to base decisions on. The reactive habit in information usage means that information in alarms and the support systems is taken for granted and no additional double checking concerning accuracy of observation is made. Characteristic for the crews acting this way was that observations are paced by alarms and thus the crew is always a little bit lagging behind the process events.

The characteristics of information usage are summarised below (Table 3). The table includes the share of crews portraying each habit in their behaviour.

Table 3. Description of interpretative, confirmative and reactive habits of information usage in a simulated LOCA situation.

	Information usage	**Share of crews**
Interpretative	Variety of sources, redundancy and diversity in information sources, dialogue in interpretation of information	38%
Confirmative	Multiple sources but all the information taken for granted	45%
Reactive	Variation in information sources not sufficient, relying on singular signs	17%

7.1.2.3 Crews' habits in interpreting the process situation

There were differences in crews' habits in situation interpretation based on which the same classification of interpretative, confirmative, and reactive could be made. The analysis was carried out based on verbalisations of the process situation in episode 1 concerning initial detections and scram.

In the analysis of verbalisations and communications there were differences between crews from reading aloud alarm texts to contemplating the general characteristics and also consequences of the situation.

If the operating crews referred to safety functions in their verbalisations, the situation identification was considered to be a threat to mass balance. This indicates a thorough and holistic functional understanding of the process situation and thus the habit was analysed to be interpretative. The same judgement was made if the size and the location of the leakage were somehow referred to. The interpretative habit in identification of the process situation means that the information concerning process, e.g. parameter values are not only considered for face value but also their functional meaning is taken into consideration explicitly. This may be, e.g., the meaning of a certain parameter value from the point of view of the overall safety. This provides resilience in the system because it enables treatment of situations which do not fit the pre-existing typology of possible events.

Confirmative habits were such that situation was identified to be some kind of leakage as this word was utilised in verbalisations but no general linking to the safety functions was made. The confirmative habit in this case reflects a strive to fit the on-going situation to an existing typology of possible events and thus it does not build capability in the system to survive unexpected events. The confirmative habit in process situation identification means that process events are identified, but not really functionally. A habit was seen to be confirmative if there was no evidence of the crew contemplating e.g. size of the leak.

If, in the other end of the spectrum, the verbalisations concerned only alarm information which is directly readable in the alarm system, the interpretation was considered to be that there is a process disturbance, and the habit was considered to be reactive. The reactive habit in process situation interpretation meant that the crew did not explicitly consider the nature nor severity or functional meaning of the incident. The characteristics of situation interpretation are summarised below (Table 4).

Table 4. Description of interpretative, confirmative and reactive habits of interpreting process situation in a simulated LOCA situation

	Interpretation of process situation	**Share of crews**
Interpretative	Interpretation by considering functional meaning of process events	17%
Confirmative	Identify the process events based on an existing typology of possible events e.g. a leak.	50%
Reactive	Identify that something is going on but now strive to understand or label the situation	33%

7.1.2.4 Crews' habits in dealing with automation

The third habit in which differences in crew behaviours were identified was labelled dealing with automation. All the crews observed that the plant protection signals had gone off. But in the analysis of episode 1 concerning initial detection and scram it was thoroughly analysed what was the crews' behaviour during and immediately after realising that plant protection had gone off.

The variation in behaviour of the crews was as follows: Some crews seemed to take a notice of the on-going plant protection chain which in this case was total containment isolation, and immediately check if the automatic sequence was functioning adequately. This behaviour indicates that the crew understands the function of plant protection and takes an active situation specific and questioning stance, in their relation to automatic safety system. This is evidence of agency which indicative of interpretative habit.

On the other end of the spectrum some other crews interpreted the protection signal as a direct signal to perform the scram. By no means is this a wrong interpretation but it reflects total reliance on automation and thus this habit was considered reactive.

In the middle class the containment isolation was mentioned but no active role of ensuring that it was functioning was taken by the crew. This reflects a stance in which the human and the automation remain in their separate pre-defined roles, and neither bothers the other by questioning the adequacy of its functioning. The characteristics of situation interpretation are summarised below (Table 5).

Table 5. Description of interpretative, confirmative and reactive habits of dealing with automation in a simulated LOCA situation.

	Dealing with automation	**Share of crews**
Interpretative	Human assures the automatic functions. Shared responsibility of human and automation.	8%
Confirmative	Automation functioning is observed but not taken action on. Reliance on the pre-defined roles of human and automation	33%
Reactive	Automation information is taken for granted, reflects total reliance on automation	58%

7.1.2.5 Crews' habits in decision making

The next task in which crews' acted differently was the way of deciding to conduct the scram in episode 1. Although all the crews conducted the manual scram to ensure the automatic one, the different crews took different measures of making the decision. The main difference in behaviour of the crews was that some crews took the initiative to conduct the scram by SSs' judgement of the situational demands. The other way to conduct the scram was to do it instructed by procedure. This difference means that some crews conducted the scram prior to taking the abnormal operating procedures into use.

It was analysed that SS's discretion concerning the need to conduct the scram is a sign of interpretative habit. It reflects understanding of the situational needs, anticipation, and prioritizing safety relevant tasks and most importantly, human agency in controlling the automated process.

Conducting the scram after the instruction from the procedure was considered confirmative habit because it reflects a rule-based attitude to decision making: The conductance and e.g. timing of actions is controlled by the procedure.

In the habit of decision making reactive habits were not discovered, but a reactive habit might be hesitance in conducting the scram which would mean that there is insufficient system level control. The characteristics of habits in crew decision making are summarised below (Table 6).

Table 6. Description of interpretative, confirmative and reactive habits of crew decision making in a simulated LOCA situation.

	Decision making	Share of crews
Interpretative	SS makes decision to scram the process. Human as an active, present agent in decision making.	50%
Confirmative	Scram is conducted paced by the procedure. Actions are controlled by the procedure	50%
Reactive	Not identified in the data	-

7.1.2.6 Crews' habits in communication

Communication habits of the crews were observed throughout the scenario but as specific indicator they were utilised in the analysis of episode 2 concerning detection of new indicators of the process situation. This episode had a special demand for communication as it is about transferring to a new procedure. In a heavily proceduralized task, in which almost all the operations are described in the procedure, the selection of correct procedure becomes the critical point in which the common understanding of the crew could be used as a way of testing the decision. But this cannot happen if the crew does not communicate adequately. In analysis of communication it was considered important that the crew would use the resource of collaboration in ensuring that they are moving into a right direction when shifting to use the procedure A0. Therefore, the same information sources were utilised as in analysis of information usage in episode 2 but in the analysis of communication the question was whether the individual crew member who became aware of the particular information made it available to the other crew members also. In other words the question was to what extent the information was communicated within the crew.

The interpretative habit of communicating all the relevant information reflects an objective of creating joint awareness of the situation. Interpretative habit in communication was about depth of issues which were talked about and dialogue. Dialogue in communication enables creation of new knowledge and is thus considered interpretative. Creation of new knowledge is especially important resilience characteristic because resilience assumes that system can survive even totally unprecedented situations about which no previous knowledge exists

The confirmative habit in communication was type of conversation which did not include real dialogue concerning process status or e.g. projected upcoming process behaviour or required crew activity. In the conversation lacking dialogue the contents are merely repetitions of own interpretations and it seems that the objective in the communication is to confirm what is already known. In the analyses of communication also the content was taken into consideration. For example if there was evidence (based on direction of gaze) that the crew was making observations concerning functional status of the process but did not communicate this level information, the communication was seen to be confirmative.

The reactive habit of saying aloud display support system information reflects an objective of mere information transfer. The reactive habit in communication was that the crew made the decision to switch procedures without contemplating together whether the process situation really requires it. They did not discuss together the process parameters. Only the display support system information was made remarks about, thus it seems that the style of communication enforces the interpretation that the process situation is what the procedure has named it. The characteristics of habits in crew decision making are summarised below (Table 7).

Table 7. Description of interpretative, confirmative and reactive habits of communication in a simulated LOCA situation.

	Communication	**Share of crews**
Interpretative	Dialogue concerning process status in the situation. Diverse and redundant information communicated. Reflects creation of joint awareness.	33%
Confirmative	Statements made aloud concerning process parameters. Reflects confirmation of own interpretations.	25%
Reactive	Process state is not explicitly mentioned. Transfer of support system information.	43%

7.1.2.7 Crews' habits in leadership

In the analysis of the behaviour of the crews in episode 3 concerning diagnosis and selecting the event based procedure, there were great differences in the ways the shift supervisors behaved in the situation. The main differences were in coordination and collaboration which in this connection are combined and labelled the function of crew leadership.

The crews’ work in this part of the scenario is quite strictly dictated by the A0 procedure. The procedure is used by following lines and making choices on which arrow to follow after each statement box. Each box contains a statement and if the statement is true the line downward is followed, if it is false the line to the side is followed. This way, at the end point of this procedure one of the A1, I4, or A2 procedure will be taken into use next.

One way of utilising the procedure by the SS was following the lines with finger and pen, saying aloud each statement, and asking some parameter values from the operators, and after arriving to the conclusion (which is the next procedures) seeking confirmation from the operators also for example by saying: “I get A1 , do you agree?”. This habit indicates apt use of procedure as a tool to control own behaviour, dialogue both with self and other crew members to avoid misinterpretations and using the operators’ collaboration as a resource in ensuring the diagnosis. Thus this habit was analysed to be interpretative. The interpretative habit entails that the SS has a leading role in the decision making but that the whole crew is involved. This brings resilience into the system because it is acknowledged that procedure shift is a critical point in the activity and the diagnosis must be assured utilising each crew members’ point of view.

