



Situation Awareness in Monitoring Nuclear Power Plants – The Process Overview Concept and Measure

**Nathan Lau, Gyrd Skraaning jr, Maren H. R. Eitrheim, Tommy Karlsson,
Christer Nihlwing, & Greg A. Jamieson**

CEL 11-01

Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto
5 King's College Road, Toronto, Ontario, Canada M5S 3G8
URL: www.mie.utoronto.ca/labs/cel/



Directors: Kim J. Vicente, Ph.D., P. Eng.
Greg A. Jamieson, Ph D., P. Eng.

The Cognitive Engineering Laboratory (CEL) at the University of Toronto (U of T) is located in the Department of Mechanical & Industrial Engineering, and is one of three laboratories that comprise the Human Factors Research Group. CEL was founded in 1992 and is primarily concerned with conducting basic and applied research on how to introduce information technology into complex work environments.

Current CEL Research Topics

CEL has been funded by Atomic Energy Control Board of Canada, AECL Research, Alias|Wavefront, Asea Brown Boveri Corporate Research - Heidelberg, Canadian Foundation for Innovation, Defence Research & Development Canada (formerly Defense and Civil Institute for Environmental Medicine), Honeywell Technology Center, IBM, Japan Atomic Energy Research Institute, Microsoft Corporation, Natural Sciences and Engineering Research Council of Canada, Nortel Networks, Nova Chemicals, Westinghouse Science & Technology Center, and Wright-Patterson Air Force Base. CEL also has collaborations and close contacts with the Mitsubishi Heavy Industries and Toshiba Nuclear Energy Laboratory. Recent CEL projects include:

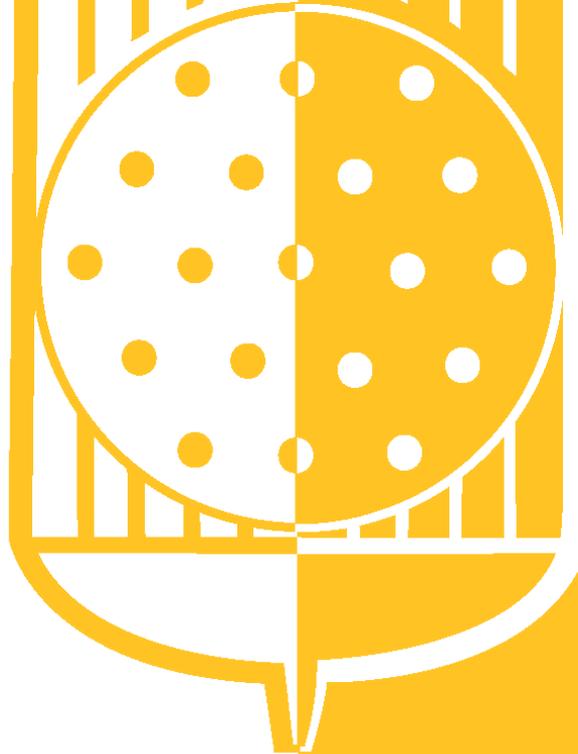
- Developing advanced human-computer interfaces for the petrochemical and nuclear industries to enhance plant safety and productivity.
- Understanding control strategy differences between people of various levels of expertise within the context of process control systems.
- Developing safer and more efficient interfaces for computer-based medical devices.
- Creating novel measures of human performance and adaptation that can be used in experimentation with interactive, real-time, dynamic systems.
- Investigating human-machine system coordination from a dynamical systems perspective.

CEL Technical Reports

For more information about CEL, CEL technical reports, or graduate school at the University of Toronto, please contact Dr. Kim J. Vicente or Dr. Greg A. Jamieson at the address printed on the front of this technical report.

HWR-954

OECD HALDEN REACTOR PROJECT



OECD

Situation Awareness in Monitoring Nuclear Power Plants – The Process Overview Concept and Measure

INSTITUTT FOR ENERGITEKNIKK
OECD HALDEN REACTOR PROJECT
P.O. BOX 173, NO-1751 HALDEN, NORWAY
www.ife.no/hrp

► **Halden Project Use Only** ◀

Situation Awareness in Monitoring Nuclear Power Plants – The Process Overview Concept and Measure

by

**Nathan Lau, Gyrd Skraaning Jr, Maren H. R. Eitrheim, Tommy Karlsson, Christer Nihlwing,
OECD Halden Reactor Project; Greg Jamieson, University of Toronto**

2011-02-11

NOTICE
THIS REPORT IS FOR USE BY
HALDEN PROJECT PARTICIPANTS ONLY

The right to utilise information originating from the research work of the Halden Project is limited to persons and undertakings specifically given the right by one of the Project member organisations in accordance with the Project's rules for "Communication of Results of Scientific Research and Information". The content of this report should thus neither be disclosed to others nor be reproduced, wholly or partially, unless written permission to do so has been obtained from the appropriate Project member organisation.

FOREWORD

The experimental operation of the Halden Boiling Water Reactor and associated research programmes are sponsored through an international agreement by:

- the Institutt for energiteknikk (IFE), Norway,
- the Belgian Nuclear Research Centre SCK•CEN, acting also on behalf of other public or private organisations in Belgium,
- the Risø DTU National Laboratory for Sustainable Energy, Technical University of Denmark,
- the Finnish Ministry of Employment and the Economy (TYÖ),
- the Electricité de France (EDF),
- the Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) mbH, representing a German group of companies working in agreement with the German Federal Ministry of Economics and Technology,
- the Japan Nuclear Energy Safety Organization (JNES),
- the Korean Atomic Energy Research Institute (KAERI), acting also on behalf of other public or private organisations in Korea,
- the Spanish Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), representing a group of national and industry organisations in Spain,
- the Swedish Radiation Safety Authority (SSM), representing public and private nuclear organisations in Sweden,
- the Swiss Federal Nuclear Safety Inspectorate ENSI, representing also the Swiss nuclear utilities (Swissnuclear) and the Paul Scherrer Institute,
- the National Nuclear Laboratory (NNL), representing a group of nuclear licensing and industry organisations in the United Kingdom, and
- the United States Nuclear Regulatory Commission (USNRC),

and as associated parties:

- Japan Atomic Energy Agency (JAEA),
- the Central Research Institute of Electric Power Industry (CRIEPI), representing a group of nuclear research and industry organisations in Japan
- the Mitsubishi Nuclear Fuel Co., Ltd. (MNF)
- the Czech Nuclear Research Institute (NRI),
- the French Institut de Radioprotection et de Sûreté Nucléaire (IRSN),
- the Ulba Metallurgical Plant JSC in Kazakhstan,
- the Hungarian Academy of Sciences, KFKI Atomic Energy Research Institute,
- the JSC "TVEL" and NRC "Kurchatov Institute", Russia,
- All-Russian Research Institute for Nuclear Power Plants Operation (VNIIAES), Russia,
- the Slovakian VUJE - Nuclear Power Plant Research Institute, and
- EU JRC Institute for Transuranium Elements, Karlsruhe,

and associated parties from USA:

- the Westinghouse Electric Power Company, LLC (WEC),
- the Electric Power Research Institute (EPRI),
- the Global Nuclear Fuel (GNF) – Americas, LLC and GE-Hitachi Nuclear Energy, LLC, and
- the US Department of Energy (DOE)

The right to utilise information originating from the research work of the Halden Project is limited to persons and undertakings specifically given this right by one of these Project member organisations.

Recipients are invited to use information contained in this report to the discretion normally applied to research and development programmes. Recipients are urged to contact the Project for further and more recent information on programme items of special interest.



Institutt for energiteknikk
OECD HALDEN REACTOR PROJECT

Title
Situation Awareness in Monitoring Nuclear Power Plants – The Process Overview Concept and Measure

Author:
Nathan Lau, Gyrd Skraaning Jr, Maren H. R. Eitrheim, Tommy Karlsson, Christer Nihlwing, OECD Halden Reactor Project; Greg Jamieson, University of Toronto

Document ID:
HWR-954

Keywords:
Situation Awareness, Process Overview, Measure, Human Factors, Psychometrics

Abstract:
This report presents a domain-specific characterization of situation awareness (SA) for monitoring nuclear power plants (NPPs) – Process Overview. Process Overview is built upon technical and operational properties of NPPs and field research on monitoring activities of control room operators. Process Overview - knowledge acquired through monitoring - is essentially a view of the plant with deviating processes in the foreground and normal processes in the background. The concept captures domain-specific details that are necessary to support design of control room technology and operationalization of SA. The utility of the Process Overview concept for operationalization is exemplified in transforming the SAGAT and SACRI measure into the Process Overview Measure. The Process Overview Measure operationalizes Process Overview as the accurate detection of meaningful changes in relevant process parameters. The Process Overview Measure is evaluated for practicality and sensitivity in two full-scope simulator experiments and for reliability in one controlled experiment. The Process Overview Measure demonstrates sufficient, practicality, sensitivity and reliability as a SA measurement tool in representative nuclear process control environment. This human performance assessment tool is valuable for evaluation and validation of control room technology and operational concepts.

Issue Date:		Name	Signature	Date
2011-02-11	Prepared by:	Nathan Lau	Sign.	2010-11-30
Confidential grade: HRP Only	Reviewed by:	Greg A .Jamieson	Sign.	2010-11-30
	Approved by:	Andreas Bye	Sign.	2011-02-09

MAIL P.O. BOX 173 NO-1751 HALDEN Norway	TELEPHONE +47 69 21 22 00	TELEFAX Administration + 47 69 21 22 01 Nuclear Safety + 47 69 21 22 01 Purchasing Office + 47 69 21 24 40	TELEFAX IND/OC div. + 47 69 21 24 90 VISIT/RID/COSS div. + 47 69 21 24 60 Reactor Plant + 47 69 21 24 70
--	------------------------------	---	---

TABLE OF CONTENTS

1.	INTRODUCTION.....	1
PART I: THE CONCEPT OF PROCESS OVERVIEW		3
2.	PROPERTIES OF NUCLEAR POWER PLANTS	3
3.	MONITORING NUCLEAR POWER PLANTS.....	4
4.	PROCESS OVERVIEW - A SITUATION AWARENESS COMPONENT IN NUCLEAR PROCESS CONTROL	7
4.1	Process Overview	7
4.2	Implications of Process Overview	8
4.2.2	Designing for Monitoring	9
4.3	Summary	10
PART II: THE OPERATIONALIZATION OF PROCESS OVERVIEW.....		11
5.	SAGAT AND SACRI FOR PROCESS CONTROL - A CRITICAL REVIEW	11
5.1	SAGAT	11
5.1.1	Description	11
5.1.2	Critique	12
5.2	SACRI	12
5.2.1	Description	12
5.2.2	Critique	13
5.2.3	Summary	15
6.	THE PROCESS OVERVIEW MEASURE	15
6.1	Methodology	15
6.2	Summary	19
7.	EMPIRICAL INVESTIGATION OF THE PROCESS OVERVIEW MEASURE	20
7.1	Statistical Analysis	20
7.1.1	Evaluation criteria (hypothesis)	21
7.1.2	Parametric and non-parametric statistics (assumptions)	23
7.1.3	ANOVA model selection.....	23
7.1.4	Statistical significance testing	24
7.1.5	Practical significance testing	24
7.1.6	Measures of association - correlation and reliability statistics	25
7.1.7	Confidence intervals.....	26
7.2	Study 1	26
7.2.1	Objective	27
7.2.2	Method	27
7.2.3	Analysis and Results.....	35
7.2.4	Discussion	50
7.2.5	Corollary Issues.....	51
7.3	Study 2	51

7.3.1	Objective	52
7.3.2	Method	52
7.3.3	Analysis and Results	61
7.3.4	Discussion	72
7.3.5	Corollary issues	74
7.4	Study 3	75
7.4.1	Objective	75
7.4.2	Method	75
7.4.3	Data Analysis and Results	81
7.4.4	Discussion	82
PART III: GENERAL DISCUSSION		84
8.	DISCUSSION	84
8.1	Nuclear Process Control	84
8.2	Situation Awareness	85
9.	CONCLUSION	86
10.	REFERENCES.....	87
APPENDIX A. BIAS IN APPLYING INTRA-CLASS CORRELATION TO ASSESS INTER-RATER RELIABILITY OF THE PROCESS OVERVIEW MEASURE		93
APPENDIX B. FREQUENCY DISTRIBUTIONS BETWEEN THE THREE RATERS WITH RESPECT TO THE PARTICIPANT RESPONSES IN STUDY 2.....		94
APPENDIX C KAPPA STATISTICS WITHOUT REFERENCE TO PARTICIPANT RESPONSES.....		97

1. INTRODUCTION

As the operations of nuclear power plants (NPPs) becomes increasing knowledge-based (e.g., see [1]), the notion of Situation Awareness (SA), along with situation assessment, becomes increasingly useful for describing and discussing operator work [2]. With a lexical definition of “knowing what’s going on” [3], SA is not simply a term for a particular psychological function, but a term about “everything” related to acquiring situation knowledge in a complex dynamic environment [2]. Though seemingly reflecting many important facets of work, SA is poorly defined or, at least, suffers from the lack of a unanimous definition or model (see e.g., [4-6] for a discussion). This is not surprising given the ambitious scope of the notion and variety of cognitive work across domains. Theoretical definitions or models of SA (e.g., Endsley’s prominent three levels of SA [3, 7]), as a natural outcome of generalization, often neglect the interaction between domain-specific properties of the work environment and the intricacy of human cognition. In turn, they appear to decouple situation from awareness – depriving the very essence of the notion (see [8-10]).

Domain-specific characterizations of SA could effectively account for the coupling between situation and awareness that would be difficult to capture with abstract models, thereby providing more precise description of cognitive work for a particular domain. Such an approach permits SA to be characterized only by cognitive work pertinent and unique to the domain, although the approach inherently precludes ready generalization to multiple industries. A considerable amount of SA research is situated in specific domains including aviation, infantry operations, command and control, and driving, but the literature contains virtually no discussion on domain-specific approaches to the study or characterization of SA. To the best of the authors’ knowledge, the literature contains only two explicit attempts to formulate domain-specific SA – one in military command and control [11] and another in unmanned vehicles [12]. In summary, the SA notion is being transferred from one domain to another without much domain-specific characterization.

For nuclear process control (or process control in general), SA research is very limited, especially by comparison to other domains such as aviation [10]. Furthermore, the literature contains no domain-specific characterization for process control; whereas, abstract or general SA theories and models often neglect important domain-specific details but introduce cognitive aspects tangential to nuclear process control (see Appendix 14 in [13] for a preliminary discussion). In essence, the current state of SA research does not adequately support practitioners in the application of the SA notion in nuclear process control.

A domain-specific account of SA for nuclear process control could depict the unique characteristics of work activities, challenges and situation knowledge that are meaningful and relevant to NPP operators. Such an account could give practitioners insights to operationalization of SA – what and how SA measurements should be collected in representative environments. In addition to operationalization, a domain-specific characterization supports practitioners in designing control room technology for situations faced by NPP operators.

This report documents a continuing effort [13] to advance both the concept and measurement of SA for nuclear process control. We adopt a domain-specific approach to investigate SA related to monitoring nuclear power plants. The report provides practitioners with insights into the challenges faced by operators to acquire SA and a tool for measuring operator SA related to NPP monitoring. It excludes extensive discussion about research on SA and cognitive psychology in the literature. Interested readers should refer to [13] for a preliminary discussion of these topics¹.

¹ We also expect to publish an extensive discussion on situation awareness and psychology research pertinent to nuclear process control in an upcoming Halden Work Report.

The remainder of this report is divided into three parts. The first part presents the concept of Process Overview – a domain-specific characterization of SA – formulated according to technical and operational properties of NPPs, and field studies of monitoring in large-scale process plants. The implications of Process Overview for designing control room technology and developing measures are discussed. The second part focuses on the Process Overview Measure formulated based on Process Overview and built upon the Situation Awareness Control Room Inventory (SACRI) [14, 15]. The Process Overview measure is evaluated in three empirical studies that speak to sensitivity, reliability, validity and future development. The final part is an overall assessment on the concept and measure of Process Overview, and the domain-specific approach to characterizing SA.

PART I: THE CONCEPT OF PROCESS OVERVIEW

Process Overview is a domain-specific characterization of SA that reflects operator knowledge acquired through monitoring process plants. It builds upon an examination of domain properties, operator monitoring behaviours and their interactions; in turn, it reveals SA characteristics that are relevant and meaningful to nuclear process control.

This part of the report presents the formulation of Process Overview by outlining the key properties of NPPs (Chapter 2) and describing the monitoring behaviours of process operators to cope with such domain properties (Chapter 3). Following the descriptions of NPP properties and monitoring activities, we present Process Overview and discuss its implications for the measurement of SA and design of human-system interfaces (Chapter 4).

2. PROPERTIES OF NUCLEAR POWER PLANTS

The properties of NPPs are the source of many of the work demands on operators, thereby, shaping the monitoring behaviours and, ultimately, defining the situation awareness needed to operate the plants. While most readers are very familiar with the nuclear domain, this section describes the relevant properties of NPPs to illustrate the explicit consideration of the domain properties in interpreting operator monitoring behaviours and subsequently formulating Process Overview.

NPPs are *causal* systems that *continuously* convert fuel into large quantities of electricity, through coordinated plant processes. Thus, central to the design of process plants are the causal relationships upon which the engineered components and processes are based to produce desired effects on the fuel.

The representation of the engineered processes (e.g., condensation, heat transfer) are typically *abstract*, given in terms of scientific principles and equations (e.g., thermodynamics, conservation of mass), or *analogical* to the physical phenomena. In contrast, other domains may permit a relatively concrete representation, such as the spatial-temporal representation in air traffic control.

The *scale* of NPPs is typically very large, containing thousands of inter-connected hardware components arranged to support dozens of engineered processes. At this scale, even highly reliable equipment cannot eliminate occasional component failures. NPPs continue to operate through component failures safely and efficiently because of redundant and safety features. However, failures could still lead to safety and efficiency implications or even disasters.

Automation plays a significant and active role in NPPs. There are many automation components, such as control loops, installed throughout the plants. They act on the system continuously to maintain or optimize steady state operations for intended/expected circumstances. Automation integrates numerical data and issues control signals at a speed and precision beyond the computational capability of human operators. Automation often consists of complex algorithms that work seamlessly during steady-state operations. However, automation may not adequately handle some abnormal operations that are not part of the design considerations. Automation components can fail, sometimes gracefully in circumstances within the design basis, sometimes catastrophically in circumstances beyond the design basis.

The many different equipment and automation components are *tightly coupled* (or inter-connected) in NPPs. Changes or impacts to any component almost always affect other parts of the system. The tight coupling between many hardware and software components results in a high level of *complexity* for process plants. For the purpose of this report, complexity refers to the integral effect of scale, automation and tight coupling.

Given their complexity, NPPs are heavy capital investments and typically have a *lifespan* of several decades to be economically viable. Long time spans, compounded by complexity, impose a limit on designers and engineers identifying and planning for possible events. For example, some NPPs have recently been licensed to increase power production (i.e., uprates) and extend service-life beyond the initial regulatory approval. Such significant changes to the expected plant operations and life-cycles could not have been part of the considerations to engineers at the design stage over three decades ago.

NPPs possess two other notable technical characteristics. First, the process changes and control responses are typically *slow* in comparison to other domains (e.g., aviation) because many of the engineering processes, such as mass transfer energy exchange, are unstable and inefficient in fast conditions. Second, all processes and instruments fluctuate to some degree, sometimes in very complex manner, generating nuisance variations or *noise*.

NPPs operate *continuously*, resulting in two operational characteristics. First, many activities including testing and maintenance must occur while plants are in full operation. For example, operations continue if component failures do not pose safety risks and quality degradation. Second, operational states (e.g., production level as a result of demand) and other activities (e.g., maintenance) can vary on both regular and irregular basis. Because of varying operational states and parallel activities, “normal” or expected values of many instrument readings can vary significantly. For example, values of some process parameters considered normal during full capacity, steady state operations might be considered emergency values during start-up. In essence, operating states are rarely fully pre-defined.

NPPs are *designed as closed systems*² for full capacity and steady state operation, leading to another key operational characteristic. The joint functioning of equipment and automation within the plants governs the safety and productivity of the process. Full capacity and steady state operation are safest and most productive. Disturbance to process plant operations is mostly internally driven, whereas, stress on comparatively open systems (e.g., air traffic control) is mostly externally driven. For example, in process control, productivity and safety of process plants largely depend on the functioning of equipment and automation in attaining and maintaining steady states. In contrast, in air traffic control, productivity and safety depends very much on varying traffic loads and patterns that are dictated (at least initially) by external factors. A crude classification is that steady states can meaningfully describe operations of closed, but not open, systems.

The technical design properties of NPPs give insights to many fundamental characteristics of situations faced by NPP control room operators. In brief, operators must cope with many pieces of tightly coupled equipment and automation. They need to manage closed, causal systems with plant processes that are mainly represented by abstract scientific principles or analogical concepts to the physical phenomena in the processes. The operators must account for the slow, noisy and continuous nature of process dynamics. They must also adapt to operate the plants in events or circumstances that are unexpected by the designers. The next section turns to the literature that describes how operators monitor NPPs to acquire situation awareness that is shaped by these technical and operational properties.

3. MONITORING NUCLEAR POWER PLANTS

The properties of NPP allude to the tremendous challenge of performing situation assessment. Nevertheless, NPP operators establish a collection of activities that enable them to assess the overall operating conditions of the plants.

²NPPs are not completely closed systems. For instance, significant changes in the environment such as temperature of the natural cooling sources for the plant do affect operations. However, environmental shielding is generally incorporated into system design to minimize effect on steady-state operations.

Process operators rely on a diverse set of activities and information sources to acquire knowledge about the plant operating states (e.g., [16-21]). Operators begin their shifts by conducting a shift turnover – a briefing on plant operating conditions between operators coming on duty and those being relieved. The shift turnover communicates the plant status, activities status and any special circumstances. In addition to shift turnovers, operators review logs – chronological records of significant activities (e.g., tests completed, component failures) – to gain knowledge of recent plant status. Operators conduct panel walkthroughs surveying process parameters to gain a “process feeling” as well as to observe specific process parameters and alarm displays. Operators often communicate with field operators and occasionally conduct field tours to collect information about the operating states outside the control room. Operators also participate in maintenance and testing work that requires proactive information gathering and processing. In summary, *active search* for information, as opposed to *passive discovery* of deviations, dominates much of process plant monitoring.

During monitoring, operators rely on *operating contexts* to facilitate sampling and observation of process parameters and alarms. Field studies indicate that process operators constantly gather contextual information to supplement information provided by process and alarm displays. Operators begin their shifts by building the operating context with shift-turnovers and log reviews. The log reviews include not only the shift immediately before but also those in the preceding few days. Occasionally, control room operators, typically with extensive experience as field operators, conduct field tours to acquire further contextual information. The myriad set-up activities before and during a shift confirm the importance of context for monitoring.

The importance of building operating context for monitoring is also evident from the technical and operational properties of process plants. Complexity, lengthy life spans and continuous operations of process plants limit control room interface design from capturing the knowledge of context (c.f., “cognitive underspecification” in [22]). As mentioned, instrument readings considered appropriate for one operating state can be deemed dangerous for another and thus can only be interpreted in context. Furthermore, process plants often operate through component failures. Some of these failures may lead to plant operations that are unanticipated or even unintended in the system design (c.f., normal accidents in [23]). Lengthy lifespan further contends that operational circumstances could gradually depart from the expectation of system designers. For instance, some systems may undergo testing in a manner unaccounted for by designers; hence, control room panels sometimes present abnormal instrument readings. However, these abnormal instrument readings and alarms do not indicate emergency situations but status of testing activities [18]. In other words, operational circumstances in practice are not identical to those postulated during system design, typically to the extent that engineers often cannot operate process plants without substantial operational training like professional operators [24]. For this reason, rarely can interface designers specify and process displays present all necessary information in any given operating circumstances. Consequently, process operators take the initiative to obtain the contextual information that supports the interpretation of process parameters and alarms (and to organize other operational activities such as equipment testing). In summary, building operational contexts is an adaptive behaviour to cope with the limitations of NPP control rooms, some of which cannot be anticipated during the design stage due to the lack of operational knowledge.