The confirmative habit in this scenario was something which did not have clear inclinations towards either interpretative or reactive. For example confirmative habit was about communicating some of the values to be checked with the crew members but not really dialogically or reflectively i.e. some transparency in diagnosis was made available to the crew. The confirmative habit in leadership meant that in this particular situation the end result was discussed with the whole crew but not the decision making points which lead to it. This is confirmative leadership because the aim is more to acquire confirmation for own diagnosis than to construct the diagnosis together.

Some SSs conducted the whole task silently on their own and only announced the end result: “Take A1 into use”. Typical for this behaviour was also that neither pen nor finger was utilised in reading the procedures and following the lines. It may even be suspected that the SS did not truly follow the whole chain but leant on a previously made diagnosis of the situation which determines the next procedure. Also, this way of using the procedure leaves room for possibility for a mistake as no opportunity for dialogue is created in the habit. Thus this habit was analysed to be reactive due to lack of transparency in decision making. The reactive habit in leadership was such which was not really collaborative work. In some crews the SS only announced the next procedure and did not involve the other operators in the decision making process at all. The characteristics of habits in leadership are summarised below (Table 8).

Table 8. Description of interpretative, confirmative and reactive habits of leadership in a simulated LOCA situation.

	Leadership	**Share of crews**
Interpretative	Active engagement of each operator in all the decision points. Transparency in contemplation enables to spot false conceptions.	42%
Confirmative	The end result of the decision making process is stated and confirmed by all the operators	33%
Reactive	No real collaboration. SS announces the next steps.	25%

7.1.2.8 Summary of the habits of action identified

As was presented above, several differences in the crews' habits of using EOPs were identified in the empirical data. The differences were identified for 6 different process control related tasks: information usage, situation interpretation, dealing with automation, decision making, communication, and leadership.

When the differences in the crews' habits were coded using the classes interpretative, confirmative and reactive, it was obvious that all the crews had created habits of each kind and therefore none of the crews was superior to the others in their habits of action. This may be one reason why in comparing the performance of the operating crews, the differences even out, and it is difficult to find differences. To compare on the level of habits solves research methodical this problem.

7.2 Comparison of the habits of action and the ProCom tool

In this subsection a comparison between the habits of actions and the ProCom tool are made. Clearly, both the habit identification study and the ProCom tool address the same phenomena: indicators of operating crews' ability to utilise procedures in an intelligent way (Dien 1998) in situ. But as they are essentially two different things there are also differences and complementary aspects between the findings and the indicators in the ProCom tool.

The comparison presented here bases on the identified competencies, rating factors, and behavioural markers presented earlier (Table 2) in this report.

7.2.1 General comparison

First of all, the habit identification study did not concern procedure competence as such. It just happens that the habits found to be interpretative and thus add to system level resilience resemble the behavioural markers of appropriate procedure usage developed for the ProCom tool.

The ProCom tool aims to be quite comprehensive as it covers procedure competencies from planning and execution to utilization of backgrounds and finally to adaptability. The empirical study of procedure usage at a Finnish NPP, on the other hand, only identified habit indicators which came up in the study as differences between operating crew behaviours. This means that by no means are the habits of action as comprehensive as the indicators of the ProCom tool.

The ProCom tool bases on both theory and empirical studies concerning operator competences. Therefore, the rating factors and behavioural markers are such that they cover variety of aspects related to operating crews' competences regarding procedure use. The habit identification study only reports those behaviours, within procedure following, in which differences were identified between the operating crews. These differences may be differences in competences although the concept of competence was not utilised in the study. Instead, the differences were labelled as differences in *habits of action* of the crews. It is a conceptual and theoretical question (not suitable for this report) to discuss whether or how much habits of action are related to competences but in general, it can be said that the two concepts are related.

7.2.2 Comparison of the indicators of appropriate procedure use

The ProCom tool is divided into four separate operator competences: procedure planning, procedure execution, utilization of backgrounds and adaptability. In the following each of the competences is discussed separately in comparison to the findings of the habit identification study.

7.2.2.1 Procedure planning

In the ProCom tool procedure planning means choosing the optimal procedure strategy and understanding the purpose of the procedures. On the level of behavioural marker this means discussing the applicability of the procedure strategy and explaining procedure goals and main actions in own words.

In the habit identification study, discussion of the applicability of the procedure strategy was not identified as a specific habit that would differentiate between the operating crews. Neither was discussion about procedure goal and main actions. The only habit relating to these markers somehow, was habit of situation interpretation in episode 1. By analysing the situation profoundly the operating crew may be able to identify the suitability of the procedure strategy. Therefore appropriate situation interpretation is an important pre-condition for the behavioural markers concerning procedure planning in the ProCom tool.

Therefore, it can be concluded that concerning procedure planning the ProCom tool complements the habit identification study because it takes procedure planning as an explicit aspect of procedure competence needed in NPP operations.

7.2.2.2 Procedure execution

In the ProCom tool procedure execution refers to ensuring sufficient progress, applying fold out pages, monitoring critical safety functions (CSFs), and correct transfer to other procedures. On the level of behaviour markers this means discussing major action steps, executing minor steps timely, briefings, looking ahead and preparing, announcing changes in foldout criteria, announcing changes in CSFs timely, and explaining transfer to new procedures in own words.

Some of the issues concerning procedure execution were such that they were also identified as differentiating habits between the crews in the habit identification study. Major action steps e.g. transfer to a new procedure was identified in episode three in the habit of leadership. Also the habit of communication in episode 2 can be interpreted to address the same issues. Some crews were able to use dialogue, in other words discussion, as a resource in making sure that they were following the right path in the procedure.

Application of fold out pages or monitoring of CSFs did not come out as a differentiating factor in the habit identification study. Therefore, it can be concluded that for their part the ProCom tool is complimentary but for the two other characteristics, ensuring sufficient progress and transfer to other procedures the identified habits of action support the notion of the ProCom tool that crews may behave differently in these aspects of EOP usage.

7.2.2.3 Utilization of backgrounds

In the ProCom tool utilization of backgrounds refers to applying additional background information and handling key utility decision points (KUDP). On the level of behaviour markers this means explaining relevant background information, executing steps according to the background information, and using KUDPs when needed.

In the habit identification study, the aforementioned factors were not identified as specific habits that would differentiate between the operating crews.

Therefore, it can be concluded that concerning utilization of backgrounds the ProCom tool complements the findings of the habit identification study.

7.2.2.4 Adaptability

In the ProCom tool adaptability means flexibility of execution of procedures, evaluating procedure effectiveness, anticipating possible outcomes, and independent thinking. On the level of behavioural markers this means performing alternative actions when steps not applicable, discussing procedure strategy in the light of plant safe state, discussing long term consequences, comparing different procedure strategies, and agreeing on final goal and starting action to achieve safe state.

In the habit identification study some of the issues concerning adaptability were identified as different habits among the operating crews. Information usage in episodes 1 & 2 can be interpreted to relate to the same phenomena of evaluating procedure effectiveness. The more thoroughly the operating crews gather information concerning process status, the more thoroughly they also follow whether the procedure strategy is bringing the plant to a safe state. And this information also gives them the possibility to plan alternative actions if needed. Independent thinking is reflected in the habit of decision making in episode 1. It was clear that some crews valued independency of human decision making over conducting only actions prescribed in the EOP. Anticipating possible outcomes is reflected in the habit of dealing with automation in episode 1. It was identified that one crew who was ensuring the plant protection signals was concerned with issues of longer term effect than immediately mentioned in the EOP.

It can be concluded that the findings of the habit identification study support the notion of the ProCom tool that crews may behave differently in these aspects of EOP usage.

7.2.2.5 Indicators from the habit identification study which would complement the ProCom tool

Not all habits of resilient accident management which were identified in the empirical study of EOP usage at the Finnish NPP are included in the current version of the ProCom tool as behavioural markers or factors to be rated.

First of all, information usage and gathering concerning the current process status could be present more strongly in the ProCom tool. This would help evaluating whether the operating crew is actually connecting the procedure to the process situation profoundly. Of the similar type is the habit of situation interpretation. As already mentioned, it is a pre-condition for probably more than one of the procedure competencies identified currently in the ProCom tool and it relates to connecting the procedure and the process situation in an appropriate way.

Secondly dealing with automation is not treated as a procedure competence in the ProCom tool. However, it is an important “competence” and needed in proceduralized situations also. It should be very clear to the operators, what is the division of labour between the automatic plant protection functions and the manual operations. Therefore, this could be a category of rating in the ProCom tool. The issue is especially important as digital automation spreads to the NPPs and thus identifying faults in it may also become under the jurisdiction of the operating crews.

Thirdly the habits of decision making, communication and leadership are something that can be combined under the general terms of coordination and collaboration. These aspects of procedure competencies are not very strongly represented in the current version of the ProCom tool. There are some behavioural markers, such as discussions and announcements, but as it was identified in the study of habits of actions the content of discussion varies a lot between the crews. Some crews only discuss by reading aloud alarm messages whereas some others engage in real dialogue concerning relevant process phenomena and decision making points. Therefore, it is not enough merely to announce things or have a general discussion. Discussion should be such that it brings crews understanding concerning the current process status forward in a way in which new information concerning the situation is created.

7.2.3 Conclusions of the comparison

For the most part, the findings of the habit identification study are compatible with the ProCom tool. They both address similar issues, i.e., what is appropriate behaviour of an operating crew in a situation in which actions are to a far extent dictated by procedures.

The empirical study of EOP usage at the Finnish NPP aimed to identify differences in the operating crews' habits of action which would have an effect on system level resilience of NPP operations. Some differences were truly identified, and their significance is more thoroughly discussed by Savioja et al (2014). The habit identification study is very much embedded in the context of the chosen accident scenario of the particular plant. Therefore it does not provide comprehensive results of all possible habits that increase system level resilience.