During monitoring, operators habitually employ a *top-down approach* to sampling process parameters and automation indicators based on the operating contexts [17, 25]. [26] further argues that self-directed sampling is an inherent part of expertise. The scale of process plants prohibits continuous, comprehensive and reliable sampling of all process parameters [27]. In addition, effective monitoring must account for the effects of automation that are constantly acting on the process. For instance, experienced operators not only inspect process parameters with respect to set points but also check for the possibilities of masking by attending to behaviours of the slave devices governed by automation [28].

Finally, tight coupling between components further necessitates interpreting any process parameter in relation to several others. Nevertheless, process operators can also rely on tight coupling to sample a subset of process parameters that could be sufficiently indicative of process states. In essence, operators are always deciding what process areas should be prioritized for closer observation and how one observation made in relation to the operating contexts could inform subsequent monitoring behaviours.

During monitoring, operators perform substantial *information processing* to account for the behaviours of plant processes, instrumentation and automation in the given operating context. Noise in instrumentation implies that some changes in process parameters are not significant (see e.g., [29, 30]). Furthermore, the dynamics of the process (as depicted by various parameters displayed in the operating panels) complicate the differentiation between nuance fluctuations and true deviations. For instance, many processes respond slowly to control actions and some parameters change in complex manners (e.g., stepwise function) obscuring the true states or behaviours of processes. The unique operating contexts (e.g., maintenance activities) often shape part of the parameter behaviours at a given time. For these reasons, normative values or actions satisfying all circumstances rarely exist, and both normal or abnormal parameter changes could be difficult to determine [20]. The complex process dynamics also implies that precise prediction of individual process parameter values is unrealistic. Monitoring of process parameters is, therefore, not a vigilance task of checking whether variables are within some predefined limits that directly lead to conclusions about plant states (also see [31] for empirical findings of a complex process control task in a microworld).

Process operators adapt to the complex process dynamics during monitoring. Operators often think in “action time” as opposed to “clock time” [32] probably because transitions between two plant states can differ in time consumption and parameter changes from one occurrence to another (see [21]). As a result, operators translate time onto a scale that relates their actions with respect to the process rather than some absolute reference. Alarms in process plants exemplify the challenges in managing process dynamics across varying contexts. Some alarms alert operators to real hazards of the process while others only to nuisance fluctuations. These alarms are conventionally classified as true and false/nuisance alarms, respectively, according to signal detection theory [33]. In some contexts, certain alarms provide information that cannot be classified as either true or false alarms [34, 35]. For example, maintenance or testing activities could lead to alarms that inform operators about the progress of certain activities rather than hazards of certain processes (see [36]). Hence, viewing alarms that are intended to identify “known” hazards requires considerable operator judgment. In summary, monitoring process parameters involves considerations of noise, process dynamics, and operating contexts, all of which involve substantial information processing.

Field and representative simulator studies illustrate that operators engage in many activities to manage the challenges of monitoring process plants. Operators constantly build and update operating contexts to direct their sampling and interpretation of process parameters and alarms; they apply a top-down approach to actively search for relevant information in order to cope with the complexity; and they exhibit considerable information processing and judgment in deciphering meaningful parameter changes or deviations according to characteristics of plant processes. Taken together, the fundamental challenge of monitoring is “not how to pick up subtle abnormal indication against a quiescent background; rather it is how to identify and pursue relevant findings against a cognitively noisy background” [18]. Monitoring thus resembles much more of a problem solving than a vigilance task.

4. PROCESS OVERVIEW - A SITUATION AWARENESS COMPONENT IN NUCLEAR PROCESS CONTROL

The understanding that NPP monitoring involves actively identifying and pursuing relevant findings rather than mechanistically sampling deviations and alarms has important implications in formulating a meaningful and relevant domain-specific SA component. The knowledge of the situation, SA, acquired through monitoring simply cannot be represented by direct (and simplistic) translation of perception given the complex activities involved. In essence, the SA that enables operators to cope with the challenges in managing NPPs is characterized by context building, self-directed sampling and information processing. This section presents Process Overview, the SA component specific to monitoring NPPs, and establishes its utility for informing measurements of SA and design of control rooms.

4.1 Process Overview

Process Overview refers to the situation knowledge acquired through monitoring NPPs. Formulated in considerations of NPP properties and operator monitoring activities, it captures the characteristics of operator knowledge that are often neglected by domain-independent theories or models of SA.

SA in process control should reflect the importance of operational context in monitoring. The operating contexts provide the necessary guidance on process areas to attend to, and process behaviours to expect (also see below). For operators, knowledge or information about the process (e.g., a parameter reading) is almost always anchored to some context or scenario which are often imprecise (e.g., several main systems during start-up) and ill-defined. The broad nature of operating contexts suggests that part of monitoring involves a broad search of plant information to compare actual and expected states. Further, the ill-defined nature of contexts suggests that information search strategies are often guided but not wholly normative. Consequently, situation knowledge from monitoring is an overview of the plant state that is difficult to fully specify.

SA in nuclear process control should reflect a top-down approach to sampling process parameters during monitoring. A top-down approach is necessary for the operator to manage the complexity of process plants with limited cognitive resources. Knowledge about which parameters and automation components require attention appears to be more important than knowledge of absolute parameter values and automation activities. A corollary to the top-down monitoring approach is that process operators hold neither complete nor random knowledge about the plant. They attend to pieces of information deemed relevant in the given operating circumstances. Monitoring therefore seeks out process information that is important and unexpected in the given operating context (also see, [37, 38]). That is, operator SA consists of process information that is sensitive to the operating context (e.g., shut-down) and revealing of abnormal (unexpected) process behaviours.

SA in process control should reflect the information processing and judgment that account for the complex process dynamics during monitoring. The complexity of process dynamics severely limits the accuracy, precision, and hence, the utility of "normative" values and predictions of process parameters. This shapes operator SA in three ways. First, knowledge about which parameter values and automation activities have changed in a significant manner appears more important than how much indicator values and automation activities have changed on an absolute scale. Process dynamics encourage operators to translate parameter changes from a scale of engineering units to magnitudes of operational significance. Second, the time dimension of SA is process-based rather than clock-based as operators respond to the

plant processes of which fixed time standards bear limited intrinsic value³. Third, due to limited precision in modeling process dynamics, anticipation generally occur at a “macro” level (i.e., plant states or general parameter behaviours). Furthermore, operator anticipation is expressed in the form of attention towards critically changing parameters as opposed to mental projection of parameter values (e.g., [21]).

Synthesizing these characteristics, Process Overview - knowledge acquired through monitoring - is essentially a view of the plant with processes in the foreground and background. In the background are processes that are operating under normal conditions. Highlighted in the foreground are processes deviating, or potentially deviating, from ideal conditions. Operators build an operating context to recognize the relevance of different indicators, thereby supporting identification of disturbance areas. Operators functioning under a faulty context would confuse process areas or changes that are normal with those that are abnormal. Process overview reflects the importance of obtaining accurate operating context and strategically distributing attention to compensate for limitations of cognitive resources. It is the product of an array of monitoring activities – building and updating context, observing process parameters, deciphering relevance, and ultimately perceiving the true process deviations.

4.2 Implications of Process Overview

By capturing the unique characteristics of operator knowledge, Process Overview provides specific recommendations in addition to general principles to measuring SA and designing control room technology.

4.2.1 Measuring Situation Awareness

Process Overview informs measurement of operator situation awareness by specifying the characteristics of the knowledge that is *practically* necessary to monitor NPPs given the available resources, thereby prescribing the type of knowledge that is relevant for measurement in representative settings.

Measures of SA acquired through monitoring should therefore capture (at least) three characteristics: (i) the highly contextualized nature of NPP monitoring and control; (ii) top-down approach to sampling plant indicators; and (iii) information processing requirements to perceiving indicator values.

First, Process Overview is context-dependent; therefore, any form of data collection and assessment should be performed as close to the contexts or scenario as possible. In other words, situating in the operational context could improve representativeness of the measurements collected from operators and judged by process experts.

Second, Process Overview is built from a top-down approach to sampling indicators relevant to the context and disturbances. It is necessarily incomplete given the limited cognitive resources. Thus, measures should seek knowledge pertinent to the scenarios rather than random information about the process plant. Further, measures should focus on whether operators have identified relevant process parameters rather than sampled some parameter readings.

Third, Process Overview reflects information processing necessary to account for the process dynamics and operational significance. Precise prediction is not practically necessary to operate process plants and should thus not be emphasized. Measures should aim to include operator judgment on operational

³ We can relate to this experience in our daily activities. For instance, time perception for a conversation is content-based rather than clock-based. Consequently, people can generally recall a portion of conversation more accurately by referring some content rather than time markers. (e.g., What did we talk about after discussing the dinner menu? What did we talk about fifteen minutes ago?)

significance of parameter changes and timing rather than registration of precise changes on some engineering or absolute scale.

4.2.2 Designing for Monitoring

By specifying the characteristics of the knowledge that is *practically* necessary to monitor NPPs, Process Overview indicates that designs of control rooms should facilitate the following operator activities:

- building, updating and accessing contextual information,
- top-down or selective sampling of process indicators, and
- information processing of complex process dynamics.

NPP control room and interface design should facilitate the development and presentation of operational contexts. For instance, process displays can provide features for inputting and presenting relevant maintenance activities that are a part of the operational context and that support interpretation of process parameters. The recognition of contexts in operating process plants is exemplified by increasing development and adoption of contextualized alarm systems (e.g., [36]). In summary, presenting contextual information, which is not supported in many control rooms, could improve support for monitoring NPPs.

NPP control room and interface design should facilitate top-down sampling of process indicators by emphasizing relationships between indicators as opposed to precise readings. Process Overview can capitalize on the use of emergent features, as established in the literature and incorporated in some interface design principles and frameworks (see e.g., [20, 39-41]). As opposed to precise parameter readings (usually displayed in text), emergent features can simplify operator sampling by presenting a set of related parameters together and directing operators to unexpected process behaviours (at least based on the knowledge of interface designers). Control room technology can also facilitate top-down sampling by allocating some displays to present process information that is dynamically selected by operators. Today, operators use trending displays to readily access selected process parameters relevant to the operating circumstances. Trending displays are the common tool that presents information according to the judgment of the process operators. Control room displays can support top-down monitoring further by providing features to present operator-selected information in addition to parameter trends. In brief, for monitoring NPPs, the support for identifying parameters that need close attention is more important than registering precise parameter readings.

NPP control room and interface design should support operators in deciphering truly meaningful process changes rather than detecting changes on an absolute scale. Some technologies aim to reduce uncertainty such as signal validation (e.g., [42]) while others present information relevant to process dynamics such as automation activities (e.g., [43-45]). Process Overview highlights other operational characteristics as well. It indicates that prediction of process parameters is less critical to operators in controlling the plant operation than suggested by other SA models or theories. Though ideal, support for prediction is not a priority. In addition, time-related information such as trends should be presented from an operational (in addition to engineering) perspective. For instance, event tags with operational significance can supplement engineering units (i.e., minutes or seconds) on a time scale.

Some of the design recommendations above aim to leverage operational and contextual knowledge that is simply unavailable during the design stage. These recommendations require flexible design solutions that would introduce uncertainty in verification and validation process. Such flexibility in design solutions and uncertainty in verification and validation may concern regulators in their assessment or approval process. For instance, the process information selected for starting-up NPPs is inevitably different from those for full-power, steady state operations. Some informal and adaptive behaviours of operators are unavoidable to operate NPPs as concluded by the field studies in [18]:

"Finally, we emphasize the contribution of the various informal strategies and competencies that operators have developed to carry out monitoring effectively. Although these strategies are not part of the formal training programs or the official operating procedures, they are extremely important because they facilitate the complex demands of monitoring and compensate for poor interface design decisions. Thus one could effectively argue that the system works well not despite, but *because of*, operators' deviations from formal practices." (p. 52)

Note that the incorporation of some flexible technology in the control room would be dependent on regulators in granting approval to designs from the point of view that operations of NPPs evolve after licensing.

Finally, it is important to note that some of the primary challenges faced by NPP operators are not merely a result of poor human factors engineering but of domain properties, as suggested by [28]:

"For example, if a reactor generates heat and requires substantial cooling systems, rather than simply controlling the temperature, by cooling towers, the plant designer will often employ heat exchangers and transfer the recovered heat to another plant unit, which is driven by heat, e.g., a distillation column. This makes good sense but it introduces diagnostic difficulties... This is a design error by definition since we can be reasonably sure that diagnostic complications are unintended. Moreover, it is a designer error resulting rather directly from the designer's model of the plant, which is limited in a way that an experienced operator's model is not." (p. 268)

The point here is **not** to stifle technical efficiency for "human factors". Process Overview simply reveals that some challenges to NPP monitoring are a consequence of system design decisions that may have had different outcomes if human factors had been included in the decision. It speaks to the well-established but neglected principle that a multi-disciplinary approach to design is necessary even for problems that may intuitively belong to a very specific discipline. Often, designing for multiple stakeholders without substantial compromise is achievable but requires the multiplicity of the design problems to be recognized.

4.3 Summary

Process Overview represents the situation knowledge being developed through monitoring. Built upon field studies and properties of NPPs, it rests on the premise that monitoring is a collection of information gathering activities that enable operators to identify anomalies amidst a set of highly complex processes. Process Overview is the product of monitoring wherein anomalies are depicted in the foreground and normal operations in the background. This domain-specific concept captures the characteristics of SA specific to monitoring process plants and thereby provides unique recommendations for the development control room technology and SA measurements.

PART II: THE OPERATIONALIZATION OF PROCESS OVERVIEW

Part I presented Process Overview – the formulation and description of SA in NPP monitoring. Because it accounts for the characteristics of operator knowledge acquired through monitoring, Process Overview has implications for measurement of SA.

This part of the report turns to measurement of Process Overview to be used in evaluating control room technology and the collection of empirical data on SA in representative nuclear process control environments. While there are often multiple operationalizations of a single construct, this report only presents the Process Overview Measure that is inspired by the Situation Awareness Global Assessment Techniques (SAGAT) [46, 47] and Situation Awareness Control Room Inventory (SACRI) [14, 15]. The discussion begins with a critical review of SAGAT and SACRI (Chapter 5), followed by the description of the Process Overview Measure (Chapter 6), and concludes with four empirical studies evaluating the Process Overview Measure (Chapter 7).

5. SAGAT AND SACRI FOR PROCESS CONTROL - A CRITICAL REVIEW

The Situation Awareness Global Assessment Techniques (SAGAT) is a popular SA measure that inspires the HRP's development of the Situation Awareness Control Room Inventory (SACRI). SACRI, in turn, inspired the development of the Process Overview Measure. Both SAGAT and SACRI contribute to the methodological foundation of the Process Overview Measure. This section reviews the strengths and weaknesses of both SAGAT and SACRI for measuring SA acquired through monitoring. These strengths and weaknesses reflect the congruence to the Process Overview, which is formulated based on field studies on monitoring NPPs.

5.1 SAGAT

5.1.1 Description

SAGAT is a measure built on Endsley's domain-independent three level (i.e., perception, comprehension, and projection) model of SA [3]. SAGAT is separated into the preparation, data collection, and data analysis phases.

During the preparation phase, SAGAT prescribes the Goal Directed Task Analysis (GDTA) to formulate a set of queries on perception, comprehension and projection for administration during an experiment.

During the data collection phase, participants operate a medium to high fidelity simulator [46]. Within each scenario trial, the simulator would freeze at random times and the simulator interface would become inaccessible. Operators would then respond to a random subset of queries formulated during the preparation phase during the simulator freeze. Upon completion of the queries, the trials continue. (Note that random selection of queries and timing of administration is a classical method to avoid cuing effects in experimental psychology⁴.)

During the data analysis phase, the reference answers to the administered query are determined by either indicator readings in the simulator or subject-matter-experts. SAGAT does not prescribe any specific method to calculate the final SA scores but the most common method is percentage correct.

⁴ Cueing effect refers to a shift in participant attention (consciously or unconsciously) due to the appearances or presence of some environmental stimuli. Typically, this effect is unintended in the experiments.

5.1.2 Critique

SAGAT follows the “freeze-and-query” technique and is the most popular SA measure in the literature. Consistent with its theoretical foundation, the SAGAT methodology appeals to intuition in that (declarative) knowledge about the situation supports decision-making and control. Therefore, the knowledge on perception, comprehension, and projection collected during the experiment should be a good predictor of task performance. From this perspective, SAGAT is a useful measurement technique.

SAGAT, though popular, is less readily applicable to measuring SA in nuclear process control than the literature might suggest. [48-50] review various SA measures including SAGAT. SAGAT actually represents a general measurement framework rather than an actual measure or operationalization of SA. Endsley’s model of SA only provides lexical/literary definitions for the three levels of SA without reference to research from psychology or human factors. Thus, the theory on which SAGAT is derived provides limited guidance for specifying the contents and forms of the queries. SAGAT further lacks prescriptions on many methodological details, leaving many decisions in operationalizing SA to the researchers and investigators. For instance, there is limited guidance in formulating good queries and selecting a corresponding response format. For certain knowledge about the situation, guidance might not be necessary; however, for some complex operating situations, the content, phrasing, response formats of the queries could significantly affect validity of the SA scores. While SAGAT might provide a general direction to measure SA, prescriptions on many methodological details are not available to ensure valid, sensitive and reliable measurements (see 7.1.1 Evaluation Criteria for human performance measurements).

On some methodological details, SAGAT prescribes guidance from laboratory research that may not be compatible with the constraints or operator cognitive processes in nuclear process control. For instance, laboratory research in psychology typically employs random selection of queries and timing of administration to minimize cuing effects. However, in representative nuclear process control settings, operator monitoring (as well as many other situation assessment activities) is heavily directed. Thus, responses collected from randomly administered queries are not representative of operator SA.

It is important to have a balanced interpretation on the usefulness of SAGAT. The main contribution of SAGAT is providing a general framework to measure SA. Though developed to operationalize Endsley’s three levels of SA, SAGAT are sometimes modified to measure for specific domain settings or research topics (see e.g., [49, 51]). However, SAGAT inherently omits methodological details that are necessary to operationalize SA effectively for specific industrial settings. Nuclear power plant control is one such setting. In brief, SAGAT is best treated as a measurement framework as opposed a measure of SA.

5.2 SACRI

5.2.1 Description

HRP developed SACRI to evaluate SA of NPP operators. SACRI adapts SAGAT and Signal Detection Theory (SDT) [33]. The theoretical basis of the measure is founded upon Endsley’s three levels of SA [3]. Over the past fifteen years, many HRP simulator studies have employed SACRI or its variants to measure SA. Below is a summary of SACRI separated into preparation, data collection, and data analysis phases.

During the preparation phase, SACRI prescribes the creation of an inventory of parameters that can represent all states of the plant or simulated nuclear process of interest. SACRI structures the queries to elicit operator knowledge about changes in these parameters in the past, present, and future (see Figure 1). SACRI also provides four response sets but does not prescribe any guidance on the selection (Figure

1). Unlike SAGAT, the simulator freezes are determined according to judgment of the experimenters with respect to their research interests as opposed to random selection.

Past, Present and Future SACRI Queries:

- (i) In comparison with the recent past, how has the parameter [code] changed?
- (ii) In comparison with normal status, how would you describe the parameter [code]?
- (iii) In comparison with now, predict how the parameter [code] will develop over the next few minutes.

Response Sets for SACRI Queries:

- (i) Increase/Same
- (ii) Decrease/Same
- (iii) Increase/Same/Decrease
- (iv) Increase in more than one/Increase in one/Same/Decrease in one/Decrease in more than one/Drift in both directions

Figure 1: Query and response formats of SACRI.

During the data collection phase, the participants operate a medium to high fidelity simulator. Within each scenario trial, the simulator freezes according to the pre-determined times to administer the queries from the inventory. Operators would then respond to the queries without any access to the simulator interface. Like SAGAT, the queries are randomly selected and administered to the participants. Upon completion of the queries, operators continue with the scenario trials.

During the data analysis phase, the reference answers for the queries are determined by reviewing simulator logs. The general criteria are that the changes should be (i) observable on displays, (ii) detectable on plots of appropriate scale, (iii) large compared to the baseline established by normal simulator runs, (iv) illustrative of predominant parameter trends, and (v) enclosed in an approximately three-minute interval (for past and future queries). Operator responses to the queries are then classified according to SDT in terms of hit, miss, false alarms and correction rejections. For the final scores, SACRI applies Equation [1] and [2] to calculate sensitivity, A' , and bias, β , respectively.

$$A' = 1 - 0.25 \left(\frac{P_{false_alarms}}{P_{hit}} + \frac{P_{miss}}{P_{correct_rejection}} \right) \quad [1]$$

$$\beta = \frac{(P_{non-deviation} \times P_{false_alarm}) + (P_{deviation} \times P_{hit})}{P_{deviation}} \quad [2]$$

5.2.2 Critique

SACRI reflects some properties of Endsley's domain independent model of SA in the context of monitoring NPPs. It is an attempt to adapt SAGAT by specifying methodological details relevant to nuclear process control.

SACRI prescribes a query structure that is compatible with monitoring NPPs. In particular, the query structure explicitly capitalizes on the fact that operators monitor for changes rather than precise values of parameters [15] (also see e.g., [20]). Further, the query structure does not specify any clock time regarding the past and the future. This is also consistent with Process Overview in that operators tend to think in "action time".

SACRI recommends that query administrations should be timed according to scenario characteristics. Further, [15] concluded from four empirical evaluations of the measure that restrictions based on scenario characteristics may be necessary for the selection of queries within the inventory. Though not explicitly related back to the operator work, the recommendations are consistent with the top-down approach to monitoring as discussed in Process Overview. That is, operators are selective in their search for process information.

Four HRP empirical studies were conducted to develop and evaluate SACRI [14, 15]. The studies concluded that SACRI is sensitive and reliable in representative nuclear process control settings. Several subsequent HRP experiments did employ SACRI and its variants to measure SA. However, there is no other formal evaluation of SACRI beyond the original four empirical, developmental studies. Researchers appear to accept SACRI as a useful SA measure in the nuclear domain.

SACRI does demonstrate five weaknesses, some of which even caution the interpretation of the results from the four empirical, developmental studies. First, SACRI prescribes random sampling of queries from the inventory for administration. Because the researchers found random selection could lead to irrelevant queries, they recommend restricting sampling of process parameters based on scenario characteristics [15]. In effect, developing a general inventory or plant process parameters appears to be an superfluous task because characteristics of the scenario would ultimately dictate the queries to be administered for a particular trial. When restrictions on selection of process parameters for the queries are relaxed, parameters may be irrelevant in the scenarios. Measuring knowledge of such parameters could increase noise and make interpretation of results difficult. In brief, random selection of queries and development of a parameter inventory are ineffective to eliciting knowledge acquired through monitoring NPPs.

Second, the general response format of SACRI appears to be in conflict with SDT. The three Alternative-Forced-Choice (AFC; i.e., decreased, stayed the same, and increased) response format of SACRI cannot be classified into the four categories of hit, miss, false alarms, and correct rejections. [13] examines the conflict between application of SDT and the response format of SACRI in detail. The original SACRI documents [14, 15] did propose 2-AFC response formats (e.g., increased or stayed the same; decreased or stay the same) that are compatible with SDT. However, these response formats appear incomplete to cover the behaviours of any process parameters. This incorrect SDT classification of operator responses could affect the conclusions of the original development studies. For scoring 3-AFC, the SDT literature recommends calculating percentage correct as the final score, thereby omitting the bias score (e.g., [33])⁵.

Third, SACRI prescribes a query structure that includes three categories of time – past, present and future. Eliciting knowledge about parameter changes on a full time spectrum appears appropriate from a theoretical standpoint. However, it neglects the practical challenges of the process control setting. Process Overview suggests that process dynamics often prevent precise prediction of parameters. Further, with complex process faults, prediction is sometimes unrealistic before accurate diagnosis. As such, eliciting prediction of parameter changes may not represent the most pertinent situation knowledge necessary to operate NPPs [21]; thus, the data collected may contain a significant amount of noise (i.e., irrelevant knowledge and guesses).

Fourth, SACRI methodology prescribes that reference answers to queries should be determined by reviewing simulator logs based on a set of general criteria (see above). The potential problem with determining reference answers post trial is the loss of contextual information that is important to decipher signal from noise. As indicated by Process Overview, individual parameter information without context is

⁵ The literature also propose Choice Theory [104, 105] to calculate sensitivity and bias for three or higher alternative-forced-choice tasks. However, due to limited publication in experimental psychology, the authors are unable to readily apply and evaluate Choice Theory for calculating the final scores.

rather meaningless. Therefore, the reference answers may not be representative of the actual situation being experienced in real time. In fact, experimental research staff (including process experts) at HRP expressed answering queries difficult due to the lack of context (in personal communication).

Finally, when determining the reference answers to the queries, SACRI defines the past and future time intervals of the queries to be approximately three minutes before and after the simulator freeze (i.e., the administration of the queries). The time interval specification is inconsistent with Process Overview. As indicated by Process Overview, operators do not think in “clock time” because complex variation of the process prohibits precise prediction; thus, time perception is often subjective and situation dependent, specific to dynamics governed by the scenarios. This three-minute interval definition of past and future also appears inconsistent with the query structure in that the queries administered to the participants do not contain such time intervals. If a three-minute interval is applicable and consistent with nuclear process control, it is unclear why the query omits the time interval (the specifying of which could reduce noise and increase reliability).

SACRI illustrates the potential of adapting SAGAT through applying process control research. However, the SACRI methodology does not exploit the full range of the available research on human operators in process control. Further, some parts of SACRI are incompatible with operator activities as indicated by Process Overview, while others are in conflict with SDT.

5.2.3 Summary

Review of SAGAT and SACRI provides methodological guidance for operationalizing Process Overview. SAGAT is a well-established and intuitive SA measurement framework for collecting operator knowledge of the situation. However, SAGAT does not prescribe sufficient methodological details to truly support valid, sensitive and reliable SA measurements for specific domains. SACRI adapts the SAGAT framework to measure SA in nuclear process control. SACRI specifies many methodological details to improve validity, sensitivity and reliability. While some parts show compatibility with research in process control, SACRI has some critical theoretical and methodological weaknesses. SAGAT and SACRI do offer insights to the operationalization of SA in nuclear process control. The next section turns to the Process Overview Measure, which capitalizes on the research on both SAGAT and SACRI.

6. THE PROCESS OVERVIEW MEASURE

The Process Overview Measure operationalizes Process Overview as the accurate detection of meaningful changes in relevant process parameters. Process parameters are *relevant* when they effectively represent the operating contexts (e.g., shutdown) and reveal potential process anomalies. Parameter changes are *meaningful* when they represent the systematic trends as opposed to (uninformative) fluctuations.

6.1 Methodology

During the preparation phase of full-scope simulator experiment or evaluation session, process experts are instructed to perform three inter-connected tasks – development/review of scenarios, selection of relevant parameters, and identification of simulator-freeze points for query administration. These tasks are described in the following paragraphs.

Process experts are typically responsible for developing test scenarios with characteristics that are useful for the purpose of the study (see [52] for a discussion.) For instance, to evaluate an alarm system for monitoring, the test scenarios must contain process events or faults leading to alarms. The Process

Overview Measure does not prescribe any guidance for developing scenarios because the dominant consideration should be the purpose of the empirical study. However, the Process Overview Measure does rely on process experts to develop representative test cases that are relevant to experimental topics and sufficiently challenging to operators.

Process experts also select process parameters according to the scenario characteristics. Relying on their knowledge in developing the scenarios, process experts should select a set of *relevant* process parameters that represent the operating context and process events (including faults) in the scenario. In other words, the operators/participants successfully completing the scenarios are expected to know the behaviours of these parameters while monitoring the process. The awareness of these parameter behaviours in the recent past is elicited through administrations of queries in the form specified in Figure 2. For empirical studies where process experts do not develop test scenarios but still implement the Process Overview Measure, they must review the scenarios carefully.

Process Overview Query Structure:

Recently, the parameter [code] has:

Process Overview Response Alternatives:

Increased/Stayed the same/Decreased

Figure 2: Query and response format of the Process Overview Measure.

Relevant parameters can typically be classified as (i) context-sensitive or (ii) fault-sensitive according to Process Overview [53]. Context-sensitive parameters reflect the overall plant states based on the operating contexts given to the operators at the beginning of the scenarios. For instance, during start-up at a certain power level, operators often sample a set of key parameters periodically to determine the general progress. The cuing effects for context-sensitive queries should be negligible as these parameters are emphasized during their professional training and work practice. Fault-sensitive parameters reveal the process faults introduced by the scenarios and therefore require close observation. Operators may not sample these fault-dependent parameters during normal operations. Thus, fault-sensitive queries may be subject to cuing effects, prompting consideration of the method by which these queries are introduced (see below). Note that the two classes of parameters could overlap.

Process experts also select timing of simulator freezes in the scenarios to administer queries about the relevant parameters. The timing of these freezes (i.e., administration of the queries) should be determined according to selection of process parameters and scenario characteristics. Some context-sensitive parameters become relevant or irrelevant as the scenarios progress. Context-sensitive queries that are relevant for the entire scenarios may be administered at random times. In contrast, fault-sensitive queries often require strategic timing of freezes because the queries need to coincide with the introduction of the faults without resulting in cuing effects. Two general methods are available to counteract cuing effects of administering fault-sensitive queries. The first method relies on the strategic timing of alarms that occur immediately after the freeze to nullify cuing across participants. The second method relies on administering the queries at the end of the scenarios when the cues from the queries cannot influence operator performance. Either or both method may be employed in any implementation of the Process Overview Measure.

The three process expert tasks - scenario development/review, parameter selection, and freeze timing – are inter-connected and often iterative. For instance, the scenarios may be re-designed to provide effective strategic timing of freezes that can nullify cuing effects. Flexibility across those three tasks

should be leveraged to optimise the quality of the SA measurements with respect to the purpose of the empirical studies.

During data collection (i.e., while operators/participants are running the test scenarios), the simulator should freeze according to the timings specified by the process experts during data preparation phase. During simulator freezes, the participants answer the corresponding set of queries without any access to process displays. From here onwards, the participants' answers are labelled as "responses". At the same time, the process experts supporting the data collection answer the queries with access to all the process displays. From here onwards, the process expert answers are labelled as "reference keys". In addition to collecting the responses and reference keys to the queries, the simulator should log the parameters throughout the scenario for potential verification needs after the experiment.

After the data collection, final scores are the proportion correct (or matches) between the responses and reference keys (collected from the participants and process experts, respectively). These scores may then be analyzed statistically.

Table 1 summarizes the key similarities and differences between the three measures. The Process Overview Measure adapts SACRI and SAGAT in five ways to reflect the characteristics of Process Overview.

Table 1: Similarities and differences between SAGAT, SACRI and the Process Overview Measure.

	SAGAT	SACRI	Process Overview Measure
Construct/ Theory	Model/3 Levels of SA by Endsley [3]	Model/3 Levels of SA by Endsley [3] and some research on NPP monitoring	Process Overview
Elicitation Method	Queries during simulator freezes in each trial	Queries during simulator freezes in each trial	Queries during simulator freezes in each trial
Query Characteristics	Level 1, 2, 3 SA (i.e., perception, comprehension and projection)	Behaviours of process parameters in the past, present and future	Behaviours of process parameters from last meaningful change to the present
Query Development	Goal Directed Task Analysis	System documentation and discussion with process experts to build an inventory for a specific power plant	Scenario analysis by process experts
Query Selection	Random selection based on job classes	Random selection (with constraints) from inventory	Strategic selection according to scenario characteristics
Timing of Query Administration (i.e., freezes)	Random timing	Random timing	Strategic timing based on scenario characteristics
Response Format	No requirement but typically with categorical choices (i.e., multiple choice)	Select one of the four prescribed sets of Alternative-forced-choice (see Figure 1)	3-Alternative-forced-choice (increased, stayed the same, decreased)
Reference key to the queries	Post-trial assessment based on simulator data logs and judgement of subject-matter experts	Post-trial assessment based on simulator data logs	Real time assessment by process experts
Scoring	Percentage correct	Non-parametric formula for calculating Sensitivity and Bias	Percentage correct

First, the Process Overview Measure prescribes that the selection of process parameters for the queries and timing of the query administration are specified according to scenario characteristics because operators do not sample parameters randomly or comprehensively. In contrast, SAGAT prescribes random selection of queries and timing of administration, whereas, SACRI recommend random selection of queries only. Selection of relevant process parameters may introduce cueing effects, biasing performance. The Process Overview Measure recommends the use of scenario events and structures to counteract cueing effects.

Second, the Process Overview Measure prescribes queries about parameter behaviours only pertaining to the recent past. In contrast, SACRI elicit knowledge about parameter behaviours in the past, present, and future. (Note that SAGAT does not prescribe a fixed query structure and is thus not included in this comparison about time orientation of queries.) SACRI queries about past and present parameter behaviours are both consistent with Process Overview. However, SACRI queries about the present can be challenging to interpret. Specifically, operator response may depend on their interpretation of “normal status” with respect to the process faults. If the process faults are factored into the operators’ consideration, the undesirable parameter behaviours are “normal” given the abnormal circumstances. Of course, the parameter behaviours are not “normal” relative to the ideal situations. For example, it is normal that water flow rate into the reactor reduces when a reactor circulation pump fails. Further, it is

also normal that reduced water flow leads to reduced reactor power level. In this example, one may instruct operators to classify the flow rate into the reactor as abnormal because it is a direct result of a pump mal-function. However, classifying reduced reactor power level as normal or abnormal cannot be handled with clear and simple instruction. That is, it would be abnormal if the reactor power level does not reduce; at the same time, reactor power level that does not reduce would be normal relative to ideal operating conditions. In other words, determining the normal status of parameter readings often require qualifying statements about the operating conditions. However, the number of operating conditions and thus qualifying statements are infinite. For this reason, queries about the present status of parameters are impractical for nuclear process control. To avoid semantic and conceptual confusion, the Process Overview Measure excludes queries about the present are omitted. Future/prediction queries are also eliminated, as they are deemed inconsistent with the information needs of process operators. Operators anticipate process parameters that need close observation but rarely predict their specific behaviours, especially in complex operating conditions (also see Chapter 3 and 4.

Third, the Process Overview Measure prescribes that process experts determine the reference answers in real time with observation of participant control actions, support of process displays and full knowledge of the scenarios. In contrast, SAGAT⁶ and SACRI recommend post-hoc assessment of parameter changes. During the scenario trials, contextual information of the scenario, operator control actions, process behaviours and related parameter values are all readily available to the process experts for assessing changes in any process parameters. In contrast, post-hoc assessment of parameter changes based on graphs from simulator logs significantly increases the difficulty of incorporating such operational information. Further, real time assessment generally reduces the labour required to determine reference answers to the queries⁷. For instance, generating trend graphs of the parameters necessary for post-hoc assessment of parameter changes can be effortful.

Fourth, the Process Overview Measure relies on process experts to determine the time intervals representative of the situation when assessing changes in process parameter. The Process Overview Measure assumes sufficient consistency across process experts⁸. SAGAT lacks any detailed prescriptions on query structure, and thus, omits discussion pertaining to time intervals for determining past and future behaviours of process parameters. SACRI pre-defines a time interval of three minutes for determining reference answers to the queries (i.e., assessing parameter changes). As mentioned, pre-defining a time interval for all operating circumstances to assess parameter changes contradicts Process Overview.

Fifth, the Process Overview Measure employs proportion correct as the final performance index of Process Overview. SAGAT also commonly employs, though does not restrict to, proportion correct as the final performance. On the other hand, SACRI employs sensitivity and bias, which appear to be useful indices of performance. However, the existing query structure of SACRI cannot appropriately employ SDT. (For examples of measures that integrate bias scores based on SA queries, see [54, 55].)

6.2 Summary

The Process Overview Measure adapts SAGAT and SACRI to measure SA specifically in nuclear process control. Process Overview, a domain specific characterization of SA, provides conceptual guidance to many methodological adaptations, prescribing methodological details that would improve

⁶ Note that SAGAT does employ process experts to determine reference answers to queries that cannot be obtained from the simulators including Level 2 – Comprehension questions.

⁷ The labour reduction benefit is dependent on the particular setup of experimental staff.

⁸ Study 3, an inter-rater reliability study, tests the merits if this assumption.

and/or ensure validity, sensitivity and reliability of SA measurements. To assess the effectiveness of the method, the remainder of this report turns to empirical evaluation of the Process Overview Measure.

7. EMPIRICAL INVESTIGATION OF THE PROCESS OVERVIEW MEASURE

The Process Overview Measure applies a domain-specific construct – Process Overview - to modify both the SACRI and SAGAT methodologies. Despite leveraging on research specific to process control and “tested” SA measurement methodologies in the literature, the development of Process Overview Measure, thus far, is only an analytical activity that necessitates empirical evaluation.

Furthermore, critiques of SAGAT and SACRI (see above) demonstrate weakness in their empirical assessment for measuring SA in nuclear process control. SAGAT prescribes insufficient methodological details and conflicts with some domain-specific properties. Evaluation of the SAGAT method in representative process control settings is also absent. SACRI, on the other hand, prescribes more methodological details than SAGAT resulting in relatively consistent SA measurements across studies. SACRI was also empirically evaluated in HAMMLAB experiments, although some inherent weaknesses in the SACRI methodology challenge the validity of the empirical evidence. In essence, the Process Overview Measure cannot rely on the theoretical and empirical bases of SAGAT and SACRI to circumvent empirical evaluation.

The Process Overview Measure is evaluated in two full-scope simulator studies and one controlled study. The two full-scope simulator studies (sections 7.2 and 7.3) investigated the sensitivity of the Process Overview Measure to experimental manipulations, relationships with other measures and practical implementation challenges (e.g., real-time scoring). These full-scope simulator studies provided highly representative platforms for evaluating the Process Overview Measure for its intended purposes. However, the studies did not offer sufficient control for detailed investigation (due to the number of research topics and complexity of the experimental tasks involved). The controlled study, on the other hand, afforded a detailed investigation of the measure (7.4).

7.1 Statistical Analysis

The empirical evaluation of the Process Overview Measure in representative process control environment differs from the statistical analysis approach presented in typical psychology/human factors experiments and psychometric studies. Most psychology and human factors experiments compare treatment levels and/or study interactions between treatments without the complexity of full scope simulators; whereas, many psychometric studies examine questionnaires or scales (e.g., personality or intelligence) recruiting a relatively large number of participants ($n > 100$) without involving heavily controlled experimental conditions⁹. Given resource constraints in conducting representative process control studies (e.g., availability of licensed operators), the psychometric properties of the Process Overview Measure are investigated through human factors experiments that recruit few participants in repeated-measured designs for evaluating control room technology/concepts. Therefore, data analysis in this report, especially for the two representative simulator experiments, follows a pragmatic approach, adapting standard statistical procedures to evaluate the Process Overview Measure in nuclear process control. To support readers in interpreting the analysis, results and discussion, this section presents the adopted data analysis approach to evaluating a measure in representative process control settings. It begins with criteria for evaluating a measure, continues with assumptions of ANOVA modelling, and ends with interpretations of specific statistics.

⁹ Some psychometrics studies do involve experimental manipulations or interventions. For instance, an experimental manipulation in an IQ study involved providing additional tutoring time for one group of child participants [106].

7.1.1 Evaluation criteria (hypothesis)

Validity and reliability are the two general assessment criteria for any psychological testing or cognitive performance measurement method. Furthermore, practicality and sensitivity are frequently included in the discussion of establishing validity and reliability. A review of psychometric literature would indicate that a complete assessment of any measure requires extensive empirical research (see e.g., [56-58]). For this reason, the empirical studies in this report are intended to provide an initial evaluation of the Process Overview Measure.

7.1.1.1 Practicality

Practicality is critical for deploying any measure to collect data that are necessary for establishing validity and reliability of an instrument. Specifically, a measure must prescribe a process that can be practically integrated into the preparation, data collection and analysis phases of high-fidelity simulator experiments. Such experiments involve many personnel, some of whom are not experts on human factors (e.g., process experts). The implementation must capitalize on the expertise of team members who are non-human factors professionals. Further, the data collection procedure must not lead to any disapproval from the recruited operators (also see face validity below). Congenial relationships with professional operators are vital to data quality and future experiments. For the Process Overview Measure, practicality is judged subjectively, based on experience of the experimental research team and participants.

7.1.1.2 Sensitivity

Sensitivity refers to the degree to which a psychological testing method can differentiate between experimental manipulations and/or individual participants. In statistical terms, the measurements collected must result in variances or a distribution that reveals differences between independent variables. Therefore, sensitivity is a critical property of a measure for two reasons. First, sensitivity is necessary to reveal the differences between the various control room technologies and/or operational concepts for evaluation. Second, sensitivity is necessary to establish both validity and reliability, which are determined by examining the systematic effects and variance captured in the data. For the Process Overview Measure, the statistically significant effects of experimental manipulations are employed as an indicator of sensitivity. However, significant effects can be a misleading indicator of sensitivity because (i) experimental effects (or their lack of) can be a result of effectiveness of experimental manipulations and/or sensitivity of measure, and (ii) experimental effects can be spurious, especially for large experimental designs. Consequently, experimental effects as an indicator of sensitivity must be interpreted in relation to validity and reliability.

7.1.1.3 Validity

Validity refers to the degree to which a psychological testing method measures what it claims to measure. The discussion on validity of a measure is generally divided as follows:

1. Face validity refers to the degree to which a measure appears to collect data about the construct.
2. Content validity refers to the degree to which a measure covers the entire construct.
3. Concurrent and discriminant validity refers to the degree to which a measure correlates with another measure of the same construct and differentiates from another measure of a different construct.
4. Criterion validity refers to the degree to which a measure predicts another construct or variable.
5. Construct validity refers to the degree to which a measure collects data representative of its own construct (i.e., what it claims to measure). The literature often equates construct validity as

validity (e.g., [56]). In general, all the other forms of validity are established to attain construct validity.

As mentioned, establishing absolute validity of a measure is extremely challenging, generally due to the lack of any true (or validated) reference or criterion. Further, because outcome of naturalistic simulator study is rather unpredictable, clear-cut testing of precise, a priori hypotheses of the experimental manipulations on the measure is likely inappropriate (see [59, 60] for a discussion). For the initial evaluation of the Process Overview Measure, the examination of validity primarily focuses on confirming experimental effects that serve as an indicator of sensitivity. Confirming experimental effects or validating sensitivity of the Process Overview Measure is a priority because evaluation of various control room concepts is most central to HRP simulator experiments.

Specifically, the concurrent, discriminant and criterion validity are examined through comparing significant effects of experimental manipulations and correlations between the Process Overview Measure, measure of situation understanding (i.e., Scenario Understanding, see below) and task performance. Furthermore, qualitative data from the participants are also employed as supporting and refuting evidence of the experimental effects revealed by the Process Overview Measure. This approach seeks both confirming and refuting empirical evidence in the studies to assess the validity of the experimental effects and the Process Overview Measure.

The scope of evidence considered is limited to situation understanding and task performance although the two representative simulator studies included other dependent variables (measures). These other measurements include subjective ratings of workload, trust, and self-assessment. Unlike Scenario Understanding and task performance that are generally interpreted as higher is better, the constructs of workload, trust and self-assessment have neither clearly established relationships with Process Overview nor ideal levels for process operators. For this reason, measurements on constructs other than Scenario Understanding and task performance are omitted in this evaluation of the Process Overview Measure. The measurements of these constructs would involve substantial discussion beyond the scope for an initial evaluation of the Process Overview Measure; however, they could be useful for empirical modelling to study relationships of Process Overview with other human factors constructs in the future.

Face validity and content validity of the Process Overview are investigated based on opinions of the process experts and operator participants. Because implementation of the Process Overview Measure extensively involves process experts, concerns with face validity and content validity are inherently addressed by the methodology.

7.1.1.4 Reliability

Reliability refers to the degree to which a psychological testing method *consistently* measures a phenomenon. That is, the measure must prescribe a method that can collect data with consistently same results and interpretations. The discussion on reliability of a measure is generally divided as follows:

1. Test-retest reliability refers to the measurement consistency of a psychological testing method between different time of administration.
2. Parallel form reliability refers to the measurement consistency of a psychological testing method between two (or more) versions.
3. Inter-rater reliability refers to the measurement consistency of a psychological testing method between multiple raters.
4. Internal-consistency reliability refers to the measurement consistency across items/questions administered in psychological testing method.

As mentioned, the literature tends to discuss psychometric properties from the perspective of developing measurement scales (e.g., personality). Thus, some types of reliability are less applicable than others in the context of evaluating control room technology. Further, some reliability types cannot be practically evaluated in representative experimental settings. Test-retest reliability is omitted in the evaluation of the Process Overview Measure as resources are not available to repeat experiments¹⁰. Parallel form reliability and internal consistency reliability are not applicable to the Process Overview Measure, which administers queries of different parameters (or different questions) according to scenario characteristics. For the Process Overview Measure, the reliance on process experts to determine the reference answers brings inter-rater reliability to the fore. Inter-rater reliability is assessed through a controlled study that provides Cohen's Kappa statistics (see Study 3, section 7.4).

7.1.2 Parametric and non-parametric statistics (assumptions)

There is an on-going dialogue in the scientific community on the choice between parametric and non-parametric statistics (e.g., [61]). Parametric tests are generally more powerful than the non-parametric counterparts. Parametric tests are also available for analyzing complex factorial designs. Furthermore, parametric tests are generally robust to violations of assumptions. On the other hand, non-parametric tests minimize any potential bias that may arise from parametric tests when assumptions are violated.

As there is no conclusion to this debate presently, the data analyses here follow a pragmatic approach. In order to avoid improper use of parametric statistical tests, we test for pre-defined data distribution patterns that are known to introduce bias to the ANOVA results. Thus, parametric tests are applied unless distributions are asymmetrical or skewed in opposite directions across the experimental manipulations. ANOVA is not robust to these particular violations of the normality assumption [62], and whenever such deviations occur, the data will have to be transformed or treated non-parametrically [61].

When manipulated variables are analyzed in a between-subject design (i.e., completely randomized factorial (CRF) block design), statistical assumptions include homogeneity of variance, against which ANOVA is generally robust [63]. However, ANOVA is not robust to the case where the means are correlated with the variance in the experimental effect. When this violation appears, the effect is verified with non-parametric statistics.

When a manipulated variable has more than two levels in a within-subject design (i.e., randomized block factorial (RBF) design), the statistical assumptions include sphericity and compound symmetry [61, 63]. ANOVA is not robust to these two assumptions. The strategy chosen here is to employ the multivariate approach to repeated measures first. If this test uncovers a systematic effect, we know that the sphericity and compound symmetry assumption is not an issue, and that the power of the multivariate approach was sufficient. Further univariate testing is therefore unnecessary. If, however, the multivariate test suggests that an effect occurred by chance, the potentially more powerful Greenhouse-Geisser adjustment will be used (see [52]).

7.1.3 ANOVA model selection

Experiments employing professional operators as participants often involve repeated-measures (within-subject/RBF design) due to the limited availability of these professionals. For this reason, most data sets are collected and analyzed as RBF designs, which are generally perceived to result in more powerful statistical tests than CRF/between-subject designs. However, to be precise, the power of statistical testing in RBF design depends on the correlation between measurements across experimental

¹⁰ Repeating any full-scope simulator experiments purely for the assessment of test-retest reliability of a measure is highly unlikely in any research programme.

conditions [64]. That is, the increased power of RBF design is a consequence of the high tendency of the participants behaving or responding in a manner more similar to themselves than others amongst scenario trails. [65] reveals that the advantage of RBF over CRF designs occurs when the measurement correlations (or intra-class correlation within subjects), ρ , is above .25. When the intra-class correlations within subjects are lower, data analysis employing CRF is in fact more powerful than RBF designs (because of higher degrees of freedom for the error terms in CRF). However, the designs for analysis are different from the designs for data collection. [61, 62] further discuss the advantages and disadvantages of adopting different experimental designs between analysis and data collection. The data analysis in this report adopts a conservative but pragmatic approach: the dataset is analyzed as CRF (rather than RBF) when the intra-class correlation is below 0.1, demonstrating negligible association within-subjects across experimental conditions (see section 7.1.6 for a short discussion on correlation).