The ProCom tool aims to be a comprehensive tool with which operating crews' competency in handling proceduralized situations can be evaluated online. Therefore it contains markers concerning different aspects of procedure usage. However, as the match between the habits identified in the Finnish study and the behavioural markers developed for the tool is not a complete one to one, it is possible that the ProCom tool could be further developed with the findings of the habit identification study.

All in all, it can be concluded that the habits of action identified support the use of the ProCom tool in evaluating the procedure competence of NPP operating crews because similar issues were identified as differences in operating habits as are now mentioned as behavioural markers in the ProCom tool.

As a final remarks concerning the ProCom tool it may be so, that not all behavioural markers can be pre-defined to exist in an evaluation sheet. In conducting the habit identification study of EOP prescribed accident management it became evident that crews sometimes behave very differently and the same behaviours may be indices for different functions. This is possible to identify when taking the whole context of activity into consideration. E.g. reading aloud alarm information may in one situation be appropriate b but in another insufficient. Therefore, it is important to consider the level of *meaning* also in the evaluation of procedure competence. Answering simply yes or no the instances of behavioural markers is not always enough to evaluate procedure competence. It may depend on the meaning and purpose of the behaviour whether it is appropriate or not.

8 Experiences from using eye-tracking for studying procedure use

This chapter summarizes initial experiences from using eye-tracking (ET) methods for studying procedure use. These experiences come from a study on resilient procedure use, where 12 crews of nuclear control room operators each performed four emergency scenarios (Eitrheim et al., 2013; Hildebrandt et al., 2015). The scenarios were about two hours long, and in each scenario there were added complications so that strict procedure following would not lead to success.

8.1 How eye tracking works

An eye tracking device essentially delivers two types of raw data: the gaze position in x-y coordinates, i.e. where the participant is looking, and the pupil diameter, i.e. how far the participant's pupil is dilated. In addition, head-mounted eye trackers like the one used in the study (Figure 10) also record a scene video on which the gaze coordinates can be overlaid, as well as audio.

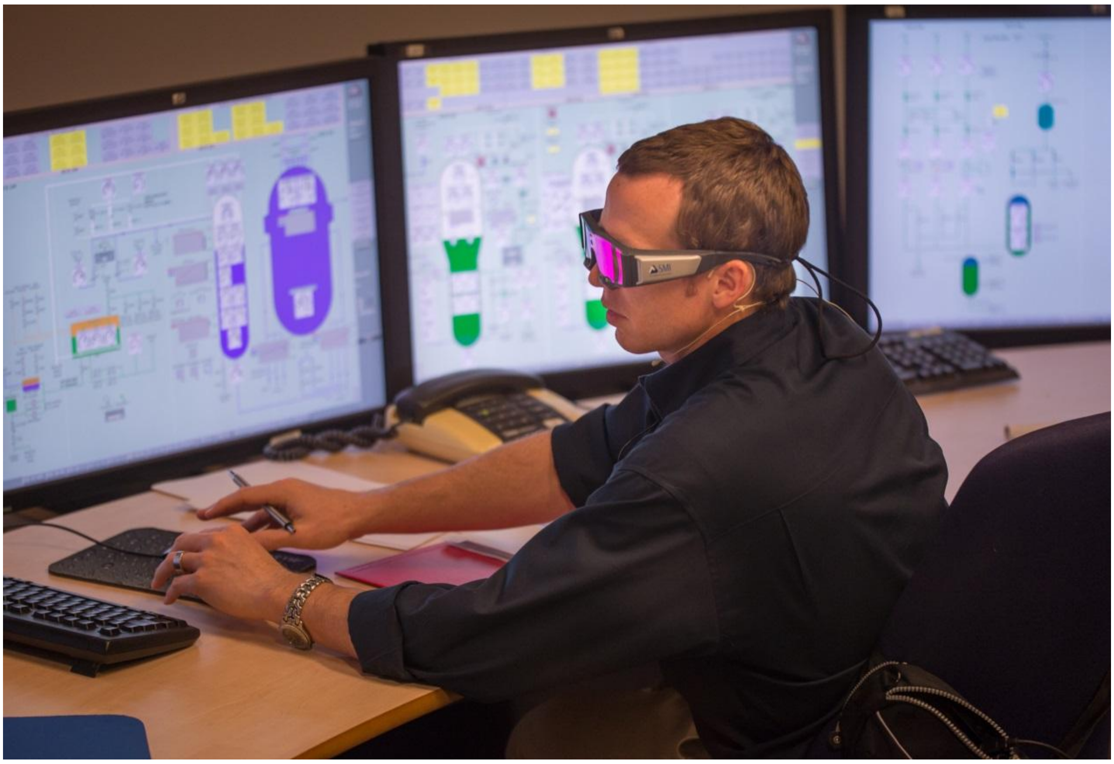

Figure 10. Participant in the resilient procedure use study wearing eye tracking glasses with a built-in scene camera.

This data can be viewed in real-time or as a recorded video stream. For further analysis, the gaze data can be mapped onto a static reference image in order to obtain aggregated gaze data across time. There are several techniques for visualizing such data, for example heat maps (Figure 11, Figure 12).

8.2 Eye tracking applications

Eye tracking is used in a variety of domains, including usability research, marketing research, and human factors research.

In the resilient procedure use study (Hildebrandt et al., 2015) there are three principal goals for using eye tracking. Firstly, we aim to examine the recorded gaze video qualitatively in order to identify scan patterns associated with different types of procedure use. Once these scan patterns have been categorized, a quantitative analysis can be performed whereby the change in scan patterns can be related to experimental factors, individual factors, task-related factors or interface-related factors.

Secondly, gaze data can be mapped onto a static reference image in order to obtain aggregated gaze maps over time, and to enable quantitative gaze data for statistical analysis. A common way to visualize aggregated gaze data is heat maps (Figure 11 and Figure 12). In these heat maps, areas that have been viewed the most are highlighted by colour overlays. This gives the analyst an initial, qualitative impression of which interface elements have been used the most. When mapped onto a procedure page, it can for instance reveal areas that have been used very little, even though they are very important. The mapping does not necessarily have to be onto an exact visual representation of the stimulus. Instead the data can be mapped onto a more abstract structural or functional representation of the interface, e.g. for an emergency procedure, we might use categories such as “step”, “notes and warnings”, “foldout page”, “response-not-obtained column”, “procedure background”, and “process information”. For further quantitative analysis, so-called “Areas of Interest” (AOIs) can be marked out on the reference images. The analysis software can then calculate a range of measures, such as average dwell time, percentage of fixations, and fixations per visit. This data can be used statistically, e.g. to compare identification times for interface elements between two interface designs, or to understand the structure of scan patterns by calculating likelihood transitions (for instance, after an automatic reactor trip, what are the most likely interface elements of information displays the operators will view). Again, all this data can be related to experimental factors, individual factors, task-related factors or interface-related factors.

Figure 11. Gaze data mapped onto a static reference image, and visualized as a heat-map. The heat-map shows the distribution of fixations on the reference image. Red patches signify the areas with the most visual focus. Based on short-sample data, for illustrative purposes only.

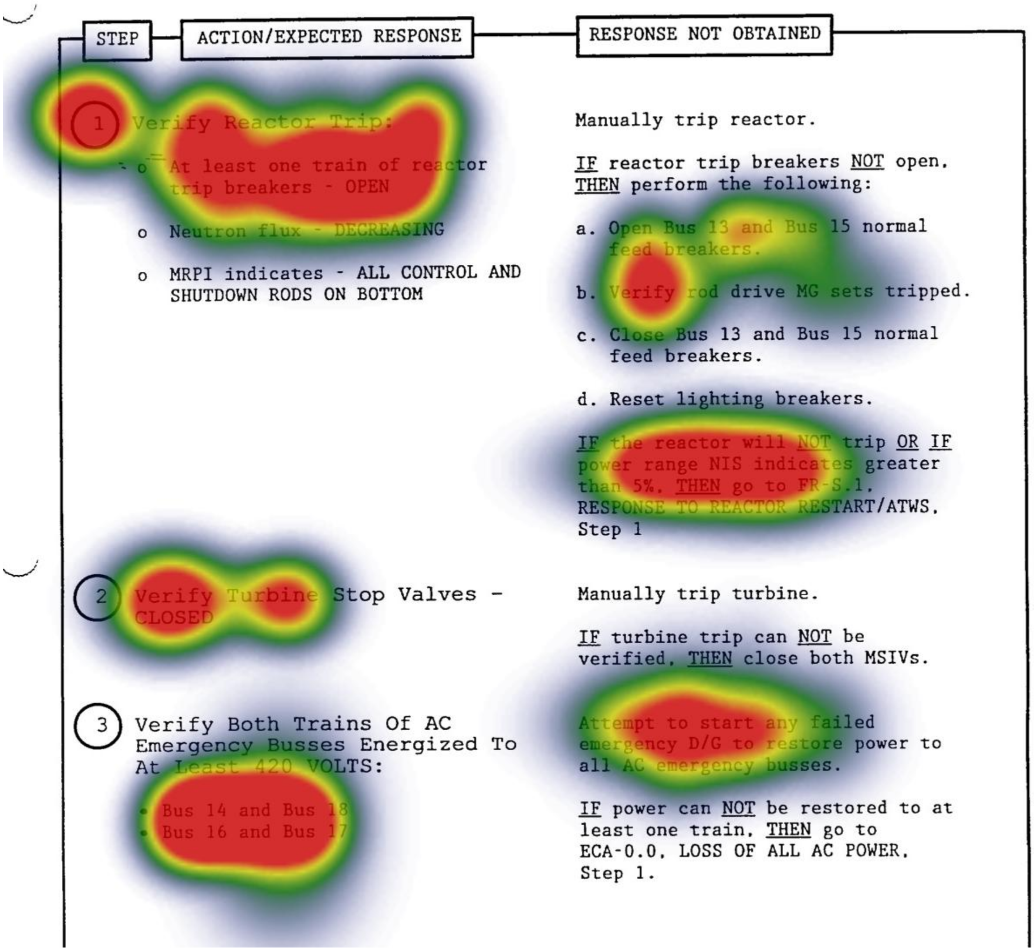

Figure 12. Gaze data mapped onto a static reference image, and visualized as a heat-map. The heat-map shows the distribution of fixations on the reference image. Red patches signify the areas with the most visual focus. Based on short-sample data, for illustrative purposes only.

A third application for eye tracking is to infer cognitive states such as information load. Studies have shown that pupil dilation can be computed into a measure of cognitive workload (Klingner et al., 2011; Kramer, 1991; Beatty, 2008). Measuring cognitive workload via pupil dilation has a number of advantages over observational measures, participant self-report and task measures. These advantages include that measurement is continuous and non-disruptive, and that no subjective ratings are needed. On the other hand, it is necessary to use effective algorithms to eliminate the influence of other factors that affect pupil dilation, such as ambient lighting conditions and fatigue. The combination of various information, such as what task the operator is working on, what information or display element they are using, and the cognitive workload, can provide a comprehensive picture for understanding the factors that cause high cognitive workload, and how to mitigate them effectively.

8.3 Apparatus and method

This section describes how a head-mounted eye tracking system was used in the resilient procedure use study. Note that the use of eye tracking was an add-on to the study, and was not part of the original data collection plan, which relied on observation and analysis of simulator log file protocols. Therefore the use of eye tracking in this study should be considered a pilot application, as a test of the technology and to find an optimized data collection and analysis approach. The study leaders had no prior experience with eye tracking.

The system used in this study was the SensoMotoricInstruments (SMI) Eye Tracking Glasses. It can be seen in Figure 10 and Figure 13. The SMI glasses track the position of both eyes at a frequency of 30hz, and also record pupil dilation. A scene camera is mounted in the nose-piece, and audio is also recorded.

Figure 13. Close-up view of the SMI Eye Tracking Glasses used in this study. The hole in the nose bridge of the glasses houses the scene camera.

Initially recording was performed on a 13-inch laptop PC. The laptop was attached to the back of the participant's chair. Because the glasses were tethered to the laptop, the mobility of the participant was limited. Halfway through the study, we changed to a more mobile solution, where data is recorded on a smartphone device that can be carried in the participant's pocket, or clipped to their belt. This device not only improved the participant's mobility and comfort, but also provided improved battery capacity, allowing up to 4 hours of continuous recording (compared to 2 hours for the laptop solution). The smartphone device also allows live streaming of the gaze video to the observation gallery, though this feature was not used in the current study.

At the start of the study, we were concerned about the comfort of the participants when wearing the device for extended periods. We therefore instructed the participants to take off the glasses whenever they felt discomfort. There are a few factors that can cause discomfort for the participants. These include the weight of the device, its fit on the nose, the pressure it exerts (the device is loosely strapped around the participant's head, because any movement during of the glasses during recording would interfere with the calibration), and a slight warming of the device due to the electronics inside. We expected the participants to take off the device after around 30-45 minutes due to these factors. However the majority of participants wore the device for the whole duration of the run (on average 2 hours). None of the participants reported excessive discomfort.

In this study, we only used the device on participants who do not already wear glasses, i.e. participants who have good uncorrected vision or use contact lenses. About half of the participants wore glasses, i.e. on average we had two participants who could wear the device.

We alternated the use of the device between the participants not wearing glasses, but aimed to get the most recordings from those operators who handle the procedures. There was only on crew where all operators wore glasses, and we did not record eye tracking with this crew. There are a number of other factors that can limit the trackability of participants, such as conditions like astigmatism, or drooping eye-lids. In some cases these caused low tracking accuracy, and consequently we moved the eye tracker to another participant. Recently the manufacturer has launched a set of clip-on lenses for participants using glasses. We have tested this system, and it is promising for increasing the participant pool for future studies. The latest version of the glasses comes with a lighter frame, less constraint of the field of view, and a higher sampling rate of 60 Hz.

An important issue for eye tracking is the ease, speed and accuracy of calibration. The study was run under considerable time pressure, so it was important that eye tracking did not add a lot of extra time and complexity. The first half of the study was therefore run with 1-point calibration. The participant looks at reference point, and the experimenter clicks on the corresponding position as shown by the calibration software on the recording device. This method allowed set-up times of about 3-4 minutes. The accuracy is fairly high for the distance that the calibration was performed on (in our case about 80cm in front of the participant's face). However, for distances further away, e.g. looking at an overview display about 4-6 meters away, we found a sometimes considerable constant error in the vertical axis. Although this could have been compensated for in the analysis and coding phase, this would have been cumbersome and time consuming. We found that 3-point calibration practically eliminated the parallax error and provided high accuracy at all distances. 3-point calibration adds very little time to the calibration process, but results in much improved accuracy. The only complication to consider is that the participant should not move his/her head during the calibration procedure. Once calibration was completed, we found it useful to start the recording and collect some reference data by walking across the room and asking the participant to look at objects at various distances. These recordings allow the analyst to spot any calibration problems when the data is coded, and compensate for them if necessary.

One issue we were concerned about was the issue of calibration drift during extended periods of recording. Drift refers to the gradual change in the calibration point over time that could often be a problem in older system, necessitating a recalibration during recording or during coding. However we did not observe drift issues in our recordings. The only calibration problems were caused by participants accidentally moving the glasses, e.g. when scratching their head.

We have not yet started systematic coding of the data onto reference images. Initial coding test showed that coding takes 3-6 times as long as the original recording. That means a 2-hour experimental run would take about 6-12 hours to code. As this will make complete coding of the material (about 90 hours of raw material) unfeasible, we decided on a different strategy. Instead of coding the whole run, we will code selected episodes, for instance the first 10 minutes following a reactor trip, or periods where the qualitative analysis has shown the crew to struggle with the scenario. To obtain aggregated data about interface usage, we are considering several different approaches, including coding a series of 2-minutes samples selected randomly from each trial.

8.4 Initial insights

Lacking any form of systematic analysis at this stage, all insights we can report at this stage should be treated as highly preliminary and speculative. The observations below were gathered from viewing the raw recordings (Figure 14). For each topic, further analysis will be needed to understand its causes and consequences.

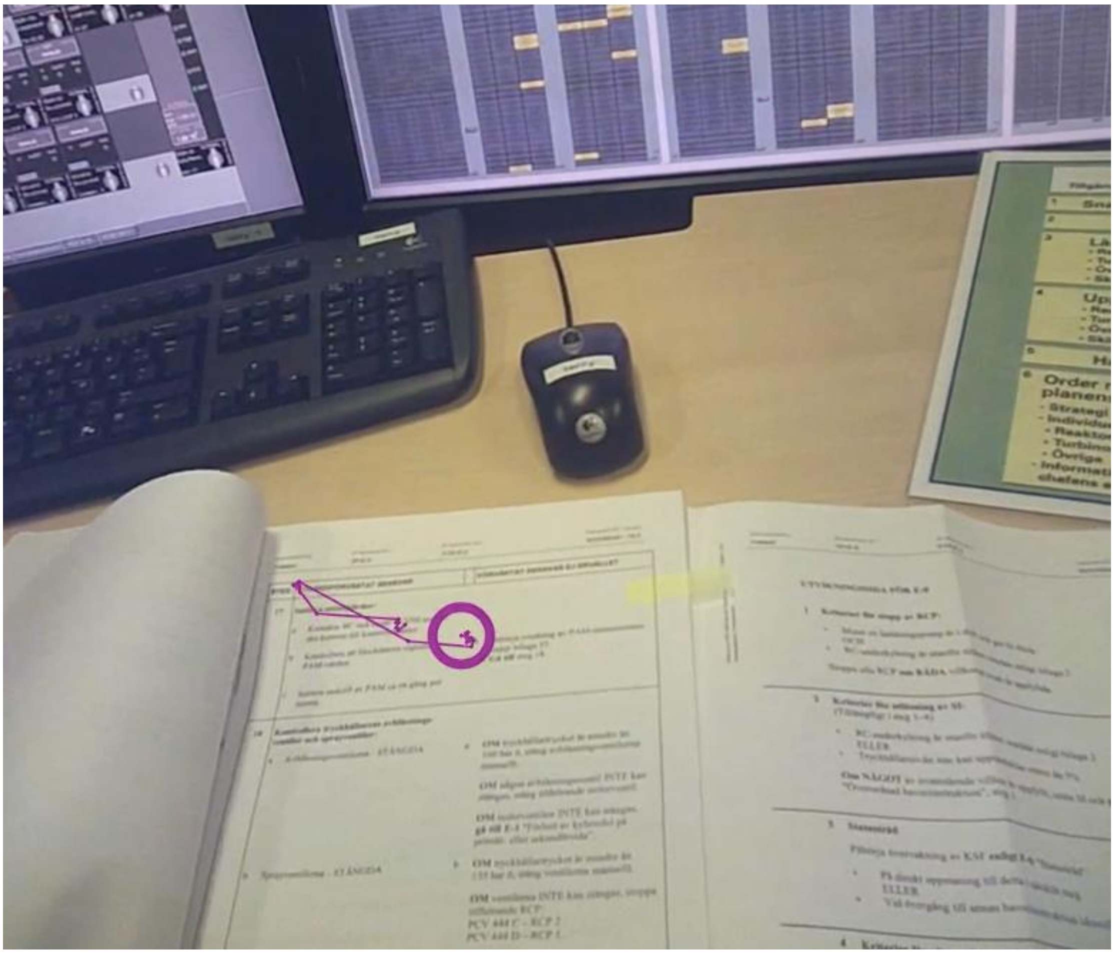

Figure 14. Screenshot of gaze video, showing gaze point with tail (gaze path over the last 2 seconds).