7.1.4 Statistical significance testing

The American Psychological Association established the Task Force on Statistical Inference (TFSI) to clarify controversial methodological issues and provide new guidelines for the application of statistics in psychology [66]. According to TFSI, it is always better to report actual p-values than to make dichotomous accept-reject decisions with a predefined significance level. In this report, all results included the p-values, which directly express the likelihood that statistical results have occurred by chance. In addition, the convention presented in Table 2 was also used to guide the interpretation.

Table 2 Interpretation of p-values

p-values	Interpretation
$p \leq .01$	Strong indication that the effect is systematic
$.01 \leq p \leq .05$	Indicates that the effect is systematic
$.05 \leq p \leq .10$	Suggests that the effect might be systematic
$p > .10$	Indicates that the effect occurred by chance

Even though many statistical textbooks include one-tailed significance testing as an option, such tests are often regarded as speculative and are therefore excluded by statistical software packages. [67] developed three criteria that should be fulfilled to justify use of a one-tailed test:

1. Observed difference in the unpredicted direction should be psychologically meaningless.
2. Results in the unpredicted direction should never be used differently than results that occur by chance.
3. The directional hypothesis should be deducible from psychological theory, but results in the opposite direction should not be supported by coexisting theory.

These criteria are extremely hard to satisfy in applied research settings. Consequently, we used two-tailed tests of statistical significance in all situations to be on the safe side.

7.1.5 Practical significance testing

TFSI encouraged researchers to present effect sizes for primary outcomes. Effect size indicators reveal the practical significance of the findings, and inform future power analyses and meta-analyses [66]. However, the interpretation of effect sizes is not straightforward. [68] identify the following problematic issues:

- The impact of an effect is a function of its size, but also the perceived value of a unit change on the dependent variable. For example, a 1 percent reduction in the death rate for cancer is a small effect of a new drug, but would still be regarded scientifically noteworthy due to the

magnitude of the cancer disease and the importance of saving lives. Thus, practical significance is more than effect size.

- The levels of a fixed treatment are arbitrarily determined by the experimenter, and do not represent the population of all possible treatment levels. It is therefore impossible to estimate the population component of variance in the dependent variable attributable to fixed treatments. Thus, effect sizes for fixed effects can only be generalized across studies that have chosen identical levels of experimental manipulation on the fixed treatments.
- The total variance estimate in the denominator of the effect size is strongly affected by subjective judgment. That is, researchers aggregate the data according to their needs and calculate partial effect size indicators [69].
- Comparison of effect sizes across experimental designs with different error terms, sample sizes, and/or number of levels on the manipulated variables can be misleading.

[68] still concludes that it is informative to report effect sizes along with the level of statistical significance for primary results. However, the effect size must be interpreted within specific research contexts and in light of the weaknesses listed above. We used the partial eta-squared to report effect sizes.

7.1.6 Measures of association - correlation and reliability statistics

The interpretation on the size of correlation and reliability varies with experimental contexts and statistics. In this report, the statistic for studying correlating relationships is the Pearson's product-moment correlation coefficient, r , adopting the interpretation presented in Table 3 [70, 71]. For representative experiments, correlation coefficients below .1 generally represent negligible relationships; whereas, coefficients substantially above .5 indicate very high shared variance, which may signal a lack of discriminant validity.

Table 3: Interpretation of Pearson product-moment correlation, r^2 .

Pearson product-moment correlation	Strength of relationship
$.1 < r^2 \leq .3$	Weak
$.3 < r^2 \leq .5$	Moderate
$.5 < r^2$	Strong

The statistics for studying the agreement between raters or inter-rater reliability is Cohen's Kappa, κ , adopting the interpretation presented in Table 4 [72]. For tasks involving substantial judgment (e.g., medical diagnosis [72]), κ typically ranges from .4 to .7.

Table 4: Interpretation of Cohen's Kappa, κ .

Kappa coefficient	Strength of agreement
≤ 0	Poor
$0 < \kappa \leq .2$	Slight
$.2 < \kappa \leq .4$	Fair
$.4 < \kappa \leq .6$	Moderate
$.6 < \kappa \leq .8$	Substantial
$.8 < \kappa \leq 1$	Almost perfect

7.1.7 Confidence intervals

According to TFSI, interval estimates should be given for principal effects. It is also recommended to include interval estimates in graphical representations whenever possible [66]. The rationale for this seems to be that interval estimates give information about precision as well as statistical significance. The most popular interval estimate is the 95% confidence interval and no other alternatives will be considered here.

For within-subject designs, confidence intervals around the cell means are irrelevant for inferences about mean differences [73-75]. This is because the variance among subjects affects the size of the confidence interval even though subject differences are removed from the error term in the repeated measures analysis of variance. [76] demonstrated that the majority of researchers misinterpret graphical representations of confidence intervals for within-subject effects by making incorrect inferences about mean differences.

It is possible to calculate and graph customized within-subject confidence intervals that make it possible to draw correct inferences about mean differences [74, 77]. However, the customized confidence intervals do not provide precise estimates of the cell means. There is limited utility in presenting less accurate interval estimates for within-subject effects in order to enable precise graphical illustrations of information that is captured by statistical significance testing (p-values). Therefore, customized within-subject confidence intervals are not included.

Nevertheless, interval estimates provide valuable information regarding the precision of findings and should be presented for principal effects. Interval estimates for the cell means of within-subject effects are therefore reported in tables below the graphs of the effects.

7.2 Study 1

Study 1 investigated human-system interfaces and operator performance for existing operational concepts in a full-scope simulator environment. Specifically, Study 1 covered the following topics:

- Ecological Interface Design for nuclear process control [13, 78-82];
- Graphically-enhanced mimic displays [83]; and
- Operator performance (including SA) [13, 84].

The full-scope simulator experiment provided a platform to test the Process Overview Measure for evaluating interface design concepts under various realistic operating conditions.

7.2.1 Objective

Study 1 provided empirical data on the Process Overview Measure concerning (i) the sensitivity to experimental manipulations of human system interfaces and operating scenarios, and (ii) the relationships to other human performance indicators. Study 1 also offered the opportunity to check the feasibility of the protocol prescribed by the Process Overview Measure.

7.2.2 Method

7.2.2.1 Participants and operator roles

Six licensed operator crews (n=6) were recruited from a BWR power plant identical to the simulated process. Each crew consisted of one reactor operator (RO) and one turbine operator (TO), responsible for the primary and secondary side of the simulated process, respectively. In two cases, participants currently working as ROs operated the secondary side. This substitution should not affect generalization of the results given that all ROs must previously or currently hold TO licenses. The Ecological displays are only developed for the secondary side; hence, the results and discussion in this section only pertain to the performance data collected on the TOs. (The performance data collected on the ROs was not analyzed here.)

7.2.2.2 Test and simulation environment

The experiment was performed in the HAlden Man-Machine LABoratory (HAMMLAB) on a high fidelity simulator of a 1200 MW boiling water reactor with one turbine and a feedwater tank.

HAMMLAB has three operator workstations and a large-screen display (Figure 3). The RO and TO workstations were used for this study and had eleven 19" LCD screens each (Figure 4). On seven of the screens, the operators could select freely among 64 mimic-displays that represented the sub-systems of the plant. The mimic-displays gave access to detailed process information and were used to interact with the process. The operators could also generate trend graphs of selected system parameters on these screens. (Note that these trend diagrams require operators to select the system parameters and differ from the mini and pop-up trends, which plot preselected parameters.) The remaining four screens on each workstation were spatially dedicated to alarms and event lists. The large-screen display in the middle of the Control Room provided a mimic-based overview of the plant for both operators. Navigation was carried out via keyboard and mouse. The operating procedures used in the experiment were standard paper-based procedures developed for the simulated plant.

An experiment leader, process expert, and laboratory technician managed the simulator, scenarios, and data collection from the Observation Gallery (see, Figure 3). Operator interactions with the interface were logged in the simulator, while activities in the control room were audio and video recorded.

The process expert had several important roles in the experiment, including interface design, scenario development and human performance evaluation. Our expert was an engineer with 15 years of experience from a Swedish nuclear power plant, working 4 years as Turbine Operator, 4 years as Reactor Operator, 4 years as Shift Supervisor, and 3 years responsible for reactor core operation. During these years he was also involved in the construction of the plant, simulator development and operator training. In addition, he had 10 years of experience as a process expert and control room systems designer in HAMMLAB, and contributed strongly to the development of the Boiling Water Reactor (BWR) simulator used for this experiment.

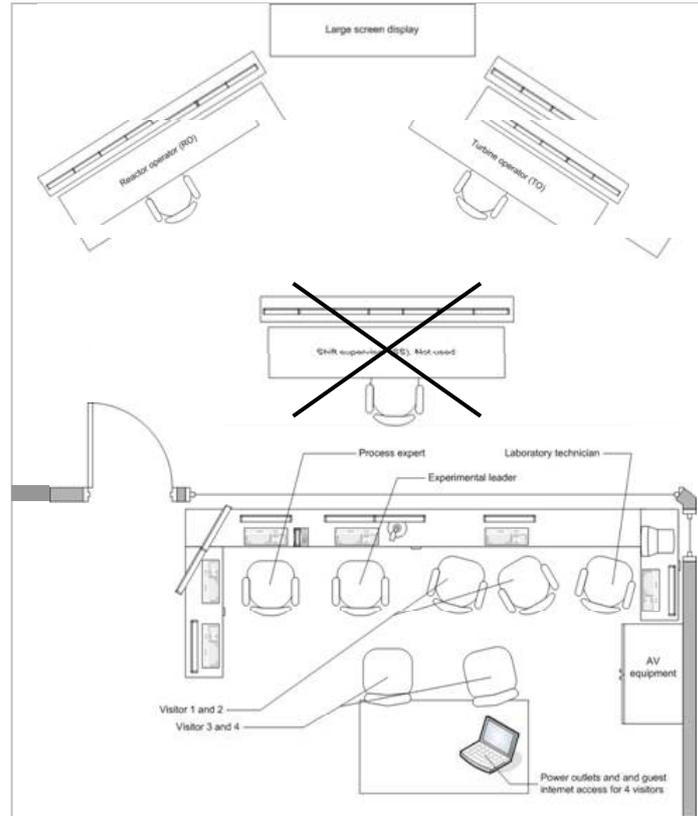


Figure 3: HAMMLAB Control Room and Observation Gallery



Figure 4: LCD panel layout for the turbine operator workstation in the Ecological display condition.

7.2.2.3 Experimental Manipulation

This study consisted of three experimental manipulations – display types, scenario types and scenario phases.

7.2.2.3.1 Display Types

Three display types – Traditional, Advanced, and Ecological – were selected for comparison. Prior to illustrating their distinct characteristics, we first describe the shared features of the interfaces.

The design scope was limited to the turbine side (e.g., Figure 4). Given this scope, alarm information was communicated in the same manner across the display conditions (i.e., they have the same alarm displays). Furthermore, all three types of displays share the general layout shown in Figure 5, in which the gray areas are the same across the three display types. The interaction style was also consistent across display type. Specifically, operators could click on a plant component icon in the format field of any display to access/view the equipment status and related control variables that appeared in the automatic and/or process control fields. To execute control actions, operators could key in desired values

in corresponding variable entry fields. In contrast, for displays that were within the design scope of the ecological displays (i.e., the turbine side), the format fields (i.e., the white area in Figure 5) varied according to the description as follows.

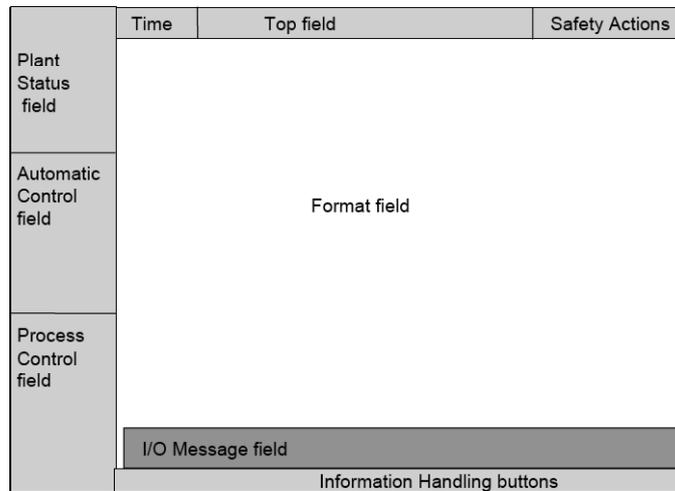


Figure 5: Display layout.

7.2.2.3.2 Traditional Displays

The Traditional displays are the computerized version of the hard-wired wall panels originally installed in the operating nuclear plant (Figure 6). The Traditional displays roughly represent the “state-of-practice” design, characterized by mimic diagrams of the facilities with numeric outputs of instrumentation (i.e., mimic-based displays). Although the design process might not adhere strictly to a particular framework, the design was largely informed by task analyses and possibly some user input as mandated by regulators [85].

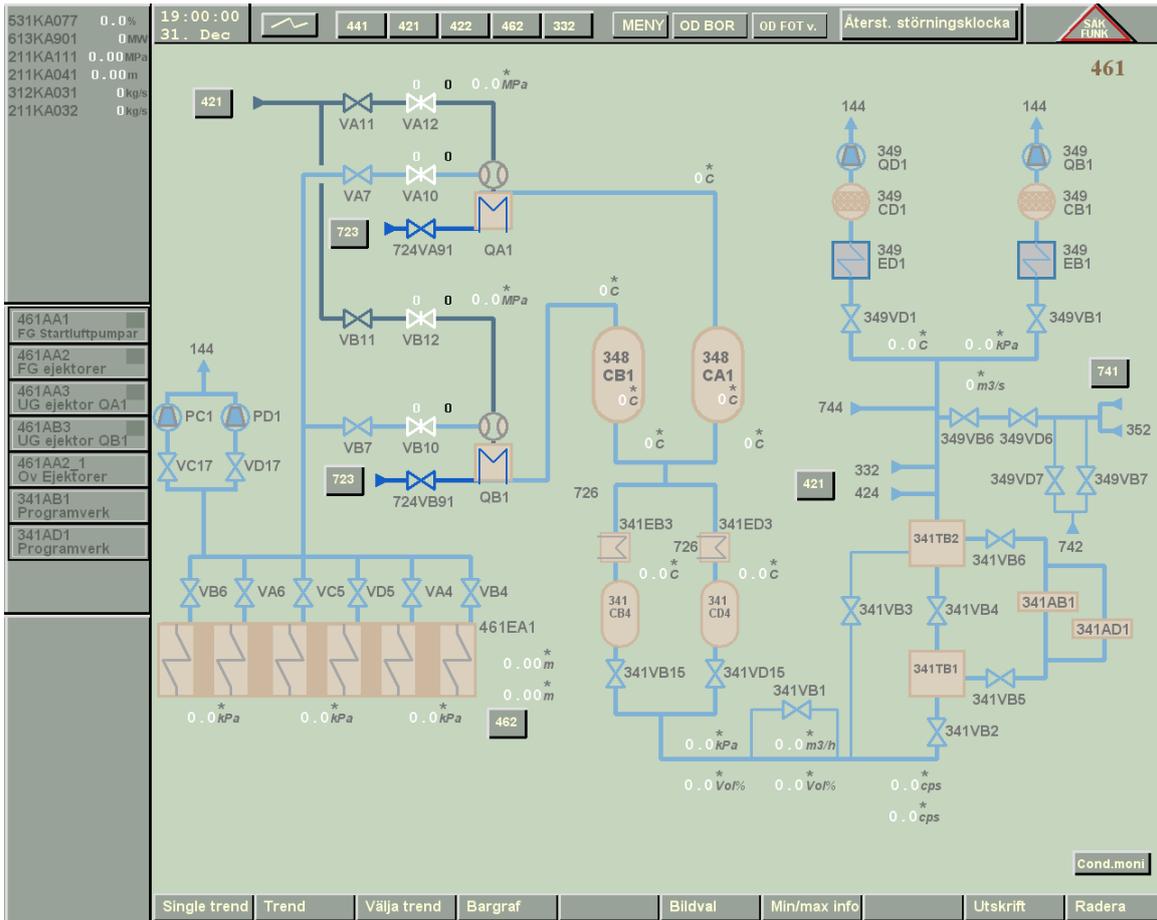


Figure 6: An example of a Traditional display.

7.2.2.3.3 Advanced Displays

The Advanced displays are a graphically enhanced version of the Traditional displays (Figure 7). The Advanced displays retain the mimic-diagrams of the Traditional displays; however, they also contain some configural graphics (e.g., [86, 87]) and “mini-trends” strategically developed and inserted by process experts. [83] provides a detailed account of the graphical enhancements over the Traditional displays. For this display condition, the large screen display contained some advanced visualization features that were absent from both Traditional and Ecological display conditions in addition to the condensed mimic diagrams with key instrumentation outputs.

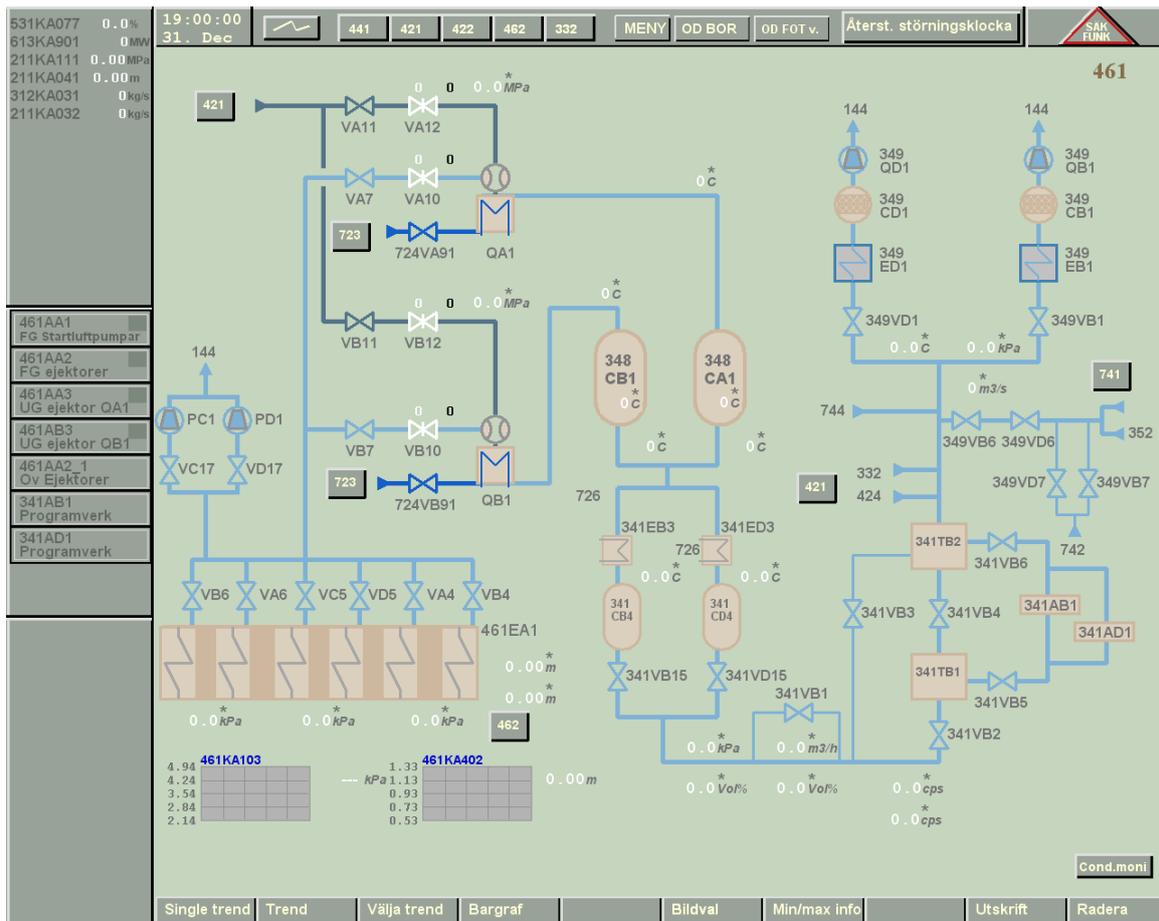


Figure 7: An example of an Advanced display.

7.2.2.3.4 Ecological Displays

The Ecological displays [82, 88] were designed according to the EID framework. In brief, EID is a theoretical framework for designing human computer interfaces for complex systems that claims to enhance operator performance by specifying information requirements and perceptual features based on formative work analysis and cognitive controls, respectively. Figure 8 shows an example of an ecological display. As mentioned, the design scope was limited to the secondary side. The participants had access to the Traditional displays for plant processes that were not represented by the Ecological displays. Furthermore, the operators had access to the large screen display in the Traditional display condition.

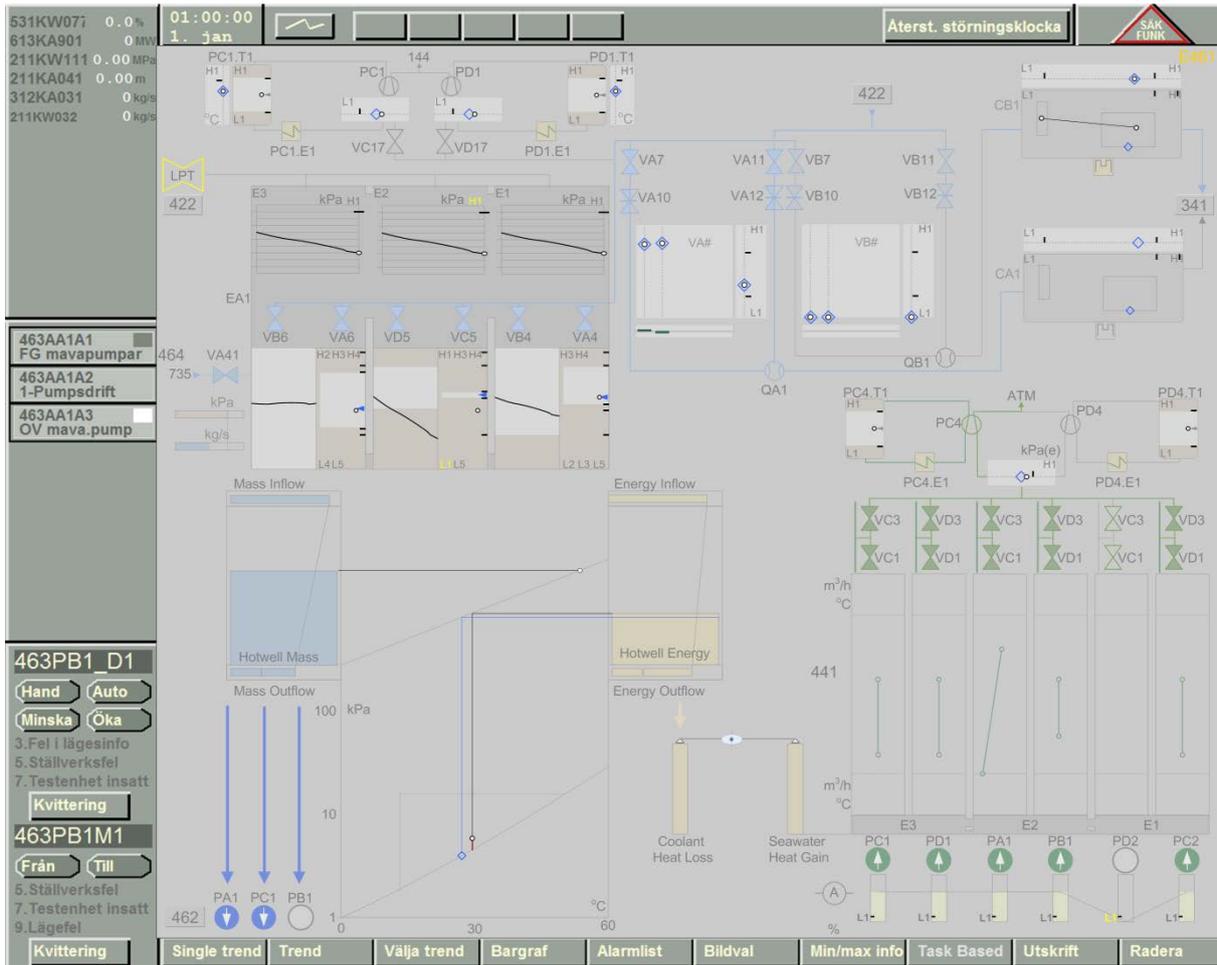


Figure 8: An example of an Ecological display.

7.2.2.4 Scenario Type

This study contained three Procedure-guided and three Knowledge-based scenarios. For the purpose of this study, Procedure-guided scenarios were defined by a set of disturbances that could be resolved by referencing plant procedures. Scenarios in which disturbances could not be resolved by procedures were classified as Knowledge-based. In other words, equipment failures anticipated by the utilities and job responsibilities familiar to operators characterized the Procedure-guided scenarios, while unanticipated and unfamiliar ones characterized the Knowledge-based scenarios. [13] documents the details of each scenario.

7.2.2.5 Scenario Phase

Each scenario started with a “Detection” phase, a time period just before the first alarm sounded, and then ended with a “Mitigation” phase that consisted of all subsequent events. (Figure 9 illustrates the detailed structure of the scenarios.) The two phases afforded separate assessments of the effectiveness of the displays in supporting both monitoring and intervention.

7.2.2.6 Experimental Design

A 3x2x2 RBF (within-subjects) design was employed with treatments of display type (Traditional, Advanced and Ecological), scenario type (Procedure-guided and Knowledge-based), and scenario phase (Detection and Mitigation). The treatments were completely crossed and counterbalanced using a Latin-

square technique. The final experimental design assigned the six crews to the six scenarios that were divided into two phases (N=72).

7.2.2.7 Procedure

The participation of each crew was divided over three consecutive days. The first day was dedicated to the training program after obtaining informed consent and demographic information. Six hours of training occurred on the first day. The second day started with a one-hour training session to refresh the materials presented on the first day, followed by three scenarios with fifteen-minute breaks in between. The third day started with three scenarios also with fifteen-minute breaks in between, followed by a debriefing/closing session.

For all scenarios, crews were asked to maintain the original power level and safe operation. A process expert registered scores to corresponding performance items at various points of the scenarios by observing the participants while they monitored system states and resolved disturbances (see measure of task performance below, section 7.2.2.8.3). The participants also responded to the subjective task-complexity questionnaire during a short simulator freeze and at the end of each scenario. The simulation freeze occurred at the end of the Detection phase, which took up the first five to ten minutes of the scenario as depicted in Figure 9. The scenario then continued with the Mitigation phase, which was marked by the onset of the first alarm within the first minute. The Mitigation phase usually lasted for 30 to 40 minutes, followed by another administration of the subjective task-complexity questionnaire at the end of the scenario.

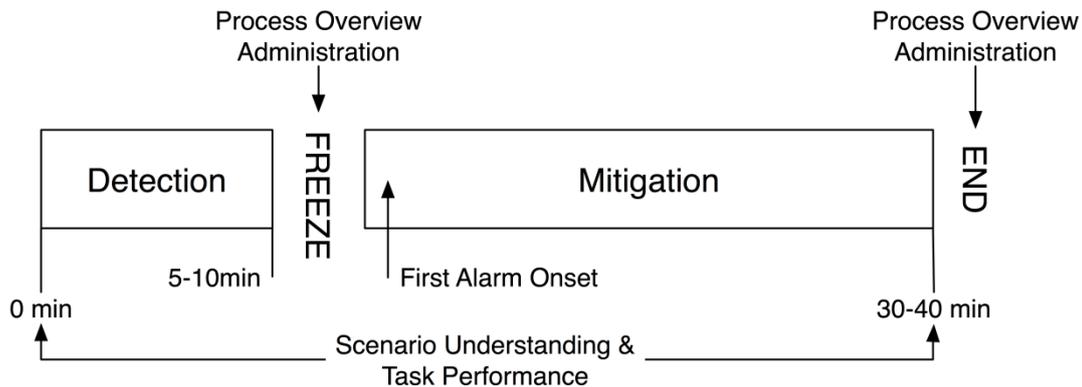


Figure 9: Basic structure of the scenarios in Study 1.

7.2.2.8 Measures

Study 1 employed several operator performance indicators to illustrate the full effect of the experimental manipulations. [13, 80, 84] documents the details of the performance measures and effects of the experimental manipulations. Besides the data collected through the Process Overview Measure, this report also presents the results of scenario understanding and task performance measures that are relevant to exploring sensitivity and construct validity of the Process Overview Measure.

7.2.2.8.1 Process Overview

Process Overview is operationalized with the Process Overview Measure (Chapter 6). In Study 1, the Process Overview Measure included both context- and fault-sensitive queries about the parameter changes in the recent past. The queries were administered during one simulator freeze and at the end of every trial (Figure 9).

To nullify the effect of cuing for fault-sensitive queries, the Mitigation phase (i.e., second phase of the scenario) began with an alarm that alerted the operators to faults developing during the Detection phase (i.e., first phase of the scenario).

7.2.2.8.2 Scenario Understanding

Scenario Understanding [13] refers to the operator understanding of the process faults introduced in the scenarios, essentially reflecting the quality of diagnosis. Study 1 operationalized Scenario Understanding with the Halden Open Probe Elicitation (HOPE) measure [13].

HOPE employs process experts to rate operators understanding of the scenario on a four-point-anchored scale. The process experts can perform their ratings based on (i) observation (e.g., conversation between operating crew members) from the experimental gallery, and (ii) telephone inquiries to operators about the plant states. Process experts or experimental leaders, disguised as plant management, telephone operators to inquire about plant states at strategically predefined times of the scenarios. HOPE prescribes a list of “open” probes, which do not refer to any scenario-specific characteristics, to avoid cuing effects. For example, a probe can be: “Is everything normal?” During the telephone calls, the simulator is frozen but the operators can access all process displays.

For Study 1, a process expert conducted two ratings in each scenario phase. The final scores for each phase was the average of the two ratings.

7.2.2.8.3 Task Performance

Task performance refers to the overall effectiveness of control room operations demonstrated by the operators. Study 1 operationalized task performance with the Operator Performance Assessment System (OPAS) [13, 52, 89].

OPAS provides a structure for the assessment of whether operators carry out their task work in accordance with scenario solutions prescribed a priori by experts in control room operation. Prior to data collection, process experts analyzed the scenarios and developed optimal solutions by identifying items that expressed the desired performance. During the experiment, a process expert registered the points earned by operators in completing the predefined activities within each performance item based on observations of operator verbalization, physical behaviours, problem solving, and system states. In Study 1, the items were not divided into any category. The employed performance index is the unweighted average of all performance items defined for a scenario. The OPAS index reflects the degree of conformance between operator performance and predefined optimal solutions to scenarios.

7.2.2.9 Evaluation Criteria (Hypothesis)

In Study 1, the Process Overview Measure should demonstrate sensitivity and practicality for assessing monitoring performance in a representative nuclear process control setting.

Sensitivity is demonstrated by significance testing that corroborates with other empirical data in the study and general indications from the literature. Specifically the Process Overview Measure should reveal:

- slight to moderate correlation (see e.g. [70]) with measures of Scenario Understanding and task performance (as an indication of validity),
- effects of the human-system interface manipulation (as the literature (e.g., [86]) suggests that graphical representation beyond mimic-based displays could improve monitoring),
- effects of the scenario manipulations (as the literature (e.g., [90]) suggests that different types of situations could impact monitoring, particularly for fault-sensitive parameters), and

- effects that are consistent with the Scenario Understanding, task performance, and debriefing data.

Feasibility is demonstrated by qualitative feedback from the process experts and participants in Study 1.

7.2.3 Analysis and Results

The analysis and results of Study 1 are documented in four parts: (i) descriptive statistics to qualitatively inspect distributions of the Process Overview measurements and search for threats to robustness of ANOVA; (ii) correlations with other measures to identify convergent and discriminant relationships; (iii) ANOVA results to assess sensitivity of the measure; and (iv) findings from other measures to confirm the ANOVA results.

7.2.3.1 Descriptive Statistics

For Study 1, the descriptive statistics are presented in the following sequence: (i) context-sensitive Process Overview, (ii) fault-sensitive Process Overview, and (iii) overall Process Overview (i.e., aggregating both context- and fault-sensitive items).

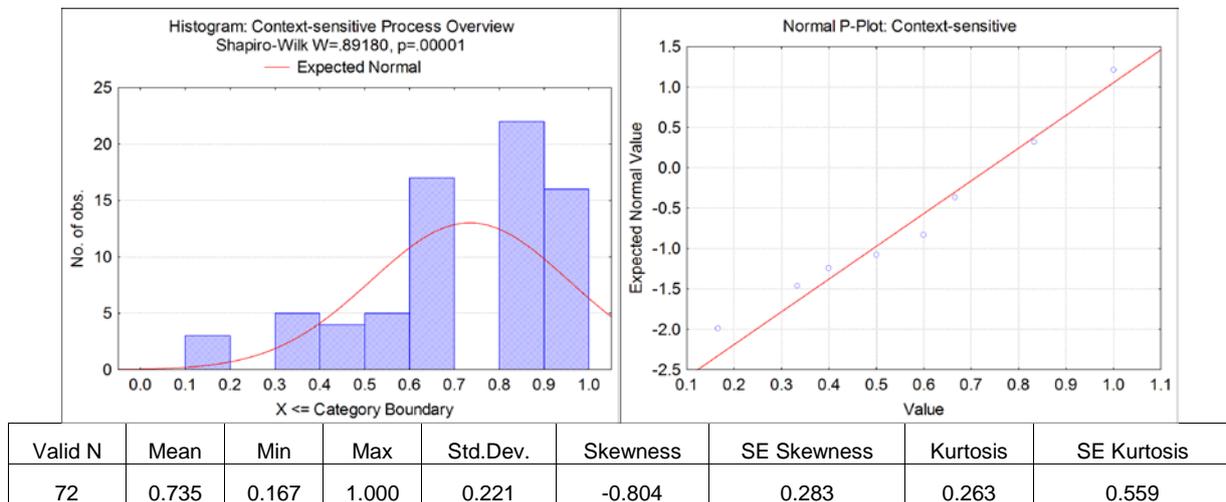
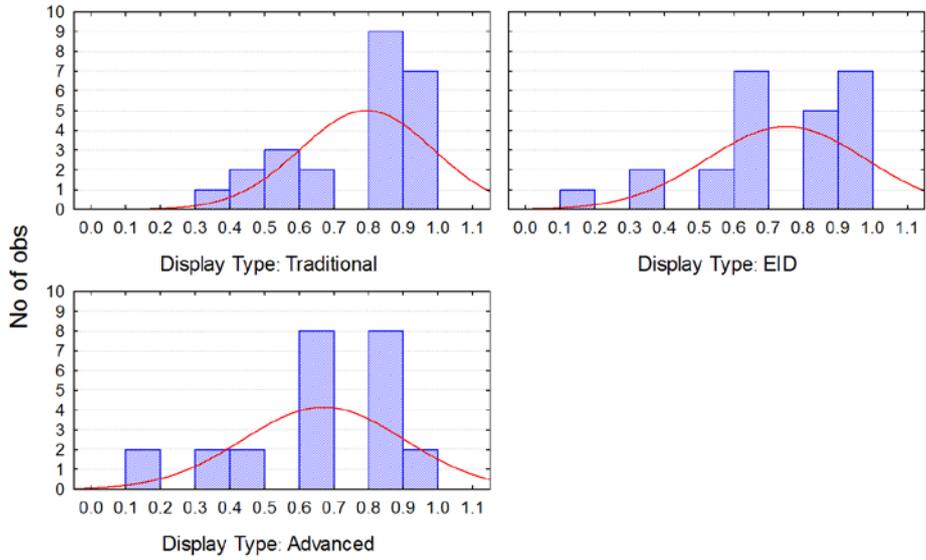


Figure 10: Descriptive statistics of context-sensitive Process Overview measurements in Study 1.

The context-sensitive data of the Process Overview Measure (Figure 10) is not normally distributed ($W=0.89$, $p=0.00$). The distribution is negatively skewed, exhibiting a slight ceiling effect. Because routine monitoring practice typically includes these parameters, slight ceiling effect is not concerning. The intra-class correlation for the measurements of context-sensitive Process Overview amongst the participants is $ICC(3,1)=-0.02$, indicating negligible within-subject effects across experimental conditions and suggesting statistical analysis based on CRF design as opposed to a RBF design (see section 7.1)

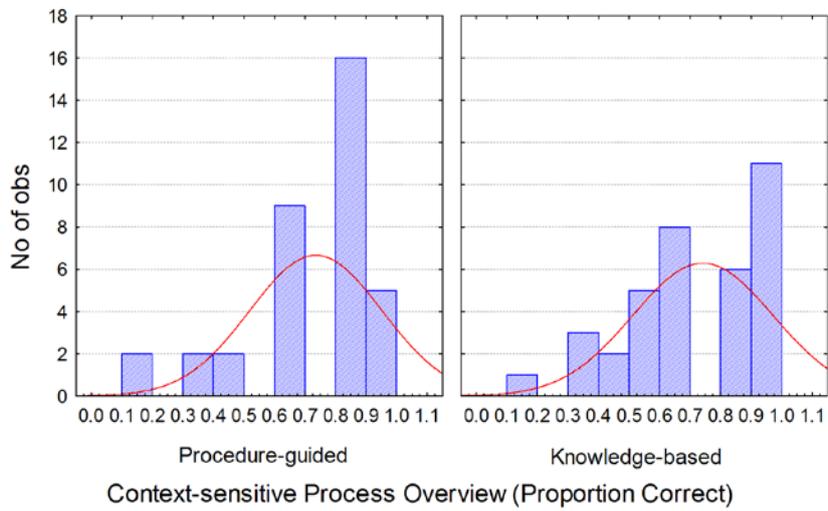
Histograms by Display Type: Context-sensitive Process Overview



Context-sensitive Process Overview (Proportion Correct)

Figure 11: Histograms by display type for context-sensitive Process Overview in Study 1.

Histograms by Scenario Type: Context-sensitive Process Overview



Context-sensitive Process Overview (Proportion Correct)

Figure 12: Histograms by scenario type for context-sensitive Process Overview in Study 1.

Histograms by Scenario Phase: Context-sensitive Process Overview

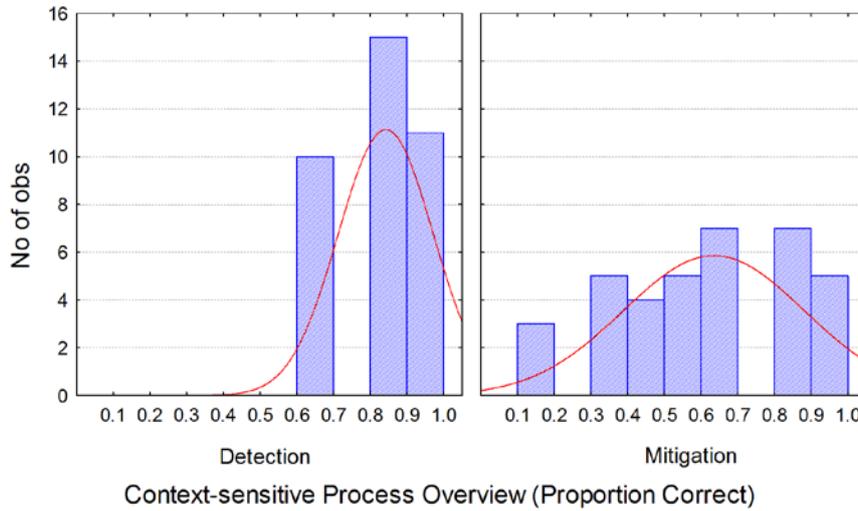


Figure 13: Histograms by scenario phase for context-sensitive Process Overview in Study 1.

Figure 11-Figure 13 present the distributions of the context-sensitive Process Overview data by individual treatments (i.e., display type, scenario type, and scenario phase). Figure 13 indicates a ceiling effect of context-sensitive Process Overview in the Detection phase. Again, the particular ceiling effect is not surprising. During the Detection phase when the first process fault is only developing, operators probably focused on context sensitive parameters, which were typically part of the routine monitoring practice. The histograms within the treatments appear similar (i.e, skewness and kurtosis in the same direction) and thus do not indicate any threats to the robustness of ANOVA results.

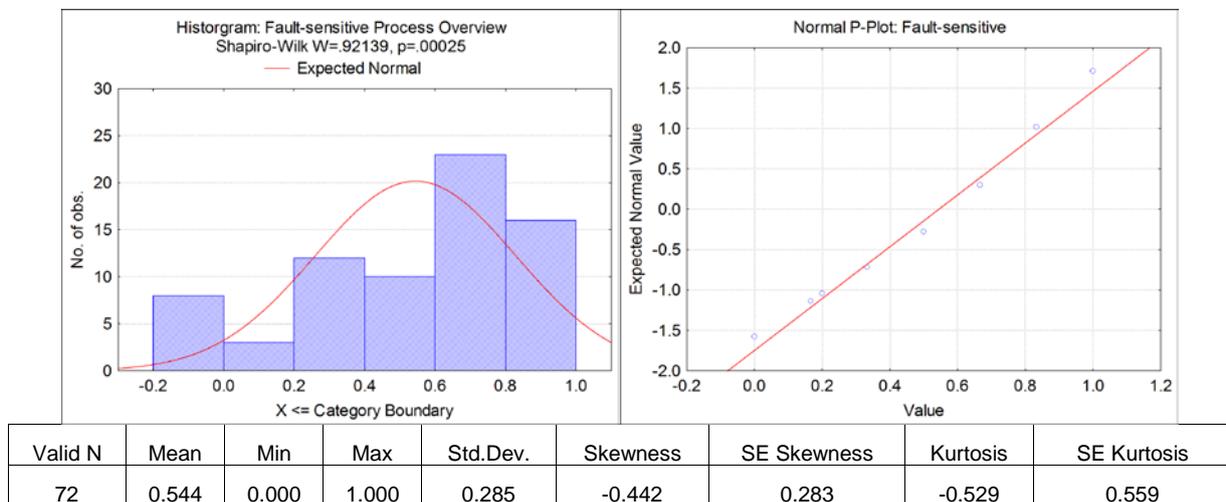
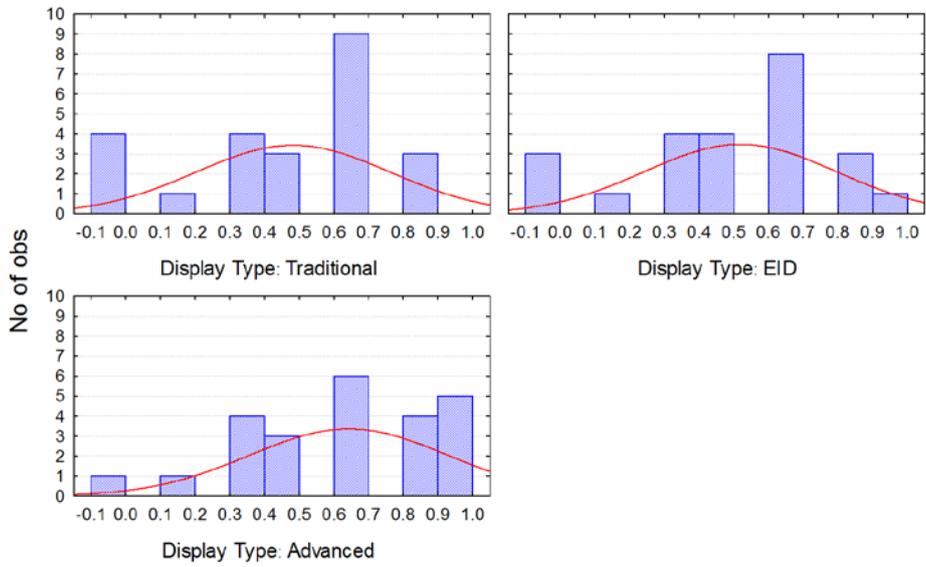


Figure 14: Descriptive statistics of fault-sensitive Process Overview measurements in Study 1.

The fault-sensitive data of the Process Overview Measure (Figure 14) is not normally distributed ($W=0.92$, $p=0.00$). The distribution is slightly negatively skewed and does not appear to pose significant ceiling effect. The intra-class correlation for the measurements of fault-sensitive Process Overview amongst the participants is $ICC(3,1)=-0.08$, indicating limited within-subject effects across experimental conditions and suggesting statistical analysis based on CRF design as opposed to a RBF design (see section 7.1)

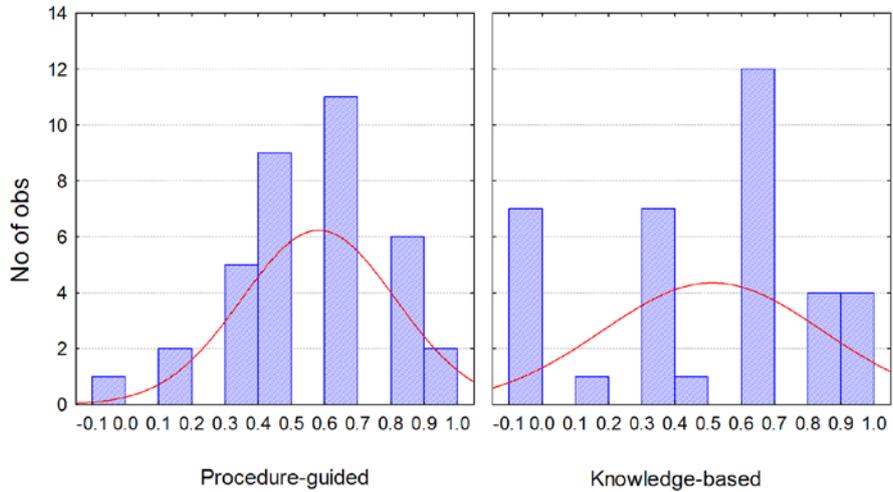
Histograms by Display Type: Fault-sensitive Process Overview



Fault-sensitive Process Overview (Proportion Correct)

Figure 15: Histograms by display type for fault-sensitive Process Overview in Study 1.

Histograms by Scenario Type: Fault-sensitive Process Overview



Fault-sensitive Process Overview (Proportion Correct)

Figure 16: Histograms by scenario type for fault-sensitive Process Overview in Study 1.

Histograms by Scenario Phase: Fault-sensitive Process Overview

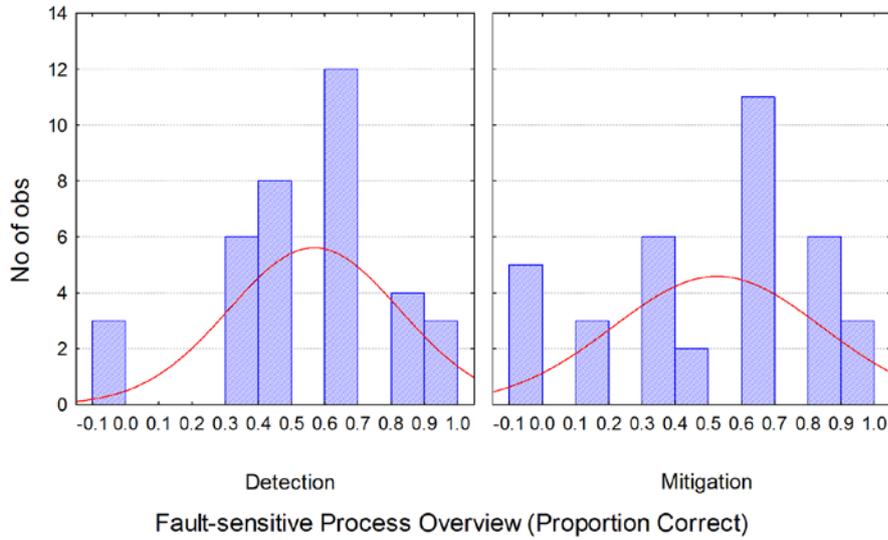


Figure 17: Histograms by scenario phase for fault-sensitive Process Overview in Study 1.

Figure 15-Figure 17 present the distributions of the context-sensitive Process Overview data by individual treatments. The histograms within the treatments appear similar (i.e., skewness and kurtosis in the same direction) and thus do not indicate any threats to the robustness of ANOVA results.