Observed issues included:

(1) *Non-sequential reading of the procedure*. Since procedures are written in a linear format, one could assume that operators also read through the procedures in a linear fashion. However, we can frequently observe back-tracking, jumping ahead, and other non-sequential visual behaviours.

(2) *Re-reading*. We observed many instance when a procedure step was re-read several times, sometimes up to 5 or 6 times. This could indicate problems with understanding the step, or thoroughness in verifying the information by re-reading a step.

(3) *Use of redundant information*. Procedure steps often require reading of indications on the information display. However we frequently observed operators seeking additional information in order to verify the initial reading. This visual behaviour corresponds to good operator work practices of seeking redundant information.

(4) *Use of other information*. We also observed that the operator reading the procedure would occasionally seek other data not related to the procedure step. This suggests that the operator is not just focused on the current procedure step, but may in parallel perform additional analysis or checks, possibly to verify their own assumption about the emergency.

(5) *Searching and navigation*. We observed that operators sometimes used a lot of time searching and navigating the procedures, e.g. scanning indexes, scanning procedure

steps, etc. This may point to problems with the structure of paper-based procedures, which makes it difficult for an operator to get an overview of the content of the procedure, and also can cause difficulties navigating the procedures.

(6) *Use of background documentation*. Occasionally the operators consulted the background documents. This activity often involved scanning and a lot of jumping in the documents. It appeared that the operators were searching for some advice that would help them understand the problems in the scenario, but often this process appeared erratic. It could also point to possible improvements in the format and layout of the procedure documents.

(7) *Usability problems.* We could occasionally observe problems in finding information or indicators. These issues can be expected due to lack of familiarity with the screen-based operator interface. Eye tracking data can help us classify these usability problems and fix them for future studies.

(8) *Overview display*. From the initial observations, it appears that the large-screen overview display is used during the beginning of the scenario, especially after the reactor trip to verify the status of safety equipment, but decreases over time. As the scenario progresses, operator seem to focus more on the operator screens in front of them, only occasionally glancing at the overview display.

(9) *Visual distraction.* We observed frequently that operators would briefly glance at other operators' screens. In particular, most of the times that the supervisors sitting at the back of the room would look up to the overview display, along the way they would have one or two fixations on the displays of the operators in front of them. There are several explanations for this, including visual distraction, or a desire for team situation awareness (checking what the operators are working on).

8.5 Operator feedback

After all scenarios were completed, we showed the participants clips from the recorded eye tracking videos. Feedback from the operators was positive. Some operators remarked that eye tracking could be a useful tool for training, giving trainers a more direct view of what information the operators are and are not considering. This could lead to more targeted interventions and improvements.

Some operators also noted that the eye tracking videos showed how much visual, and by implication cognitive, activity they perform during these emergency scenarios. One operator commented "seeing how much data we look at, it is not surprising we get tired." Another operator commented on the speed of information pick-up. He asked if the video was speeded up, since he couldn't believe that they were scanning the instrumentation at such a high speed. Some operators were interested to view the videos in more detail in order to examine what they themselves could learn from it. In future, it could be useful to replay the videos in a more formal, auto-confrontation format, and to systematically collect the operator's feedback and comments. In the current study, this was not possible due to time constraints.

8.6 Summary and future plans

The use of eye tracking in the current study was exploratory. We were interested to learn about the usability, accuracy and usefulness of current eye tracking equipment. The results so far are encouraging, although more analysis is needed before substantial statements about procedure use can be made. The usability and accuracy of current eye tracking equipment are very good, and the limitations of previous equipment seem largely overcome. Calibration and set-up can be done in less than 5 minutes, and recording times of 2 hours and more can be achieved without excessive operator discomfort. The latest generation of equipment, which came out after the study was completed, promises further improvements. Challenges remain in the area of data coding and analysis. For the moment, coding of gaze data onto reference images remains manual, requiring 3-6 times the amount of time of the original recording. There are indications that vendors are working on more automated system, using image recognition technology. However these products are not publicly available at the moment.

The use of pupil dilation data for measuring cognitive workload is enticing. We have not yet tested the various available filtering algorithms to assess if meaningful data can be extracted from our recordings, where illumination levels varied significantly, causing non-task-related changes on pupil diameter.

The qualitative review of some of the gaze plot recordings already showed some interesting and unexpected insights. These include the frequent observation of non-linear procedure reading, the high frequency of re-reading of procedure step, the glancing of other operators' workstation, and the reduction of the use of the overview display during the course of the scenario. All of these observations should be treated as preliminary, pending more comprehensive analysis of the material.

A more focused analysis of areas-of-interest may reveal interesting results about the most commonly used, as well as least-used, information sources and interface elements. Relating this data to experimental, individual and task-related factors could help generate ideas for better, more adaptive interfaces.

The operator's feedback on the use of eye tracking was positive. We were surprised how fascinated some of the participants were with this technology, and how they developed their own ideas for how to use it. In particular the use of eye tracking for more targeted training is a promising approach. Several participating plants have volunteered to host further tests of this approach, and we plan to use this opportunity for a pilot study on eye tracking for training in 2015. In this study, we will test a number of application scenarios, such as an instructor observing an eye tracking video of a trainee performing a task, and giving feedback on their information use strategies; a trainee observing an eye tracking video of an expert performing the task, and learning from their information gathering patterns; and auto-confrontation of an operator observing eye tracking clips of themselves performing a task.

More futuristic uses of eye tracking will also be considered. For instance, one could imagine integrating an eye tracking device within an operator screen, and feeding the gaze data back into the computer system. This means the process display system would be aware of which information the operator has and hasn't attended to recently. This could lead to adaptive alerting systems that call the operator's attention to those relevant information sources he/she hasn't noticed yet.

9 Summary and conclusions

9.1 Development of the Procedure Competence tool

The purpose of this project was to develop an expert rating tool for measuring operators' procedure competence. Procedure competence was defined as the operators' ability to combine procedure skills, knowledge and attitudes in practice to handle emergency situations in an effective and efficient manner, and according to specified plant standards. Due to the limited scope of the project and its emphasis on developing an applicable, easy-to-use tool, we deliberately excluded basic technical and teamwork skills needed to handle emergency situations. Thus, the tool is aimed at measuring competences in using emergency operating procedures only. For example, the tool focuses on whether the operators discuss the main goals and strategy of a procedure, but does not consider the quality of the communication itself.

Following emergency operating procedures involves multiple competencies such as knowledge and skills necessary to accomplish the pre-planned response, attentional efforts in reading and executing the sequence of actions, monitoring of the procedure effectiveness and decisions on the future path of actions in the current or other procedures. The Procedure Competence tool is organized around four overall competences: procedure planning, procedure execution, utilization of backgrounds, and adaptability. Each competence is divided into more specific rating factors, such as understanding the procedure purpose, ensuring sufficient progress and evaluating the procedure effectiveness. The rating factors are intended as guidance for what concrete skills and knowledge that should be considered when measuring the four procedure competences. These are operationalized through observable, predefined performance indicators labelled behavioural markers. The purpose of the behavioural markers is to reduce observer biases: improve reliability (consistency) across raters and improve validity (accuracy) in measuring procedure competence. The observer rates the performance of the crew for each behavioural marker on a 5-point scale ranging from 1 (very weak) to 5 (very strong).

9.2 Test applications of the Procedure Competence tool

The Procedure Competence tool was tested in HAMMLAB and at Ringhals, KSU. The inter-rater reliability, i.e., the agreement between raters using the tool, could not be fully evaluated as most ratings were performed for different crews in different scenarios. Overall, the test applications showed that the ProCom tool was sensitive to performance variability in procedure execution and adaptability. The procedure planning and utilization of backgrounds had limited applicability in the scenarios observed. The main reason for not evaluating procedure planning was that some of the operators had limited experience with the plant process and procedures in the observed scenarios. Only a few scenarios required utilization of background information. As these experiences have shown, the applicability of the tool and the performance results depend on the content and complexity of the scenario. Thus, we have listed recommendations for what should be considered when designing scenarios for testing emergency procedure competences. In addition, instructors and other users of the tool are encouraged to arrange a pre-scenario briefing to discuss the criteria for scoring the behavioural markers and any scenario-specific considerations to be made.

Both the IFE process experts and the instructors at KSU Ringhals reported that a major challenge with using the Procedure Competence tool was to separate the procedure competences from other performance aspects, such as the technical knowledge needed to understand the plant responses and teamwork aspects affecting the procedure handling such as leadership and communication among the operators. On the other hand, the limited scope of the project allowed us to work thoroughly with the specifics of using emergency operating procedures. The cross-disciplinary group of human factors researchers, nuclear process experts and practitioners from a training simulator ensured a broad and nuanced approach to observing and rating procedure competence. The instructors also expressed that the tool covers aspects of procedure competence that are not currently considered or systematically observed in their simulator training. Thus, the tool and discussions related to its development and application provided new ideas and perspectives for the instructors on what should be evaluated and how this could be achieved. They also saw the tool as a starting point for discussing competence rating between instructors and making their ratings more consistent. Although unfamiliar with using numerical scoring for evaluating crew performance, they agreed that this could be helpful to provide precise feedback and target their future training.