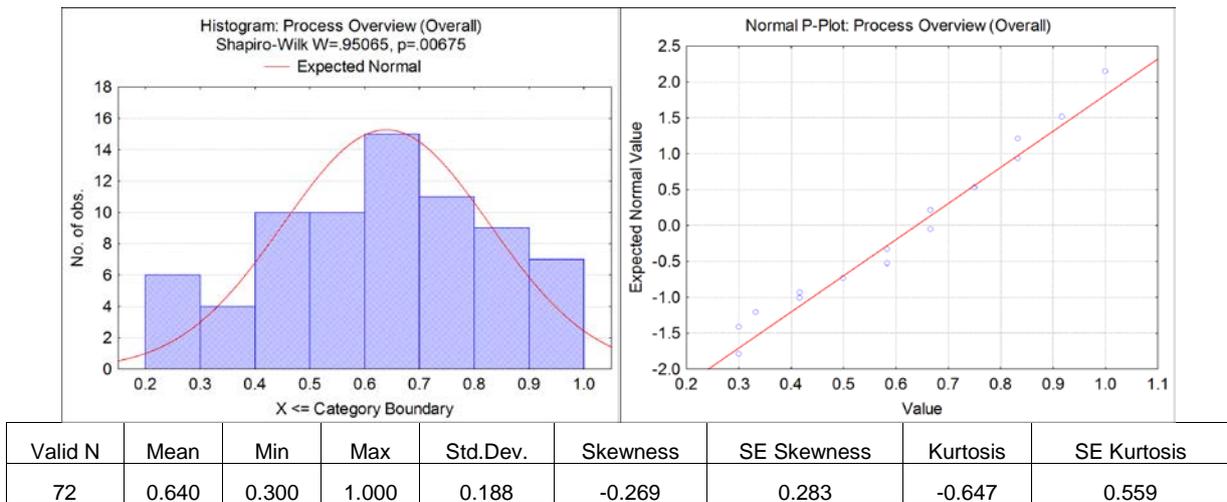


Figure 18: Descriptive statistics of Process Overview measurements (Context- and Fault-sensitive combined) in Study 1.

The aggregated data (combining context- and fault-sensitive items; Figure 18) is not normally distributed ($W=.95$, $p=.01$). The distribution does not show any alarming signs of ceiling and floor effects. The intra-class correlation for the measurements of fault-sensitive Process Overview amongst the participants is $ICC(3,1)=-.05$, indicating limited within-subject effects across experimental conditions and suggesting statistical analysis based on CRF design as opposed to a RBF design (see section 7.1)

Histograms by Display Type: Process Overview (Overall)

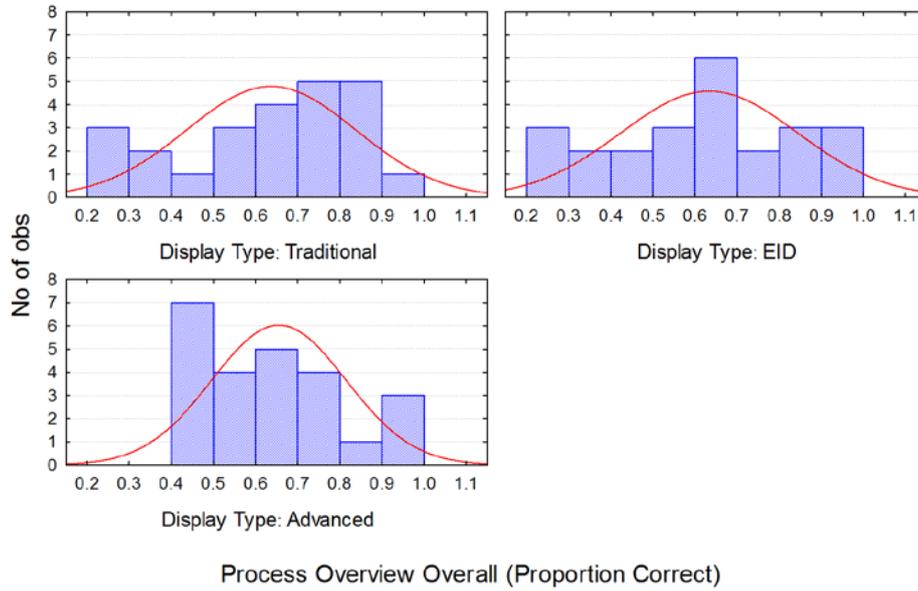


Figure 19: Histograms by display type for Process Overview (context- and fault-sensitive items combined) in Study 1.

Histograms by Scenario Type: Process Overview (Overall)

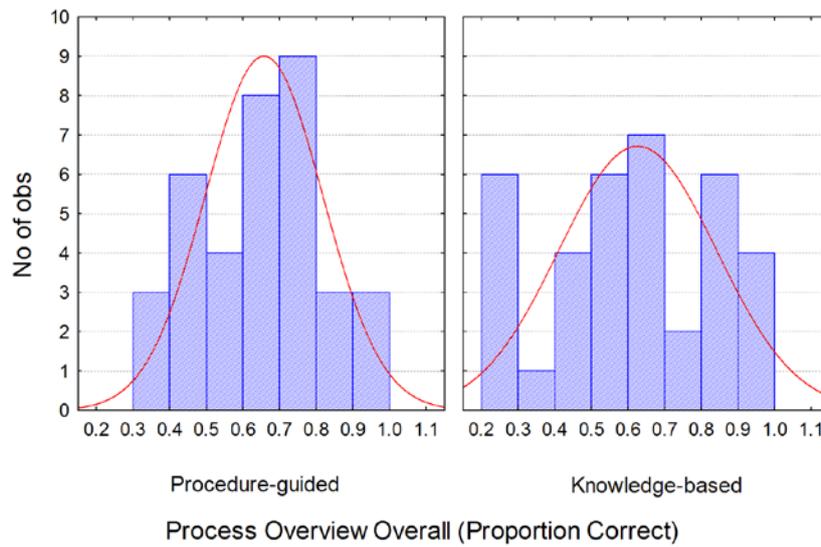


Figure 20: Histograms by scenario type for Process Overview (context- and fault-sensitive items combined) in Study 1.

Histograms by Scenario Phase: Process Overview (Overall)

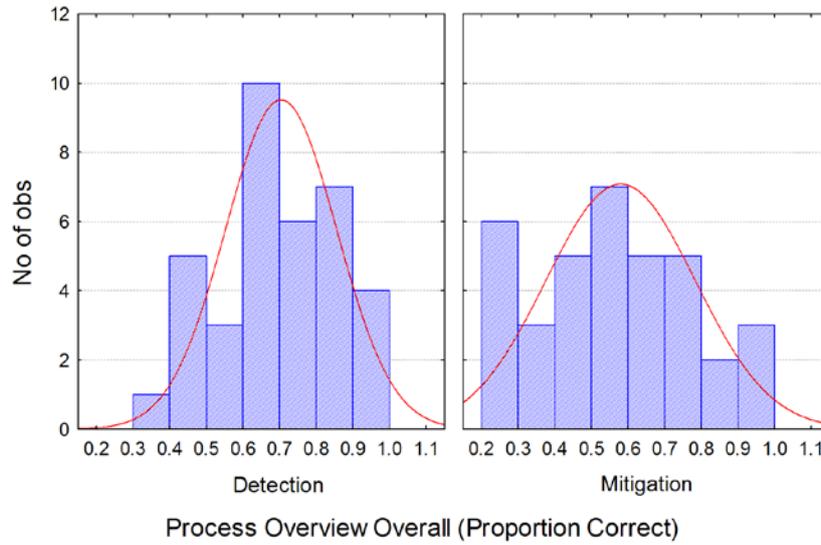


Figure 21: Histograms by scenario phase for Process Overview (context- and fault-sensitive items combined) in Study 1.

Figure 19-Figure 21 present the distributions of the overall Process Overview data by individual treatments. Figure 19 indicates small variation in form amongst the histograms of different display types. In particular, the histograms of Traditional and Advanced displays are slightly skewed in the opposite direction. Other distributions appear similar in form, indicating no threat to robustness of ANOVA. In summary, ANOVA effects ($p \leq .05$) associated with displays should be interpreted with some care. Marginal significance ($.05 < p \leq .1$) of ANOVA effects should be dismissed.

Box & Whisker Plot: All Process Overview Measurements

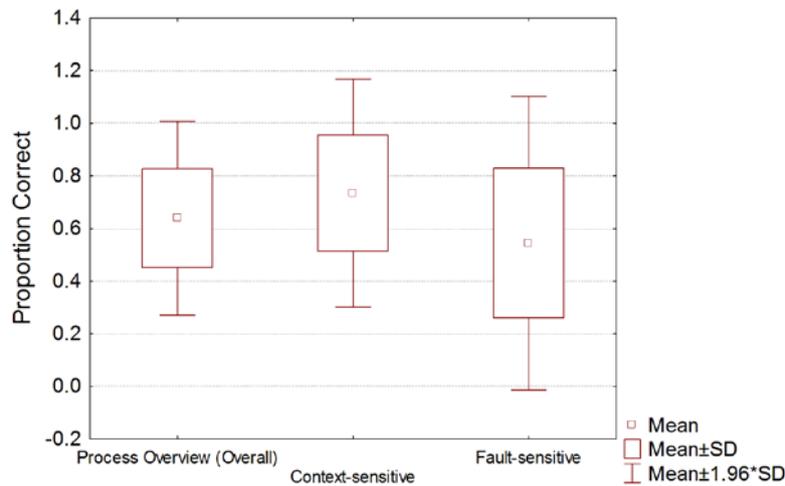


Figure 22: Box and whisker plot of Process Overview measurements in Study 1.

The box and whisker plot (Figure 22) presents an overview of the measurements. As mentioned, by comparison to the overall and fault-sensitive Process Overview, the context-sensitive items show some ceiling effects.

7.2.3.2 Correlations

Table 5: Correlation between Process Overview and other performance indicators.

<i>r(72)</i>	<i>Process overview</i>	<i>Context-sensitive Process Overview</i>	<i>Fault-sensitive Process Overview</i>
Context-sensitive Process Overview	.6574 p=.000	-	.0937 p=.434
Fault-sensitive Process Overview	.8118 p=.000	.0937 p=.434	-
Scenario Understanding (HOPE)	-.0341 p=.776	-.2308 p=.051	.1338 p=.263
Task Performance (OPAS)	-.0154 p=.898	-.0766 p=.522	.0390 p=.745

The Process Overview measurements generally do not correlate with Scenario Understanding and task performance measurements (Table 5). Context-sensitive Process Overview correlates slightly and negatively with Scenario Understanding ($r(72)=-.23$, $p=0.05$) in an unexpected direction. That is, context-sensitive Process Overview declined as Scenario Understanding improved. This finding is not very concerning because of the low shared variance ($r^2=.05$).

7.2.3.3 Sensitivity (ANOVA effects)

To test the sensitivity of the measure, the context-sensitive, fault-sensitive and aggregated Process Overview data were analyzed with ANOVA using Type III/Unique sums of squares with fixed factors of display types (Traditional, Ecological and Advanced), scenario type (Procedure-guided and Knowledge-based), and scenario phase (Detection and Mitigation) in a CRF design. Note that a CRF design was selected because of the low intra-class correlation within subjects as documented in section 7.2.3.1. (Refer to 7.1.3 for a general discussion.) In addition, confidence intervals for results analysed with a CRF design are presented below because they do not suffer from the interpretation problem associated with RBF design.)

Table 6: ANOVA of context-sensitive Process Overview in Study 1.

<i>Fixed Effects</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
Display	0.18898	2	0.09449	2.4165	0.10	0.08
Scenario	0.00076	1	0.00076	0.0193	0.89	0.00
Phase	0.76742	1	0.76742	19.6263	0.00	0.25
Display*Scenario	0.01892	2	0.00946	0.2419	0.79	0.01
Display*Phase	0.00225	2	0.00113	0.0288	0.97	0.00
Scenario*Phase	0.10631	1	0.10631	2.7188	0.10	0.04
Display*Scenario*Phase	0.02355	2	0.01177	0.3011	0.74	0.01
Error	2.34611	60	0.03910			

The ANOVA on the context-sensitive Process Overview data (Table 6) reveals: (a) the main effect of display type ($F(2,60)=2.42$, $p=.10$, $\eta^2=.08$); (b) the main effect of scenario phase ($F(1,60)=19.63$, $p=.00$, $\eta^2=.25$), and (c) the interaction effect of scenario type and scenario phase ($F(1,60)=2.72$, $p=.10$, $\eta^2=.04$). However, Levene's test only indicates homogeneity of variance for the display type ($F(2,69)=.28$, $p=.75$) but heterogeneity of variance for scenario phase ($F(1,70)=11.84$, $p=0.00$) and the scenario type and scenario phase interaction ($F(3,68)=4.57$, $p=0.01$). Furthermore, there appears to be a correlation between means and standard deviations for the two effects that violates the homogeneity of variance.

A non-parametric, Mann-Whitney U test on scenario phase ($U(36)=310.00$, $z=3.80$, $p=.00$), which is robust to deviation of homogeneity of variance, confirms the ANOVA scenario phase effect.

As non-parametric test is not available to test interaction effects, so an alternative procedure was applied. The difference scores between Detection and Mitigation phase for each scenario type were first calculated and then a non-parametric test on scenario type based on the difference scores was performed¹¹. A Mann-Whitney U test indicates Procedure-guided scenarios differ from Knowledge-based scenarios in terms of their changes from Detection to Mitigation phase ($U(18)=105.00$, $z=1.79$, $p=.07$). This provides additional confidence on the two-way interaction effect discovered by the ANOVA.

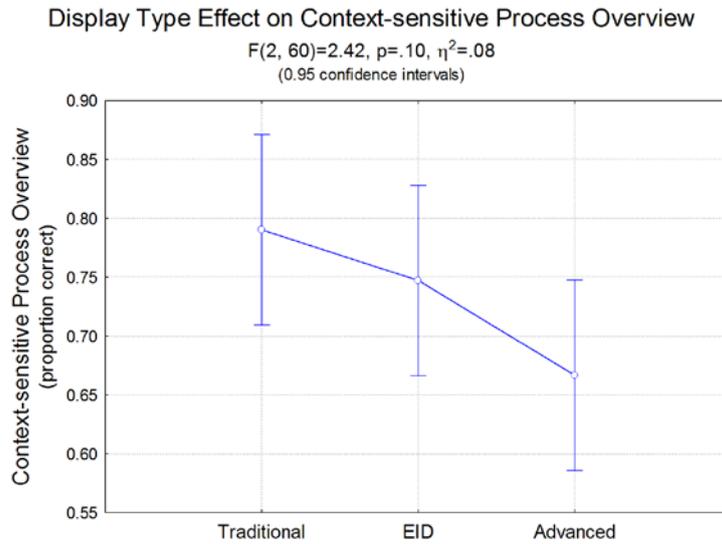


Figure 23: Display type main effect on context-sensitive Process Overview in Study 1.

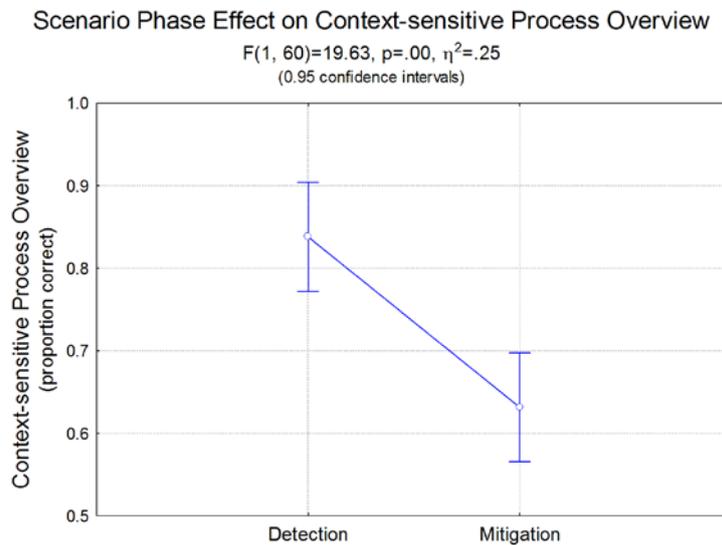


Figure 24: Scenario phase effect on context-sensitive Process Overview in Study 1.

¹¹ This method of verifying interaction effects is not a standard non-parametric statistical procedure. We adapted several common statistical methods to test for two-way interaction effects in which one of the treatments only has two levels.

Scenario Type*Scenario Phase Effect on Context-sensitive Process Overview
 $F(1, 60)=2.72, p=.10, \eta^2=.04$
 (0.95 confidence intervals)

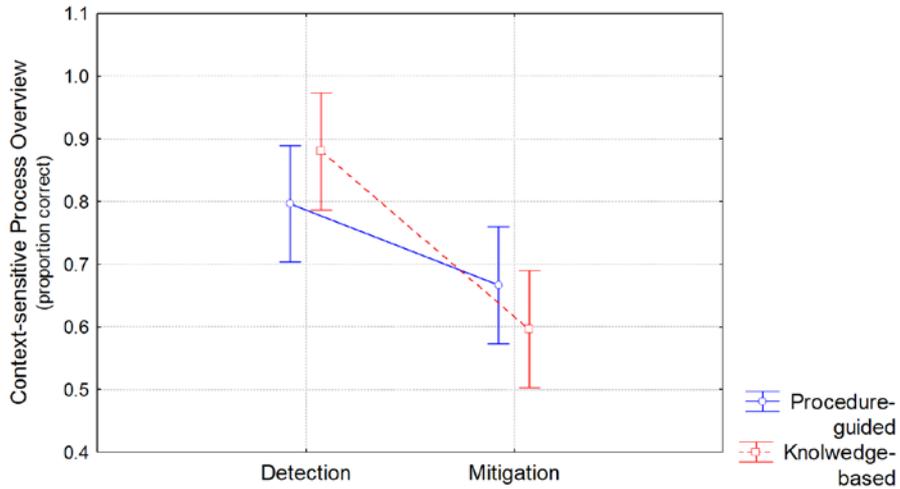


Figure 25: Scenario type and scenario phase interaction effect on context-sensitive Process Overview in Study 1.

Figure 23 illustrates that the Traditional displays provide the most support whereas the Advanced displays provide the least support for monitoring parameters relevant to the general operating contexts (e.g., full power steady-states). Figure 24 and Figure 25 illustrate that context-sensitive Process Overview declines from the Detection to Mitigation phase, and more substantially for Knowledge-based than Procedure-guided scenarios.

Table 7: ANOVA of fault-sensitive Process Overview in Study 1.

Fixed Effects	SS	df	MS	F	p	η^2
Display	0.33676	2	0.16838	2.2807	0.11	0.07
Scenario	0.08451	1	0.08451	1.1446	0.29	0.02
Phase	0.02988	1	0.02988	0.4047	0.53	0.01
Display*Scenario	0.18429	2	0.09215	1.2481	0.29	0.04
Display*Phase	0.08429	2	0.04215	0.5709	0.57	0.02
Scenario*Phase	0.01784	1	0.01784	0.2416	0.63	0.00
Display*Scenario*Phase	0.58614	2	0.29307	3.9697	0.02	0.12
Error	4.42963	60	0.07383			

The ANOVA on the fault-sensitive Process Overview data reveals the interaction effect of display type, scenario type and scenario phase ($F(2, 60)=3.97, p=.02, \eta^2=.12$). Levene’s test indicates homogeneity of variance for the three-way interaction effect ($F(11,60)=1.03, p=.43$).

Display Type*Scenario Type*Scenario Phase Effect on Fault-sensitive Process Overview

$F(2, 60)=3.97, p=.024, \eta^2=.12$
(0.95 confidence intervals)

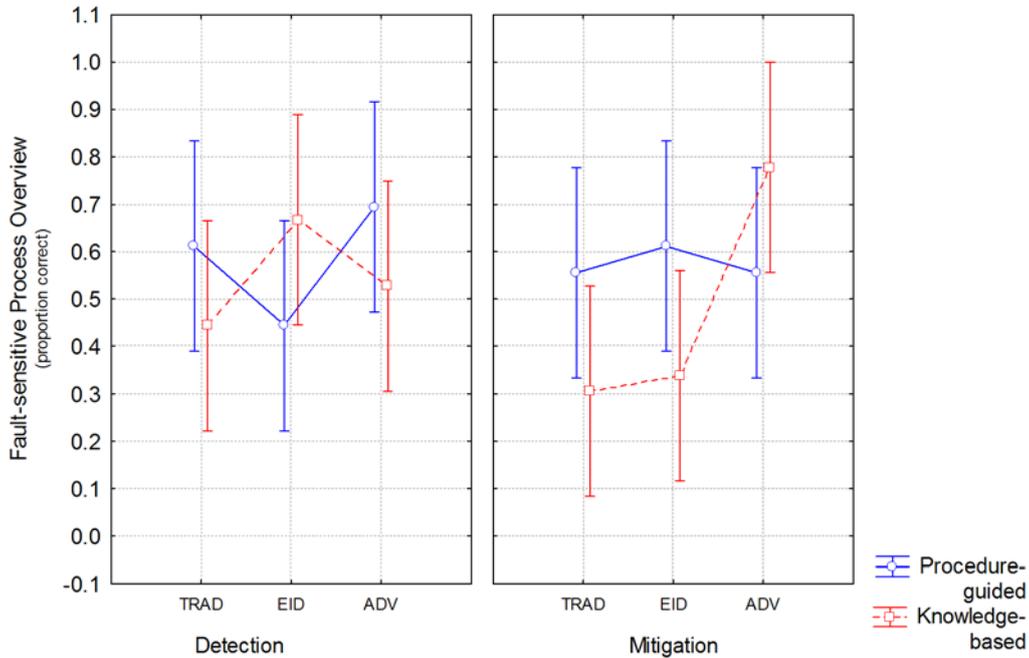


Figure 26: Display type, scenario type and scenario phase interaction effect on fault-sensitive Process Overview measurements for Study 1.

Figure 26 illustrates that the display types provide a different level of support for monitoring fault-sensitive process parameters under the different experimental scenario manipulations. Two observations are particularly noteworthy. First, in the Detection phase, Ecological displays demonstrate inferior support in Procedure-guided but superior support in the Knowledge-based scenarios relative to the Traditional and Advanced displays. Second, the Advanced displays demonstrate superior support during the Mitigation phase in Knowledge-based scenarios.

Table 8: ANOVA of overall Process Overview (context/ and fault-sensitive content combined) in Study 1.

Fixed Effects	SS	df	MS	F	p	η^2
Display	0.00641	2	0.00321	0.1013	0.90	0.00
Scenario	0.01732	1	0.01732	0.5471	0.46	0.01
Phase	0.27503	1	0.27503	8.6881	0.00	0.13
Display*Scenario	0.05047	2	0.02524	0.7972	0.46	0.03
Display*Phase	0.01475	2	0.00737	0.2329	0.79	0.01
Scenario*Phase	0.05281	1	0.05281	1.6683	0.20	0.03
Display*Scenario*Phase	0.19461	2	0.09730	3.0737	0.05	0.09
Error	1.89940	60	0.03166			

Aggregating the context- and fault-sensitive measurements, the ANOVA on Process Overview reveals: (a) the main effect of scenario phase ($F(1, 60)=8.67, p=.00, \eta^2=.13$); and (b) the interaction effect of display type, scenario type and scenario phase ($F(2, 60)=3.07, p=.05, \eta^2=.09$). However, Levene’s test only indicates homogeneity of variance for the three-way interaction effect ($F(11,60)=1.44, p=.18$) but heterogeneity of variance for scenario phase ($F(1,70)=3.30, p=0.07$). Furthermore, there appears to be a correlation between means and standard deviations between the two scenario phases.

A non-parametric, Mann-Whitney U test, which is robust to deviation of homogeneity of variance, confirms the ANOVA scenario phase effect ($U(36)=395.50, z=2.84, p=.00$).

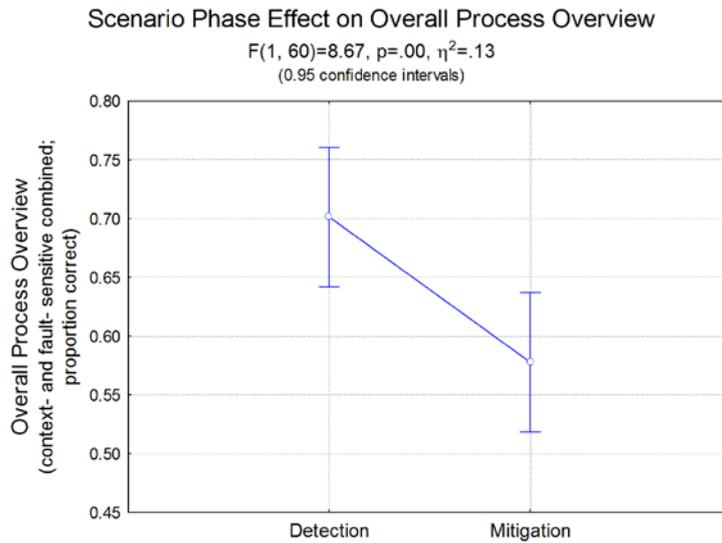


Figure 27: Scenario phase effect on overall Process Overview (context- and fault-sensitive combined) in Study 1.

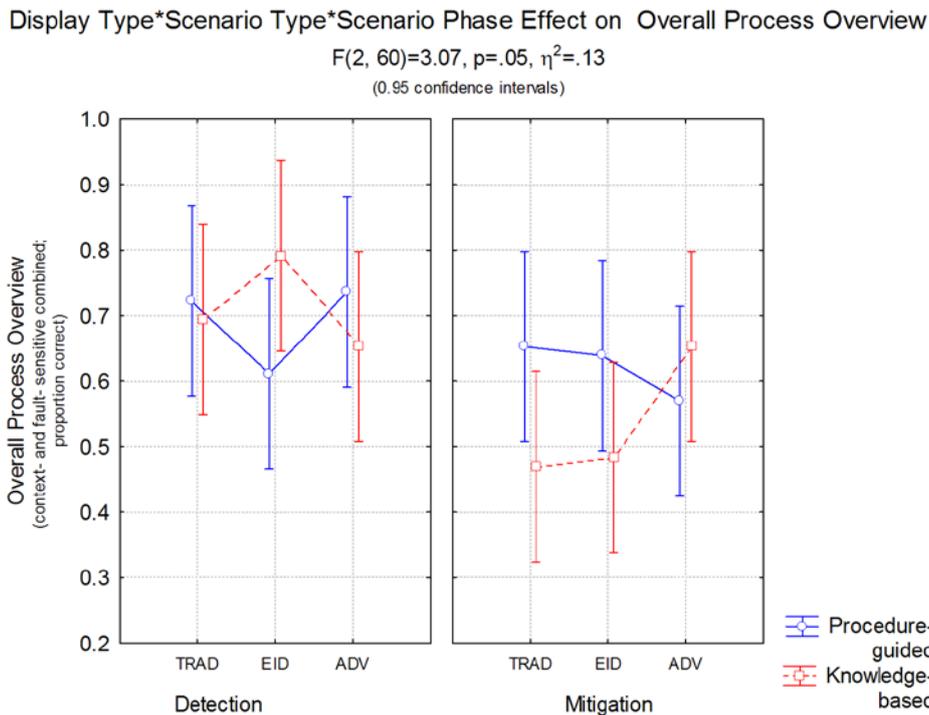


Figure 28: Display type, scenario type and scenario phase interaction effect on Process Overview (context- and fault-sensitive combined) for Study 1.

Figure 27 illustrates a decline of Process Overview from the Detection to Mitigation phase. Figure 28 presents a finding similar to the three-way interaction effect revealed by the fault-sensitive Process Overview (Figure 26). The noteworthy observations are: (a) during the Detection phase, Ecological displays demonstrate inferior support in Procedure-guided but superior support in the Knowledge-based scenarios relative to the Traditional and Advanced displays; and (b) the Advanced displays demonstrates superior support during the Mitigation phase in Knowledge-based scenarios, relative to the Traditional and Ecological displays.

7.2.3.4 Other empirical evidence

Study 1 also provides empirical results on Scenario Understanding and task performance, as operationalized by HOPE and OPAS, respectively, for comparison to the Process Overview findings.

The correlation between measurements of HOPE and OPAS is strong ($r(72)=.72$, $p=.00$), indicating a substantial overlap between the two measures and/or constructs (i.e., lack of discriminant properties).

Scenario Understanding (HOPE) data were analyzed with a multivariate ANOVA using Type III/Unique sums of squares with fixed factors of display types (Traditional, Ecological and Advanced), scenario type (Procedure-guided and Knowledge-based), and scenario phase (Detection and Mitigation) in a RBF design. Unlike all the Process Overview measurements ($ICC(3,1)<.1$), the HOPE measurements demonstrate fair within-subject variance across experimental conditions ($ICC(3,1)=.20$) to conduct analysis in a RBF design. (Refer to section 7.1.3 on selection between CRF and RBF designs.) Note that confidence intervals are omitted for presenting the results analyzed in RBF due to interpretation problems (see section 7.1.7). [13, 79] document and discuss the details of the analysis and results pertaining to Scenario Understanding in Study 1.

Table 9: Multivariate ANOVA of Scenario Understanding (HOPE) in Study 1.

<i>Fixed Effects</i>	<i>Test</i>	<i>Value</i>	<i>F</i>	<i>Effect - df</i>	<i>Error - df</i>	<i>p</i>	η^2
Display	Wilks	0.834692	0.39609	2	4	0.70	0.17
Scenario	Wilks	0.993827	0.03106	1	5	0.87	0.01
Phase	Wilks	0.196970	20.38462	1	5	0.01	0.80
Display*Scenario	Wilks	0.166899	9.98333	2	4	0.03	0.83
Display*Phase	Wilks	0.416032	2.80732	2	4	0.17	0.58
Scenario*Phase	Wilks	0.940529	0.31616	1	5	0.60	0.06
Display*Scenario*Phase	Wilks	0.060476	31.07087	2	4	0.00	0.94

The multivariate ANOVA on Scenario Understanding reveals: (a) the main effect of scenario phase ($F(1, 5)=20.38$, $p=.01$); (b) the interaction effect of display type and scenario type ($F(2, 4)=9.98$, $p=.03$); and (c) the interaction effect of display type, scenario type and scenario phase ($F(2, 4)=31.07$, $p=.00$)¹².

¹² Effect sizes for CRF are drastically different from RBF designs. Though presented, effect sizes should not be compared between CRF and RBF designs.

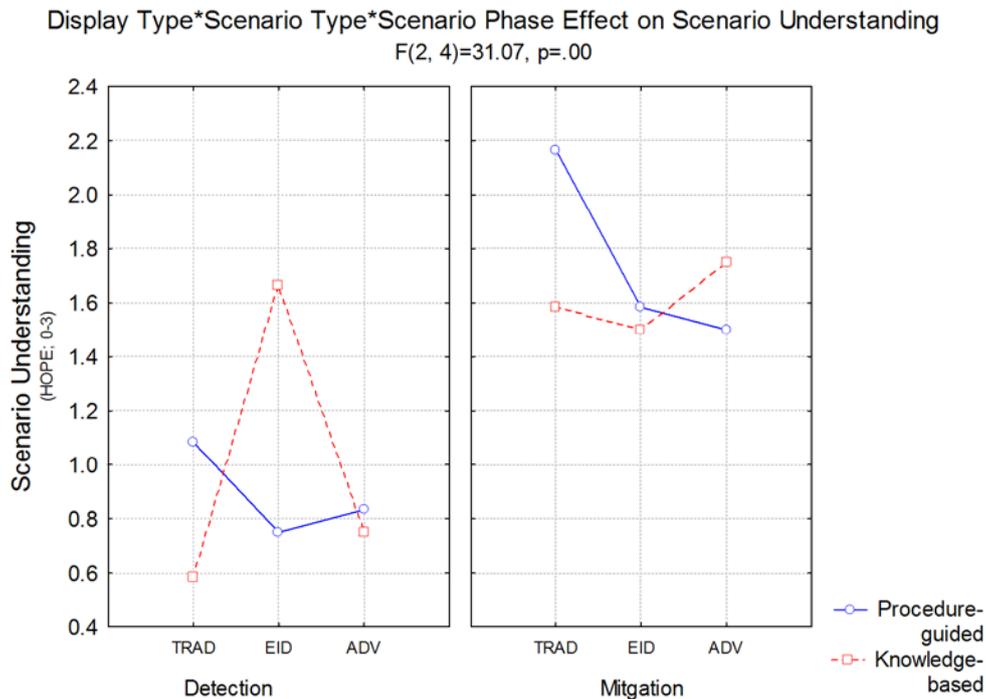


Figure 29: Display type, scenario type and scenario phase interaction effect on Scenario Understanding (HOPE) for Study 1.

Review of the lower order effects indicates redundancy to the three-way interaction effects. Therefore, discussion of the main and two-way interaction effects on Scenario Understanding is omitted. Figure 29 resembles the three-way interaction graphs of Process Overview (Figure 28). In particular, Ecological displays demonstrate notable advantage in the Detection phase of Knowledge-based scenarios. However, the advantage of Advanced displays in the Mitigation phase of Knowledge-based scenarios is only apparent with Process Overview. For Scenario Understanding, Traditional displays appear superior in the Procedure-guided scenarios.

Task performance data (OPAS) were analyzed in an ANCOVA with fixed factors of display types (Traditional, Ecological and Advanced), scenario type (Procedure-guided and Knowledge-based), scenario phase (Detection and Mitigation), a random factor of crew, and a covariate of workload (Subjective Task Complexity Scale) in a RBF design. Unlike all the Process Overview measurements ($ICC(3,1) < .1$), the OPAS measurements demonstrate modest within-subject variance across experimental conditions ($ICC(3,1) = .15$) to conduct analysis in a RBF design (see 7.1.3). The analysis controlled the workload influence on task performance because task performance effects due to workload reduction could conceal the quality of problem solving as a result of the experimental manipulations that were of particular interests to this experiment. [78] documents and discusses the details of the analysis and results pertaining to task performance in Study 1.

Table 10: ANCOVA of Task Performance (OPAS) with Workload (Subjective Task Complexity) as covariate in Study 1.

Fixed Effects	SS	Effect - df	MS	Error - df	MS Error	F	p	η^2
Display	0.99862	2	0.49931	9.47779	0.503506	0.99	0.41	0.17
Scenario	0.82189	1	0.82189	4.93964	0.437281	1.88	0.23	0.28
Phase	0.03000	1	0.03000	14.59678	0.217233	0.14	0.72	0.01
Display*Scenario	3.81322	2	1.90661	9.85857	0.969715	1.97	0.19	0.29
Display*Phase	4.39982	2	2.19991	10.54964	0.272014	8.09	0.01	0.61
Scenario*Phase	1.11333	1	1.11333	5.36694	0.571004	1.95	0.22	0.27
Display*Scenario*Phase	1.63742	2	0.81871	9.00000	0.134599	6.08	0.02	0.57

The ANCOVA on task performance (OPAS scores) with workload (subjective task complexity) as covariate reveals: (a) the interaction effect of display type and scenario phase ($F(2, 10.55)=8.09, p=.01$); and (b) the interaction effect of display type, scenario type and scenario phase ($F(2, 9)=6.08, p=.02$).

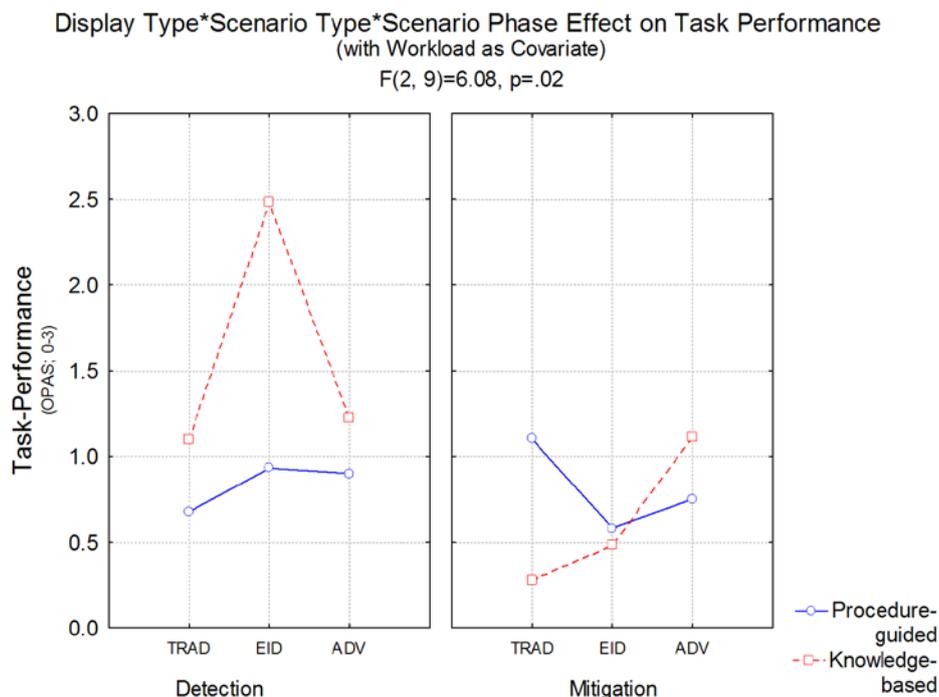


Figure 30: Display type, scenario type and scenario phase interaction effect Task Performance with Workload as a Covariate for Study 1.

Review of the lower order effects indicates redundancy to the three-way interaction effects. Therefore, discussion of the main and two-way interaction effects on task performance is omitted. Figure 30 resembles the three-way interaction graph of Process Overview. In particular, Ecological displays appear superior in the Detection phase of Knowledge-based scenarios. Traditional and Advanced displays demonstrate very slight advantage in the Mitigation phase of Procedure-guided scenarios and Knowledge-based scenarios, respectively.

Debriefing data and expert feedback are typically helpful for developing technology in complex systems and understanding experimental findings. [88] documents some of the debriefing data on Study 1. However, the debriefing data in Study 1 provide limited insight into the experimental findings. Operators generally preferred the Traditional displays and indicated that the Ecological displays were too cluttered

with information. However, some indicated that some graphical elements within the Ecological displays were useful and new visualization (in general) of the plant processes would be welcome.

7.2.4 Discussion

Study 1 provides empirical data highlighting several properties of the Process Overview Measure. From the descriptive statistics, context-sensitive items exhibit some ceiling effects, especially during the Detection phase when the number of process faults occurring is limited. This finding is not surprising, as monitoring context-sensitive parameters should be well established from training and daily work. However, researchers must be aware of the potential for a ceiling effect when employing context-sensitive parameters, especially when operators are not challenged in the scenarios.

The intra-class correlations that indicate the amount of shared variance within subjects across scenario trials are very low (i.e., $r^2 < 0.01$) for context-sensitive, fault-sensitive and overall Process Overview measurements. In other words, the operators did not exhibit much *response tendency* in the repeated testing. Operators were as similar to themselves responding to Process Overview queries in another scenario trial as they were to any other operator. In contrast, other measures demonstrate higher intra-class correlations. The low ICC or response tendency may be an indication that the Process Overview Measure is measuring a phenomenon that is heavily driven by the situational characteristics. Verification of this finding in another full-scope simulator experiments is worthwhile.

Process Overview does not correlate with Scenario Understanding or task performance. This finding is slightly surprising as modest correlations are expected for an indication of validity, although SA studies in the literature do not always observe correlation with other performance indicators (see e.g., [91, 92]). Note that correlations between measures should generally **not** be very strong (i.e., $r^2 > .5$; see e.g., [71]) as that is often an indication that the measures are only superficially different, measuring the same underlying construct (i.e., lack of discriminant properties). Surprisingly, the ANOVA findings for Process Overview, Scenario Understanding, and task performance generally converge. Thus, the lack of correlation may indicate the complexity of the different performance constructs (see e.g., [51]) and a limitation in the range of scenarios in a single study that can produce a correlation effect.

The sensitivity of the Process Overview Measure is tested using ANOVA. The ANOVA for the (overall) Process Overview Measure reveals experimental effects that corroborate the ANOVA effects of Scenario Understanding and task performance. In other words, two measures of operator performance that do not correlate with the Process Overview Measure are indicating similar experimental effects as the Process Overview Measure. The interpretation of the converging results on Process Overview, Scenario Understanding and task performance must take into the consideration of the strong correlation between Scenario Understanding (i.e., HOPE) and task performance (i.e., OPAS) in this study. In other words, the Process Overview results converge with measurements of two related or overlapping measures. Nevertheless, the results together provide sufficient confidence that the ANOVA effects of the Process Overview are a valid indication of sensitivity.

The Process Overview Measure in Study 1 can be further divided into context- and fault-sensitive contents that reveal slightly different experimental effects. Context-sensitive items of the Process Overview Measure reveal a main effect of display type and an interaction effect of scenario phase and scenario period. On the other hand, the fault-sensitive items reveal a complex three-way interaction effect of display type, scenario type and scenario phase. Due to the complexity of the experiment and lack of prior research, precise hypotheses for this differential sensitivity and detailed interpretation of this finding would be overly speculative. However, the differential sensitivity between the two types of items given the results is unlikely to be spurious. For context-sensitive items, the results indicate less complex interaction effects, consistent with the concept that the context-sensitive items should be indicative of the

performance of monitoring practice well-established at work and training in order to build the “background” to support identification of anomalies in a given context or operating states (i.e., monitoring effectiveness in familiar and expected operations). On the other hand, fault-sensitive items yield complex interaction effects, resembling those revealed by Scenario Understanding and Task Performance measurements. This is reassuring because the fault-sensitive items should be indicative of the performance of monitoring behaviours dictated by unplanned/unexpected events in the situation (or scenario) necessary for further investigation (i.e., effectiveness for monitoring for unexpected events).

The debriefing data from the participants/operators did not reveal any opinions on the experimental manipulations that conflict with the statistical findings. The process expert did not express any problems in defining the items during scenario analysis or responding to the queries for during the experiment. In brief, the protocol of the Process Overview Measure is feasible for full-scope simulator studies.

In summary, Study 1 provides evidence that the Process Overview Measure is sensitive to manipulations involving display technology and scenario types in full-scope simulator experiments. The experimental effects revealed by the Process Overview Measure corroborate with findings from other measures, providing confidence in the conclusion about its sensitivity. Study 1 does not suggest any practical concerns about the measurement protocol.

7.2.5 Corollary Issues

The Process Overview Measure demonstrates sensitivity and practicality for assessing monitoring performance in representative experiments through Study 1. However, for generalization, demonstrating these properties in other high fidelity simulator experiments is necessary to ensure that the measure is useful for a large scope of representative experiments.

Study 1 yields empirical data that prompts for verification. Specifically, it is meaningful to confirm: (i) the low intra-class correlation (or variance) within subjects; and (ii) the lack of correlation between Process Overview and Scenario Understanding as well as task performance.

Study 2 is another full scope simulator experiments employing the Process Overview Measure that speaks the corollary issues stemming from Study 1.

7.3 Study 2

Study 2 investigated a diverse set of human factors topics for futuristic operational concepts in a full-scope simulator experiment¹³. Specifically, Study 2 covered the following:

- Staffing strategies for future plants [93];
- Out-of-the-loop performance in highly automated plants [94, 95];
- Multi-layered process overview and large screen displays [96];
- Operational culture on human performance [97]; and
- Situation Awareness in nuclear process control (this volume and [98]).

The full-scope simulator experiment provided a platform to further test the concept and measure of Process Overview that could confirm the empirical results in Study 1 and thereby provide confidence for generalization.

¹³ Note that the full experimental design for the simulator experiment is a 2x2x2x2 within subject design completely crossed. The treatments are: (i) scenario period, (ii) staffing solution, (iii) automation interface, and (iv) overview displays.

7.3.1 Objective

For the purpose of this report, the objective of Study 2 is to verify findings from Study 1 on the Process Overview Measure for generalization. Specifically, the measure should demonstrate:

- Sensitivity to experimental manipulation of human system interfaces,
- Sensitivity to experimental manipulation of scenario difficulties,
- Sensitivity to prior control experience of the operators/participants about the nuclear process, and
- Validity of the experimental effects based on other empirical data in the experiment.

In addition, we continue to investigate qualitatively:

- Value-added/informativeness over other measures of human performance (i.e., discriminant validity)

7.3.2 Method

7.3.2.1 Participants

Nine licensed operator crews (n=9) were recruited from three different Swedish BWR power plants. Plant 1, 2 and 3 provided four, two, and three crews, respectively. Plant 1 and 3 operated with a similar nuclear process to the one of the HAMBO simulator whereas Plant 2 operated with one identical to the simulated nuclear process (also see 7.3.2.5.4). Each crew consisted of one reactor operator (RO), one turbine operator (TO) and one shift supervisor (SS) employed at the same plant. Note that some participants within each crew worked together on a daily basis while others did not. ([93, 94] present the demographic details of all the participants).

7.3.2.2 Test and simulation environment

The experiment was performed in HAMMLAB on the HAMBO simulator [99].

The control room set-up in HAMMLAB included:

- A process workstation with eleven 30-inch LCD screens for operators in roles that involved direct interaction with automation and intervention with the nuclear process (see section 7.3.2.5.1).
- A management workstation with four 30-inch LCD screens for operators in a managing role that involved monitoring the overall process development and performing administrative tasks (see 7.3.2.5.1).
- A large screen display to provide an overview of the process state.

Figure 31 presents an example of the control room set-up for this study. The exact human-system interface configuration and control room layout varied with experimental conditions (see 7.3). As mentioned, this report excludes the overview display manipulation. The experimental design adopted for analysis aggregates the data of the overview display manipulation. The variations caused by these factors were thereby counterbalanced per test condition, and should not be confounded with the experimental manipulations of interest.

For the process workstation, two LCD screens were available to each operator assigned to the station for accessing mimic-based displays. Operators could select freely from the full set of 64 mimic-based displays, which were designed for detailed monitoring, fault diagnosis, and human intervention. These displays also featured parameter trending. Alarm information was presented on a separate LCD screen that was shared among the operators assigned to the process workstation. Selected alarms were also included elsewhere in the human-system interface, such as in a small window below the mimic-displays

and integrated on the large screen overview display. The remaining LCD screens were assigned according to the experimental conditions described in 7.3.2.5.1. Screens that were not needed in a given test condition were turned off.

For the management workstation, one LCD screen was dedicated to mimic-based displays while another was dedicated to alarm information. The remaining two LCD screens were either assigned or turned off according to experimental conditions described in 7.3.2.5.1.

Navigation within displays was carried out via keyboard and mouse. The operating procedures used in the experiment were standard paper-based procedures developed for the simulated plant. All participating operators had a personal telephone for communicating with plant maintenance and management.



Figure 31: Example of control room set-up (traditional operator roles and transparent automation displays)

7.3.2.3 Plant Automation

Plant automation was implemented as computer scripts that could monitor basic process deviations in normal plant states, and execute start-up or shutdown procedures of the simulated plant. For this study, automation was intentionally obstructed by process faults throughout the scenarios, and would “pause” when technical problems hindered the completion of current tasks or threatened the availability of higher-level functions (see [94, 95] for details). Because mitigation of system failures was beyond the capability of plant automaton, the operators must intervene with the following interaction options with plant automation: (a) resolve the system failures first and then restart automation, (b) restart automation first and then deal with system failures while automation continued to run up the plant, (c) restart automation and ignore the system failures (which might be appropriate in some situations), (d) continue to run up the plant manually and resolve the system failures in parallel, or (e) shutdown the plant for safety reasons.

7.3.2.4 Experiment staff

An Experiment Leader (EL), two Process Experts (PEs) and a Laboratory Technician (LT) managed the simulator, scenarios and data collection from the observation gallery (Figure 32).

The two process experts¹⁴ were responsible for operating the simulator and plant automation. They also implemented event sequences for the scenarios. One process expert acted as field operator, electrician, plant management and other roles as necessary by responding to phone calls from the control room. This process expert (also Rater A in Study 3, see 7.4.2.1) registered reference answers to the Process Overview queries and rated participants on other performance scales. The second process expert (also Rater C in Study 3, see 7.4.2.1) was responsible for managing technical operations of the simulator associated with the experiment.

The experiment leader had overall responsibility for the data collection. During the scenario runs, the experiment leader coordinated the gallery staff and was the formal point of contact between the participants and the research team.

The laboratory technician managed the digital audio and video recordings, operated the computerised questionnaire system, and checked the simulator logging system.

For some of the experimental runs, a third process expert was present in the observation gallery to conduct an inter-rater reliability study (see 7.4). The third expert was given a workstation between the EL and LT with access to the same observational data as the other expert rater (Figure 54).