9.3 A comparison of the Procedure Competence tool and empirical findings on the use of emergency operating procedures by VTT

The Procedure Competence tool was also compared to the variations in operating practices observed in an empirical study conducted by VTT. Overall, their findings support the Procedure Competence tool as similar issues were identified in the habit identification study as are included in the behavioural markers in the Procedure Competence tool. For some performance aspects, the Procedure Competence tool complements the habit identification study, for example within procedure planning, use of fold out pages and background information. The habits of actions identified also included aspects that are not considered in the Procedure Competence tool, for example the gathering and use of process status information, how operators deal with automation, makes decisions and communicate with each other. These differences reflect the challenge with defining and measuring procedure competence, and how it relates to other performance aspects.

9.4 The use of expert ratings and eye tracking methods for measuring procedure competence

The Procedure Competence tool is aimed at supporting expert ratings of performance in emergency scenarios. Expert ratings could be supported by and complemented by use of other methods for assessing operator competencies. A common approach is to use operators' self-evaluation through questionnaires and/or interviews after the scenarios. These are subjective measures that can provide additional insights about the operators' understanding and strategic work and help identifying the causes of the behaviours observed in the scenario (diagnosticity). Such techniques were not considered in this project. Instead, we explored the use of an objective measurement technique, eye tracking, for assessing the use of emergency procedures. Recent generations of eye trackers have brought improvements in the ease of use, accuracy, and efficiency of data visualization and analysis.

The more subjective approaches for measuring procedure use such as expert ratings and objective methods like recording eye movements can be seen as complementary. As Figure 15 illustrates, some aspects may be uncovered by use of expert ratings only, whereas other aspects are more accurately studied by use of eye tracking equipment. Together, these techniques could inform each other and provide more nuanced insights about the procedure competences of the crew and future training needs. Expert ratings seem especially suitable for identifying indicators that could serve as predictors of operational outcomes in future situations, such as the understanding of the procedure purpose and main goals and the monitoring of the execution progress towards these goals. The processes for reaching such performance outcomes cannot be captured through use of eye tracking. Eye movement recordings could indicate what information the operators are looking at in the procedure, but not how this information is interpreted.

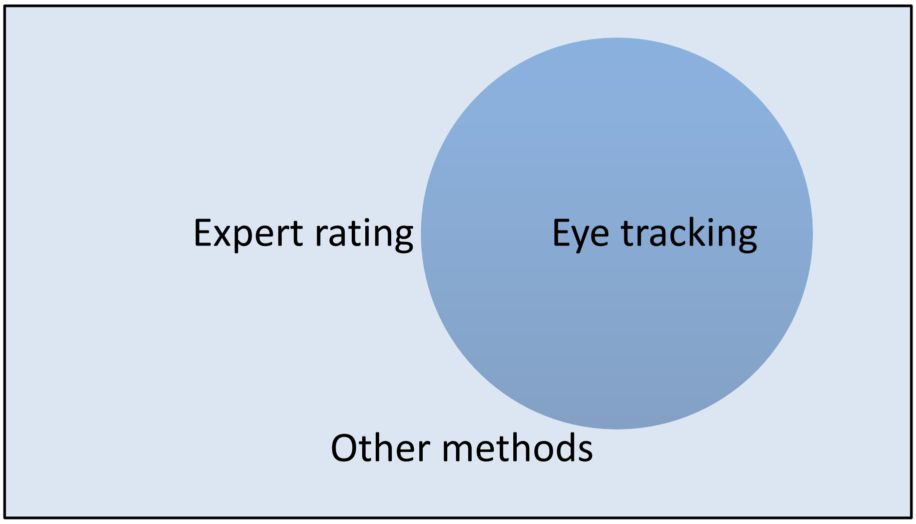

Figure 15 illustrates how procedure competence can be measured by use of expert ratings, eye tracking methods, a combination of these techniques or by use of other methods.

Some devices offer live streaming of gaze video (where the participant is looking) to the observation gallery. This provides a detailed view of what the participant is looking at which could support the expert ratings of procedure competence during the scenario. For example, the eye tracking data may confirm strategic behaviours observed by the expert. At the same time, the expert observations may enrich the interpretability and increase the diagnosticity of the eye movement data, for example explaining that procedure steps are read non-sequentially in order to achieve a flexible execution and efficient progress in the procedure. Eye tracking video material might also be utilized to provide detailed feedback to the operators after a scenario and target future training proposals.

A benefit with using eye tracking techniques is the unique ability to reveal detailed, objective performance data of how information is scanned in the control room. For the use of emergency operating procedures this means for example how the operators scan procedure indexes and steps, instances of re-reading the same pieces of information, non-sequential visual behaviours and the use of additional documents. These data could also be used by the operators themselves, such as replaying videos after scenarios and learning from videos of experts performing the same tasks.

9.5 Future prospects

Based on previous experiences from working with performance assessment tools and studying emergency operation, we developed an expert rating tool for measuring operators’ procedure competence. The experiences so far indicate that this tool provides an extensive approach for observing the use of emergency operating procedures. More systematic comparisons of performance scores (i.e., observers rating the same scenarios and crews) are needed to evaluate the reliability and validity of the tool.

Procedure competence is closely related to other skills and knowledge for handling emergencies, such as basic technical knowledge, the understanding of the current process states (situation awareness) and the communication with other operators. In the future, a similar interdisciplinary approach as established in this project could expand the tool to capture a broader range of competencies, or the tool could be complemented by use of existing measures or methods. Recent eye tracking technologies also seem promising for easily combining expert ratings with detailed performance data on how operators scan information in the control room.

10 Acknowledgements

NKS conveys its gratitude to all organizations and persons who by means of financial support or contributions in kind have made the work presented in this report possible.

11 Disclaimer

12 References

Beatty, J. (2008). Task Evoked Pupillary Responses, Processing Load and Structure of Processing Resources. *Psychological Bulletin*, 91(2), pp. 276-292.

Dien, Y. (1998). Safety and application of procedures, or how do "they" have to use operating procedures in nuclear power plants. *Safety Science*, 29(3), 179-187.

Eitrheim, M. H. R., Hildebrandt, M., Massaiu, S. & Holmgren, L. (2013). Resilient Procedure Use – Effects of an Independent Analyst on Resilient Emergency Operation. HWhP-045. Halden, Norway: OECD Halden Reactor Project.

Hildebrandt, M., Massaiu, S., Holmgren, L., Berntsson, O., McDonald, R. (2015). Initial Results of the Resilient Procedure Use Experiment. HWR-1122 (in progress). Halden, Norway: OECD Halden Reactor Project.

Holmgren, L. & Skjerve, A. B. (2014). Team Self-Assessment Tool (TESA). HWR-1082. Halden, Norway: OECD Halden Reactor Project.

IAEA (2006). Competency Assessments for Nuclear Industry Personnel. Vienna, Austria: Nuclear Power Engineering Section, International Atomic Energy Agency.

Klingner, J. , Tversky, B. & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, 48(3), pp. 323-332.

Kramer, A.F. (1991) Physiological metrics of mental workload: A review of recent progress. In: Damos, D. L. (ed.) *Multiple-task Performance.* London: Taylor & Francis Ltd.

Norros, Leena. 2004. Acting under uncertainty. The core-task analysis in ecological study of work. VTT Publications 546, Espoo: VTT.

Roth, E. M., Woods, D. D. & Pople Jr., H.E. (1992). Cognitive simulation as a tool for cognitive task analysis. *Ergonomics*, 35:1163—1198.

Roth, E.M., Mumaw, R.J., & Lewis, P.M. (1994). An empirical investigation of operator performance in cognitively demanding simulated emergencies (No. NUREG/CR-6208). Washington DC, USA: Nuclear Regulatory Commission.

Salo, Leena; Norros, Leena; Savioja, Paula. 2009. Using Operating Procedures in NPP Process Control. ECCE 2009 - European Conference on Cognitive Ergonomics. Designing beyond the Product - Understanding Activity and User Experience in Ubiquitous Environments. Norros, Leena; Koskinen, Hanna; Salo, Leena; Savioja, Paula (Eds.). VTT Symposium 258. VTT, ss. 258 – 266. http://www.vtt.fi/inf/pdf/symposiums/2009/S258.pdf

Savioja, Paula; Norros, Leena; Salo, L; Aaltonen, Iinka. 2014. Identifying resilience in proceduralised accident management activity of NPP operating crews. *Safety Science*. Elsevier, vol. 68, October, ss. 258 - 274 doi:10.1016/j.ssci.2014.04.008

Savioja, Paula; Norros, Leena; Salo, L.. 2012. Functional situation models in analyses of operating practices in complex work. Proceedings of ECCE 2012. European Association of Cognitive Ergonomics, ss. 58 - 64. 30th European Conference on Cognitive Ergonomics, Edinburgh, 28 - 31 August 2012 doi:10.1145/2448136.2448148

SMI Eye Tracking Glasses, http://eyetracking-glasses.com/

Woods, D. D. (1984). Some results on operator performance in emergency events. In D. Whitfield (Ed.) Ergonomic Problems in Process Operations, Inst. Chem. Eng. Symp. Ser. 90.

Appendix 1

TESA: Emergency Operation Handling and Details[1]

No.	Category	Details
1	Procedure strategies	– Strategies are familiar by the crew – Optimal Procedure strategy is reflected
2	Use/knowledge of procedure backgrounds	– Important Step Basis used and known – Backgrounds used at decision making – Knowledge parts are known and used – Key Decision Points are familiar
3	Plant Responses to events	– Plant event Implication understood – Reasons to plant responses known – Plant Responses anticipated and understood. – High ability interpret indications and to determine effects. – Alertness for divergences
4	Supervise critical plant parameters	– CSF Changes timely identified/announced. – Plant Designs Limits protected – Safety Equipment protected against damage. – Radiation releases reduced as far as possible
5	Effective procedure abilities.	– Procedure entering with main goals is announced and understood by the crew – Procedure Use is thoughtful – Plant monitoring and control is effective – Procedure Following not literal and get stuck – Breefing requested if not Understood – Questioning attitude is kept
6	Resilient procedure use	– Actions and Plant Responses followed up – Preparedness for the unexpected – Alternative Strategies developed if necessary. – Safe handling of the unexpected
7	Redundant/diverse plant indications	– Local Controls used if possible – Diverse Indications used to verify plant indications
8	Actively trying to restore safety functions/limits	– Plant Response at Degraded Conditions anticipated. – Effective Strategies developed – Consequences mitigated with appropriate actions – Plant Conditions reinforced monitoring
9	Evaluates effectiveness of actions	– Plant Response to Actions anticipated and understood – Focus to tasks remained.
10	Mitigates consequences until TSC arrives	– Responsibility taken to bring plant to safe state – Advises are known to be advises. – Orders: do not wait for others to give orders – Full Responsibility taken until TSC is assembled and SAMG is entered.

[1] This is an early version of the categories and details for emergency operation. The categories have been refined and restructured in the final version of TESA documented in HWR-1082.

Appendix 2: ProCom v2.0

Procedure Competence Rating **Crew:** **Date:** **Scenario:** **Observer:**

No	Competence	Rating factor	Behavioural markers	Weight	Score	Weight X Score	Comp grade	Observations	Improvement proposals
1	Procedure planning	Choose optimal procedure strategy	The applicability of the procedure strategy is discussed						
		Understand procedure purpose	Goals and main actions explained in own words when entering a new procedure						
2	Procedure execution	Ensure sufficient progress	Major action steps discussed and given high attention	2					
			Minor steps timely and safely executed	1					
			Briefings performed if steps are unclear	2					
			Look ahead and prepare subsequent procedure steps	2					
		Apply fold out pages	Changes of criteria in fold out pages timely announced	3					
		Monitor Critical Safety Functions (CSF)	Changes of CSF timely alerted and actions started. If red condition: immediately.	3					
			SUM						
3	Utilization of backgrounds	Apply additional background information	Relevant background information explained						
			If background applicable, steps are executed accordingly						
		Handle Key Utility Decision Points (KUDP)	KUDP used for decision making when needed						
4	Adaptability	Flexible execution of procedure	If step(s) not applicable, alternative actions performed	3					
		Evaluate procedure effectiveness	Discuss if current procedure strategy brings plant to safe state	3					
		Anticipate possible outcomes	Discuss and predict long term consequences	2					
		Independent thinking	Compare different procedure strategies and main actions	2					
			Agree on final goal and start actions in current or other appropriate procedure to achieve safe state	2					
			SUM						

How to use the procedure competence rating form

On-line evaluation

The procedure competence rating form covers four main competences: procedure planning, execution, utilization of backgrounds and adaptability. Within each of these four competences, specific rating factors and related behavioural markers are listed.

- For each behavioural marker, please use the "Score" column to evaluate the performance on a 5-point scale as indicated below:

N/A	1	2	3	4	5
N/A if not applicable	very weak	weak	satisfactory	strong	very strong

- Depending on the scenario, some of the behavioural markers may not be applicable, e.g., the scenario doesn't require any transfer between procedures. If a behavioural marker is not applicable to the current scenario, write N/A in the "Score" column and cross out in the column "Weight" (Weight = 0).

Calculate competence grade

- For the competences (1) Procedure planning and (3) Utilization of backgrounds, the Competence grade is the average of the rating factor scores.

- For the competences (2) Procedure execution and (4) Adaptability:

	Weight	Score	Weight X Score	Comp grade
	~~3~~	N/A	-	3.56
	3	4	12	
	2	3	6	
	2	2	4	
	2	5	10	
SUM	9		32	

If behavioural marker N/A, "Weight" = 0

A. Multiply the "Weight" and "Score" for each applicable behavioural marker to calculate a weighted score "Weight X Score" (e.g. 2 X 3 = 6)

B. Summarize the columns "Weight" and "Weight X Score"

C. Divide SUM "Weight X Score" on SUM "Weight" to calculate the Competence grade

Appendix 3

HANDBOOK

FOR MEASURING PROCEDURE COMPETENCE

The purpose of this handbook

The purpose of this handbook is to provide guidance and practical advice on how to measure procedure competence by using the Procedure Competence tool and eye tracking methods. The handbook is primarily aimed at training instructors, experienced shift supervisors and other experts with deep knowledge of the nuclear power plant process and the emergency operating procedures.

What is procedure competence?

Procedure competence is the ability to combine procedure skills, knowledge and attitudes *in practice* to handle emergency situations in an effective and efficient manner, and according to specified plant standards (IAEA, 2006). These competencies may be developed through a combination of education, experience, and training.

How to measure procedure competence?

One way of measuring procedure competence is through observer assessment (expert ratings). The observer may be an instructor, an experienced shift supervisor or other experts with deep knowledge of the plant process and the emergency operating procedures. For the 'non-observable' aspects of procedure competence, the observer need to draw conclusions based on the verbal exchange and reflections in the crew, in addition to the actual activities performed. For example, the ability to choose an optimal procedure strategy might be observed through the actual procedure applied and discussions of its purpose and appropriateness in the crew. These are the *behavioural markers* of the procedure strategy competence of the crew, i.e., the concrete, observable behaviours that will be rated by the observer.

The main motivation for using behavioural markers is to improve the reliability in observing and evaluating procedure competence, i.e., promote stable, consistent and precise assessments across observers. The behavioural markers may also serve as guidance on what concrete competences and practices the operators should hold and maintain through simulator training.

Scope of the Procedure Competence tool

The Procedure Competence (ProCom) tool is aimed at measuring competences in using emergency operating procedures. The ProCom tool does not cover related and possibly overlapping technical and teamwork skills necessary to handle emergency situations, e.g., to build a situation understanding through the use of redundant plant indications, understand the plant implications, communicate and coordinate activities within the team. The scope and limitations of the suggested tool are summarized below:

- The tool is aimed at evaluating the use of emergency operating procedures (EOPs). However, a similar tool could be developed for other types of control room procedures.
- Related operator skills and performance, such as basic technical competence and teamwork are not a part of the tool.
- The tool is developed for pressurized water reactors (PWRs) and needs to be adapted if used in boiling waters reactors (BWRs)
- The current version is based on Westinghouse EOPs for a Swedish NPP. There may be plant specific differences that should be addressed before application in other plants.

- The tool assumes that the operators and the observer are familiar with the EOPs, related background information and the plant process.

- The tool is generic and aimed at covering a variety of events requiring the use of EOPs. The different phases of handling these events including search for the appropriate procedure, use of background information, procedure execution and monitoring of its effectiveness should be reflected in the design of scenarios.

- To keep the tool as simple and practically applicable as possible, the behavioural markers are stated on a crew level. Based on the observations made, proposals for improvements can be provided individually and for the whole team.

The procedure competences to be evaluated

The ProCom tool covers the evaluation of four main competences: procedure planning, procedure execution, utilization of backgrounds, and adaptability. Within each of these four competences, specific rating factors and related behavioural markers are listed. The four competences are shortly described below.

Procedure planning concerns choosing an appropriate procedure, and understanding its purpose and main goals. The nuclear process control planning and overall goals are strongly guided by the operating procedures. Accordingly, the operators should verify whether a procedure provides the most optimal strategy for a given situation. In complex and novel situations, the operators may need to consider multiple procedures to be applied in a given situation, their premises and overall goals, and choose the most optimal strategy provided by one procedure or a set of procedures. To achieve safe and efficient execution of the strategy chosen, the evaluation of procedure planning also considers whether the operators express the goals and actions in their own words before entering a new procedure. This could enhance the operators' understanding of the most important procedure steps to be fulfilled and why.

Procedure execution is the realization of the procedure planning in terms of ensuring performance progress, monitoring of criteria in fold-out pages and monitoring of the critical safety functions. The behavioural markers describe the optimal and expected way of executing a procedure. Thus, insufficient or erroneous actions, such as missing or misinterpreting procedure steps, will be reflected in low rating scores on these behavioural markers. In accordance with the procedure plan, the operators should look ahead and prepare procedure steps whenever suitable, give major steps high attention and execute minor steps timely and safely. When steps are unclear, the operators are expected to perform briefings to clarify the step purpose, applicability and implications.

Each emergency operating procedure has its own *background information* document that includes information about analyses that were realized to develop the strategy of the procedure, information about the physics of the accident the procedure is supposed to deal with, and a detailed explanation of each procedure step. The operators are supposed to be familiar with the background information, explain the procedure steps and execute them according to the background information when relevant. In simulator training, the operators may also make use of the key utility decision points (KUDP) to support their decision making. The KUDPs indicate conditions in which the utility must determine an appropriate course of action.

The fourth area of competence, *adaptability*, concerns evaluation of the procedure effectiveness, predictions of long term outcomes, and flexible execution of procedure steps and strategies in situations where the procedure or parts of it cannot be applied as suggested. If minor deviations are detected, these may be handled without affecting the overall progression, for example by performing alternative actions while maintaining the overall goals of the procedure. If the strategy of the procedure need to be changed or cannot be accomplished, the operators need to define an adjusted strategy that meets the new situational demands. In such cases, the operators are expected to discuss whether the current strategy brings the plant to a safe state, compare different procedure strategies and main actions, and agree on a final goal. Depending on the length of the scenario, the operators may start actions in the current or other appropriate procedures to achieve a safe state.

Scoring of the behavioural markers

The observer is supposed to evaluate the procedure competences through ratings of concrete and specific behavioural markers. Each behavioural marker is rated on a 5-point scale where 1 indicates very weak performance, and 5 indicates very strong performance, see Figure 1 below. Many strengths with no or minor weaknesses that will impact safety of the reactor core and the public are considered to reflect strong or very strong performance. Several weaknesses that are considered relatively insignificant typically reflect satisfactory performance. Many and significant weaknesses are considered to reflect weak or very weak performance.

1	2	3	4	5
Very weak	Weak	Satisfactory	Strong	Very strong

Figure 1 shows the 5-point rating scale for the behavioural markers.

The ratings should reflect the overall impression of the behaviour during a scenario. Depending of the specific event and the actual progress in the scenario, the ratings could be based on only one instance or several instances during the time of observation. Some behavioural markers may not be applicable, e.g., the scenario doesn't require any use of background information. If a behavioural marker is not applicable during the scenario, the observer is asked to indicate N/A in the scoring sheet. N/A means that the behavioural marker could not be evaluated and should be disregarded in the calculation of procedure competence scores.

For the competences on Procedure planning and Utilization of backgrounds, the competence grades are the average of the rating factor scores. For the other two procedure competences, Procedure execution and Adaptability, there are several behavioural markers. As some of the behavioural markers are expected to have stronger plant safety implications, the scoring is weighted to emphasize and reward the most important behaviours. For example, the notification of changes in critical safety functions is emphasized above the quality of executing minor procedure steps. The calculation of weighted scores is explained in Figure 2 below.

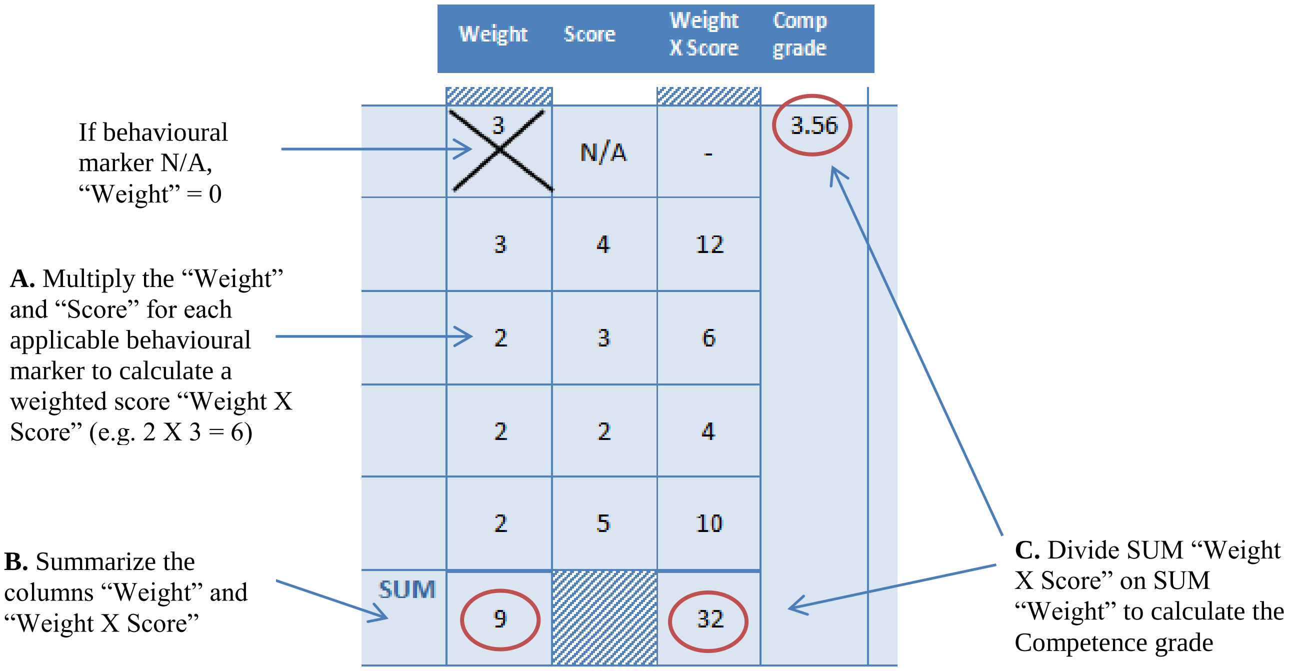

Figure 2 shows an example of how the weighted scores are calculated

Recommendations for the design of scenarios

The applicability of the ProCom tool and the basis for evaluating procedure competence is to a large extent depending on the scenario design, which provides the conditions for observing the crews. We propose the following recommendations when designing the scenarios and planning the evaluation:

- The ProCom tool considers competences for using emergency operating procedures. This can only be observed in accident scenarios with use of EOPs in major parts of the scenario.
- The scenarios should require transfer between procedures. Preferably the operators would need to change procedure at different points in the scenario.
- The scenarios should include situations where the operators would benefit from preparing subsequent procedure steps.
- The scenario design should imply changes of criteria in fold out pages and changes of critical safety functions.
- The scenarios should include challenging situation in which the procedures would not be fully applicable
- The crews should be trained to perform frequent updates or briefings whenever transferring to new procedures, when important information is received or if the situation is unclear.
- Towards the end of the scenario, a discussion might be initiated to clarify what the crew expects regarding the long term consequences and what procedures and actions they suggest to mitigate these

Preparations before evaluating procedure competence

We recommend that the observers using the ProCom tool organize a pre-scenario briefing to analyze the scenario(s) and the specific actions and behaviours required from the operators to handle the events and situations. The purpose of such briefings is to discuss the criteria for evaluating the procedure competences and the scoring of their behavioural markers. What is expected by the crews and what will the observers emphasize in their ratings of the given scenarios?

The operators participating in the procedure competence evaluation should be familiar with the ProCom tool as a guidance on what competences and practices they are expected to show and how these are evaluated.

Application of the ProCom tool and feedback to the operators

During the scenario, the observer is supposed to monitor and rate the behavioural markers as described above. Depending on the scenario design and its actual progress, the observer may indicate preliminary scores and adjust these later in the scenario. A separate column in the ProCom scoring sheet is intended for notes and observations in relation to the various ratings. Based on the ratings made, the observer may suggest concrete proposals for improvements. These could be phrased as concrete training goals for the operators and inform the planning of their future training.

Eye-tracking in procedure use training and evaluation

An eye tracker is a device for measuring eye positions and eye movement. Recent generations of eye trackers have brought improvements in the ease of use, accuracy, and efficiency of data visualization and analysis. For studies of control room work such as the operators' use of procedures, we recommend a mobile, head-mounted eye tracker that allows the participants to move around. The system should also be comfortable to wear and offer a satisfactory battery capacity for simulator training and tests. A limitation for some of the commercial products currently available is that the trackability of participants wearing glasses may be limited.

Some devices also offer live streaming of gaze video (where the participant is looking) to the observation gallery. This provides a detailed view of what the participant is looking at which could support the instructor's evaluation of the procedure execution and other performance aspects, for example the operator's situation awareness.

Eye-tracking equipment enables detailed investigations of how operators read procedures, for example:

- *Searching and navigation*. How operators scan indexes and procedure steps. This could indicate how the operators verify the main goals and important actions of a procedure or point to usability problems with navigating the procedure.
- *Re-reading*. Instances of reading a procedure step again and again. This could indicate problems with understanding the step, or thoroughness in verifying the information.
- *Use of background documentation*. Instances of consulting background documents and how the format and layout of the documents support or hinder the operators.
- *Reading notes, warnings and monitoring of criteria in fold-out pages.* What information that is considered by the operators, the timing and frequency of consulting notes and warnings, and the use of fold-out pages.

- *Non-sequential reading of the procedure*. Instances of back-tracking, jumping ahead and other non-sequential visual behaviours. This could indicate that operators are looking ahead and preparing subsequent procedure steps, executing the procedure in a flexible manner.

The eye tracking videos could serve multiple purposes. As suggested above, instructors can observe eye tracking videos of operators performing a task, provide feedback on their procedure reading patterns and suggest more targeted training to improve the performance. In the education of new operators, trainees can observe eye tracking videos of an expert performing a task and learn from their procedure execution. During simulator training and testing, operators can observe eye tracking clips of themselves to get detailed views of how they're using the procedures in different phases of the scenario.

A video illustrating the use of eye tracking in an accident scenario can be downloaded from this link: https://vimeo.com/98273509

Bibliographic Data Sheet **NKS-332**

Title	Measuring Procedure Competence. Final Report from the NKS-R(14)112/13
Author(s)	Eitrheim, M. H. R.[1], Holmgren, L.[1], Savioja, P.[2], & Hildebrandt, M.[1]
Affiliation(s)	[1]IFE, Norway; [2]VTT, Finland
ISBN	978-87-7893-413-0
Date	March 2015
Project	NKS-R / ProCom
No. of pages	58
No. of tables	8
No. of illustrations	15
No. of references	16
Abstract max. 2000 characters	This report describes the development of an expert rating tool for measuring procedure competence. The Procedure Competence (ProCom) tool is organized around four overall competences: procedure planning, procedure execution, utilization of backgrounds, and adaptability. These are operationalized through behavioural markers, i.e., observable, predefined performance indicators. The ProCom tool was tested in the Halden Man-Machine Laboratory and in a Swedish training simulator. The test applications indicate that the ProCom tool provides an extensive approach for observing the use of emergency operating procedures. The ProCom tool is also supported by findings from a Finnish empirical study of emergency procedure usage. As procedure competence is closely related to other basic technical and teamwork skills, future applications of the ProCom tool should consider these aspects by expanding the tool or using complementary measures. Recent eye tracking technologies also seem promising for easily combining expert ratings with detailed performance data on how operators scan information in the control room.
Key words	Emergency operating procedures, competence measurement, expert rating, eye tracking

Available on request from the NKS Secretariat, P.O.Box 49, DK-4000 Roskilde, Denmark.
Phone (+45) 4677 4041, e-mail nks@nks.org, www.nks.org