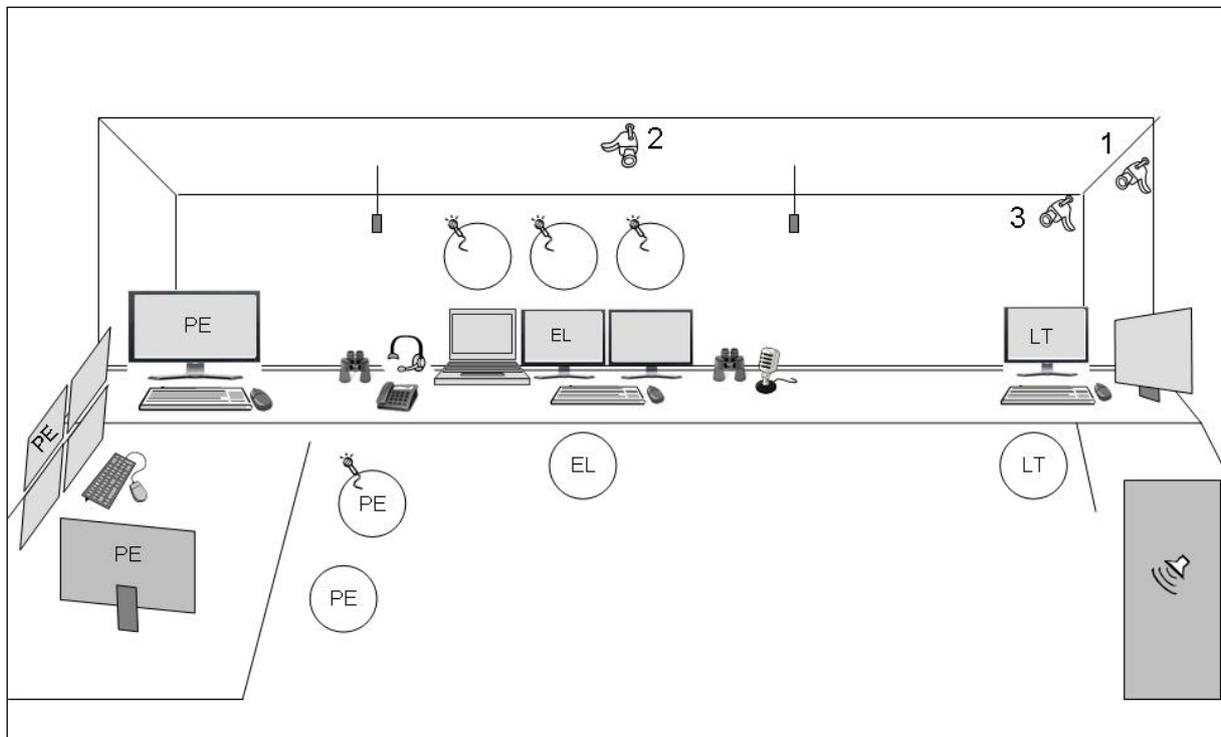


Figure 32: Observation gallery set-up for Study 2.

7.3.2.5 Experimental Manipulations

Study 2 consisted of three experimental manipulations – staffing solutions, automation interface, and scenario periods. Further, one analysis includes a predictor/factor of plant that are not part of experimental manipulations.

7.3.2.5.1 Staffing Solution

¹⁴ The two process experts were also participants of Study 3 as Raters A and C.

Two types of staffing solution – Traditional and Untraditional - were included to explore new staffing complements facilitated by high levels of plant automation. The two staffing solutions differed in the responsibility of the crew, responsibilities of individual operators, and control room set-up. For the purpose of this report, the two staffing solutions are briefly described. [93] documents the details of this experimental manipulation.

The Traditional staffing solution consisted of a Reactor Operator (RO), Turbine Operator (TO) and Shift Supervisor (SS) according to the conventional composition of an operating crew in most NPPs today. The Traditional staff solution was responsible for operating one highly automated nuclear process. The RO and TO handled the reactor and the turbine sides of the nuclear process, respectively, while the SS managed and supported the RO and TO. This study was able to assign participants to roles according to their work positions at their home plant.

The control room set-up for the Traditional staffing solution also resembled the layout of conventional control rooms in many NPPs today (see Figure 33). The RO and TO shared the process workstation. Additional displays were available to both operators in the Transparent automation interface and Future overview display conditions. The SS worked at the managing workstation. One additional display was available to the SS in the Transparent automaton interface conditions.

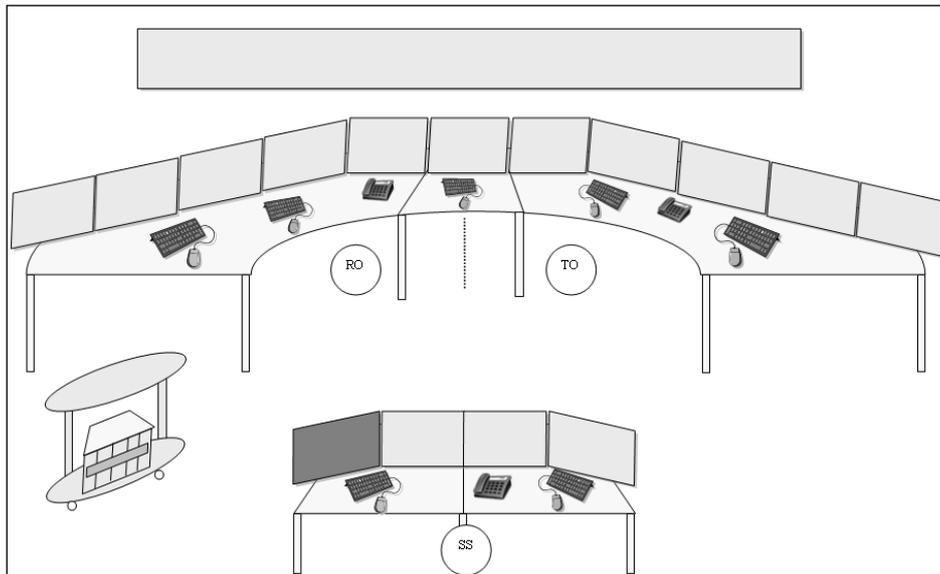


Figure 33: Control room set-up in the traditional staffing solution

The Untraditional staffing solution consisted of a Main Operator (MO), Assistant Operator (AO) and Work Manager (WM) formulated to explore staffing considerations in a hypothetical future NPP. The Untraditional staff solution was responsible for operating two highly automated NPPs in two separate control rooms. The MO was responsible for both the reactor and turbine side of simulator Plant A in HAMMLAB while the AO was responsible for the turbine side of simulator Plant B in an adjacent room. The WM was responsible for both simulator Plant A and B, managing and supporting both MO and AO. The WM was also responsible for re-allocating the AO from Plant B to Plant A to support MO under critical operating conditions.

Figure 34 illustrates the control room layout for Plant A and Plant B in the Untraditional staffing solutions. In Plant A (i.e., HAMMLAB), all LCD screens in the process workstation were assigned to the MO for operating both the reactor and turbine side except for two LCD screens presenting mimic-based displays that were only turned on for the AO when he was re-allocated from Plant B (to support the MO). Additional displays were available **only to MO** in the Transparent automation interface and Future

overview display conditions. The WM worked at the managing workstation that faced the process workstation to facilitate communication. One LCD screen of the managing workstation was dedicated to administrative tasks. One additional display was available to the WM in the Transparent automation interface conditions (as in the Traditional staffing solution). In Plant B, the AO had a workstation dedicated to the turbine side.

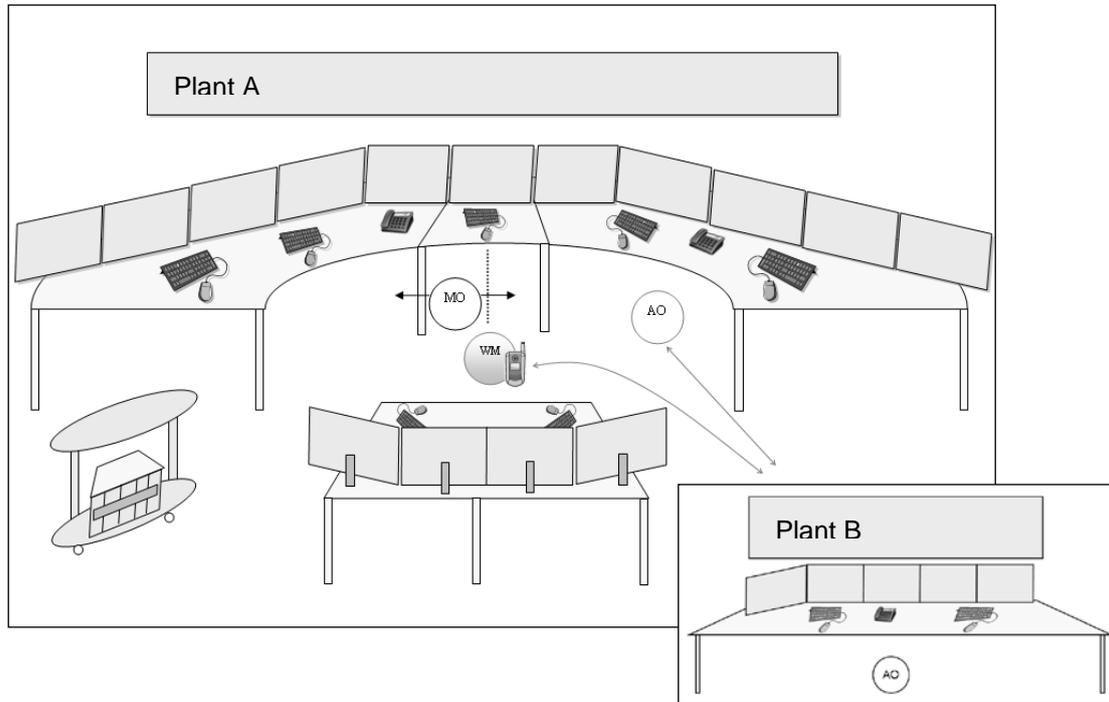


Figure 34: Control room set-up for Plant A and Plant B (smaller picture) in the untraditional staffing solution

7.3.2.5.2 Automation Interface

Two types of automation interface – Non-transparent and Transparent - were included to investigate the effects of displaying automation information in supporting operators to cope with high levels of plant automation. For the purpose of this report, the two automation interfaces are briefly described. [94, 95] document the details of this experimental manipulation.

The Non-transparent and Transparent automation interfaces shared the same interactive functions but differed in the amount of information presented about the plant automation. The Non-transparent automation interface presented minimal information about plant automation containing only (i) the interaction buttons and (ii) the unsatisfied condition when plant automation stopped.

The area enclosed by red rectangle of Figure 35 was the Non-transparent automation interface. Therefore, operators had to make inferences about automation activity from observing process events and changes in process parameters based on the expected sequence of actions as prescribed by the operating procedures.



Figure 35: Interaction with automation in the non-transparent automation interface (red rectangle)

The Transparent automation interface consisted of a plant automation overview and detailed displays presented on two 30-inch LCD screens, respectively (Figure 36). The overview display showed the goals and effects of process control tasks and the interaction buttons to control the plant automation. The detailed automation display showed historical and ongoing automation activities through excerpts from mimic-based displays and a chronological log.

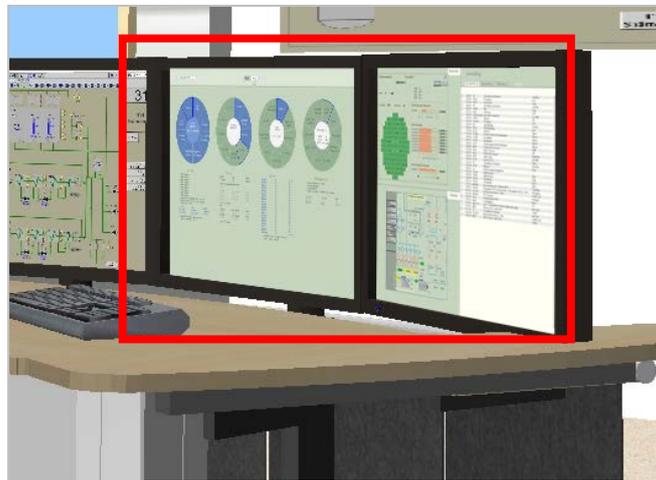


Figure 36: Transparent automation interface

7.3.2.5.3 Scenario Periods

Each scenario started with the Easy followed by the Difficult scenario period. The Easy scenario period contained process faults that were independent of each other. This scenario period was assumed to have low time pressure and reliable automation (i.e., operating within the design basis of the plant automation).

The Difficult scenario period contained more complex faults that could be either dependent or independent of each other. The causes of the faults were difficult to understand. The time pressure was typically high, particularly if unfinished tasks from the Easy period had to be completed in parallel with handling the complex faults introduced in this period. The plant automation could be working on the border of its design basis during some phases of the difficult period. Figure 37 illustrates the structure of the scenario.

7.3.2.5.4 Plant

The participating operators were recruited from three different operating BWR plants – Plant 1, 2 and 3. Plant 1 and 3 operate with a nuclear process that is similar to that of HAMBO simulator, whereas, Plant 2 operates with a process that is identical to that of the simulator. Therefore, Plant 2 operators were expected to have more control experience on the HAMBO simulator process than Plant 1 and 3 operators.

7.3.2.6 Experimental Design

To assess sensitivity of the Process Overview Measure to control experience of the participants to the simulated process, a one-way ANOVA was employed with the treatment of plant (1, 2, and 3). Because the plant factor was mainly an outcome of the recruiting process rather than deliberate experimental manipulations, the plant factor was examined with its own experimental design¹⁵.

To assess sensitivity of the Process Overview Measure to experimental manipulations, a 2x2x2 RBF (within-subjects) design was employed with treatments of staffing solution (Traditional and Untraditional), automation interface (Non-transparent and Transparent), and scenario period (Easy and Difficult). The treatments were completely crossed.

7.3.2.7 Measures

7.3.2.7.1 Process Overview

Process Overview is operationalized with the Process Overview Measure, of which the general structure is fully described in Chapter 6. This subsection highlights only the details of the implementation specific to Study 2.

In Study 2, the Process Overview Measure only included context-sensitive parameters, omitting fault-sensitive queries. There were three reasons for this choice of implementation. First, the Untraditional staffing solution required the Main Operator (MO) to assume the responsibilities of both reactor and turbine sides. As a result, the MO responded to Process Overview queries for both reactor and turbine sides. Given administration of other questionnaires during simulator freezes, including both context- and fault-sensitive queries could be too intrusive on the participants assuming the MO role.

Second, the scenarios contained several faults intended to produce frequent interaction opportunities between operators and plant automation. The faults were also introduced in increasing order of difficulty in anticipation of varying ability of participants to cope with the unfamiliarity with the high level of automation. That is, participants performing well could proceed further in the scenarios than those performing poorly. The scenarios contained several faults in the Easy period prohibiting the effective use of alarms for nullifying cuing of fault-sensitive queries. For the Difficult period, it was difficult to predict the progress into the scenarios achieved by the participants a priori given the exploratory nature of this study. Thus, fault-sensitive process parameters were difficult to select. While it was possible to emphasize faults in the beginning of the Difficult period, this would create both cuing and memory effects on operator responses. For these reasons, the implementation of Process Overview for this study omitted fault-sensitive queries.

¹⁵ All the treatments – plant, staffing solution, automation display, and scenario period – can be incorporated into a single experimental design for analysis. However, communication of the results would become unnecessarily complicated without providing additional information about the Process Overview Measure.

Third, the threat of ceiling effects of relying on context-sensitive parameters was reduced as many of the operators (i.e., participants) were not working at the plant with a nuclear process identical to the simulator. Furthermore, the experimental setup of the control room was intentionally designed to be different, as opposed to being similar, compared to conventional control rooms for exploring future operational concept.

The deliberate omission to fault-sensitive queries impacts content validity of the Process Overview Measure in Study 2. Thus, the interpretation of the Process Overview results must take this implementation choice into account.

7.3.2.7.2 Scenario Understanding

Scenario Understanding (see [13] for a preliminary discussion) refers to the operator understanding to the process faults introduced in the scenarios, essentially reflecting the quality of diagnosis. Study 2 operationalized Scenario Understanding with the Automation and Scenario Understanding Rating Scales (ASURS; [100]).

ASURS employs process experts to rate operators understanding of automation and scenario on four items. The first item assesses the operator awareness of automation purpose for the scenario on a four point anchored scale. The second and third items assess operator awareness of process and performance of automation on a five point anchored scale. The final item assesses operator understanding of the situation on a five point anchored scale. During data collection, process experts perform their ratings based on observation (e.g., conversation between operating crew members) from the experimental gallery.

For Study 2, a process expert conducted the ratings of the four items for each scenario phase. The final Scenario Understanding index is the sum of all the ratings.

7.3.2.7.3 Task Performance

Task performance refers to the overall effectiveness of control room operations demonstrated by the operators. Study 2 operationalized task performance with the Operator Performance Assessment System (OPAS) [13, 52, 89].

OPAS provides a structure for the assessment of whether operators carry out their task work in accordance with scenario solutions prescribed a priori by experts in control room operation. Prior to data collection, process experts analyzed the scenarios and developed optimal solutions by identifying items that expressed the desired performance. In Study 2, the OPAS items were separated into Detection and Operation to isolate specific aspects of task performance. By comparison to Study 1, the OPAS items in this study were more concrete with less reference to cognitive activities such as problem solving activities and strategies. During the experiment, a process expert registered the points earned by operators in completing the predefined activities within each performance item based on observations from the gallery. The employed performance index is the sum of points earned by the participants divided by the maximum available points in each scenario with a scale ranging from 0-1. The OPAS index reflects the degree of conformance between operator performance and predefined optimal solutions to scenarios.

7.3.2.8 Procedure

The participation of each crew was divided over four days. The first day was dedicated to the training program after obtaining informed consent and demographic information as well as an introduction to the study. The morning of the second day was also devoted to training. The training covered the following:

- Differences between the operators' home plant and the HAMBO simulator.
- Familiarisation with the human-machine interface (navigation, manoeuvring of components, display characteristics, alarm system, trends, and overview displays)
- Visualisation of automation activity and interactions with the automatic system.
- Overview of the test conditions in the experiment, including human-system interfaces, operator roles and work practices.
- Data collection methods and laboratory procedures.

The training program concluded with a two-part training scenario to expose operators to a realistic trial and to verify the competence level on operating the simulator. In the first part, operators were arranged into the Traditional staffing solution to operate a single nuclear process with various process faults. The first part ended with a simulator freeze. During the freeze the operators were given instruction on how to respond to a set of questionnaires. In the second part, the operators were re-arranged into Untraditional staffing solution to operate two nuclear processes with various process faults.

The data collection began after lunch on the second day until lunch hour on the fourth day. Each crew participated in eight scenarios related to starting up the highly automated simulator plant. At the beginning of each scenario, a process expert briefly informed participant the general context of the scenario (i.e., the stages of the plant start-up). [93, 94] provide detailed scenario descriptions. Each trial began with the Easy scenario period of about fifteen minutes followed by a simulator freeze to collect responses to the Process Overview queries and questionnaires. After the simulator freeze, the trial entered into the Difficult scenario period of another fifteen to twenty minutes. At the end of the Difficult scenario period, operators were administered another set of Process Overview queries and questionnaires. Figure 37 illustrates the general structure of the scenario.

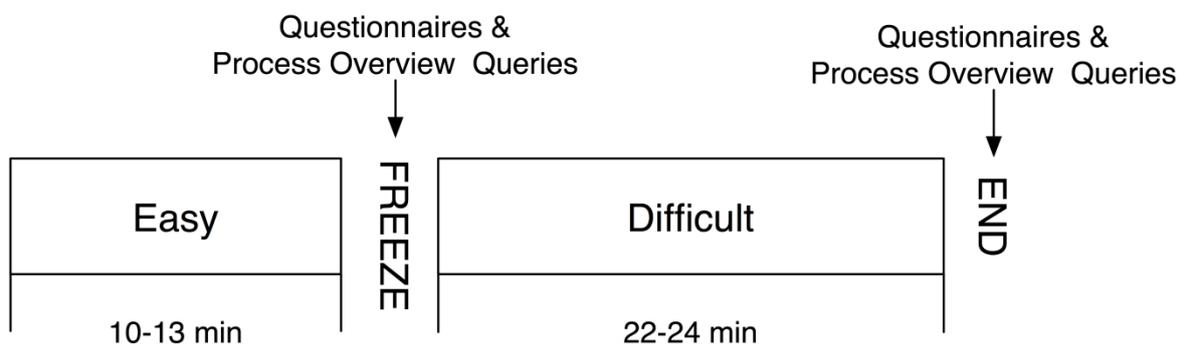


Figure 37: Basic structure of scenarios in Study 2.

To minimize overload and carry-over effects, a twenty-minute break separated every scenario. The data collection concluded with a debriefing interview after lunch on the fourth day to obtain feedback on the experiment, especially the experimental manipulations. The debriefing were about one and half hour long.

7.3.2.9 Evaluation criteria (Hypothesis)

In Study 2, the Process Overview Measure is expected to re-demonstrate the correlation, sensitivity, feasibility characteristics discovered from Study 1.

Sensitivity is demonstrated by significance testing that corroborates with other empirical data in the study and general indications from the literature. Specifically the Process Overview Measure should reveal¹⁶:

¹⁶ The literature provides insufficient materials to meaningfully postulate the effects of the staffing manipulation on Process Overview (see for [81] a discussion). The experiment is exploratory, intentionally designed to study new ways of working for process operators.

- effects of plant where operators work (as one of the nuclear power plants from which operators were recruited has an identical process to the simulator, resulting in relative advantage in terms of simulator knowledge),
- effects of the human system interface manipulation (as the literature suggests that presenting of automation behaviours could support monitoring),
- effects that are consistent with the Scenario Understanding, task performance, and debriefing data, and
- lack of correlation with Scenario Understanding and task performance (as confirmation of Study 1 finding).

Feasibility is demonstrated by qualitative feedback from the process experts and participants in Study 2.

7.3.3 Analysis and Results

The analysis and results of Study 2 are documented in four parts: (i) descriptive statistics to qualitatively inspect distributions of the Process Overview measurements and search for threats to robustness of ANOVA; (ii) correlations with other measures to identify convergent and discriminant relationships; (iii) ANOVA results to assess sensitivity of the measure; and (iv) findings from other measures to confirm the ANOVA results.

7.3.3.1 Descriptive Statistics

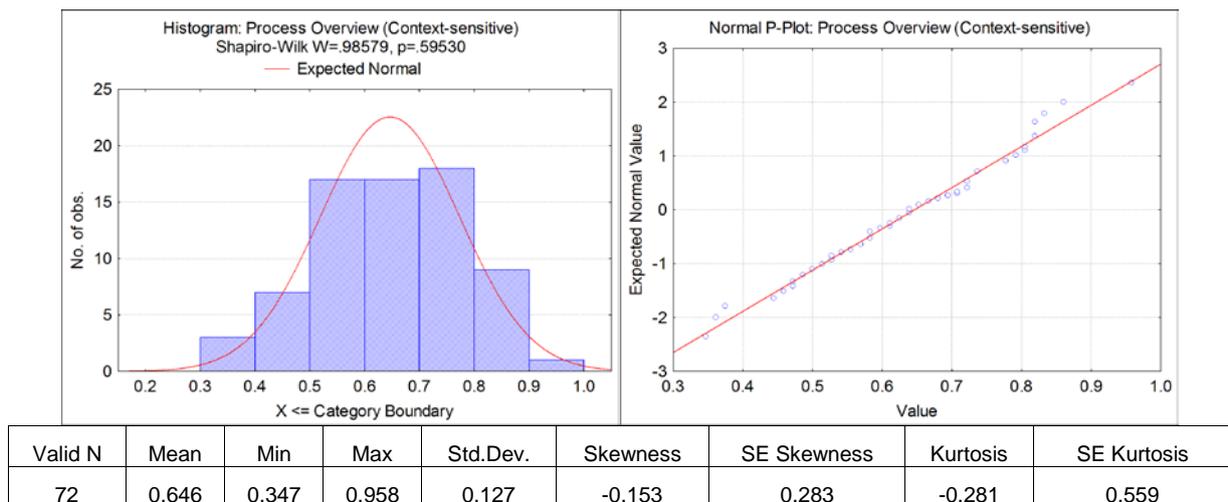
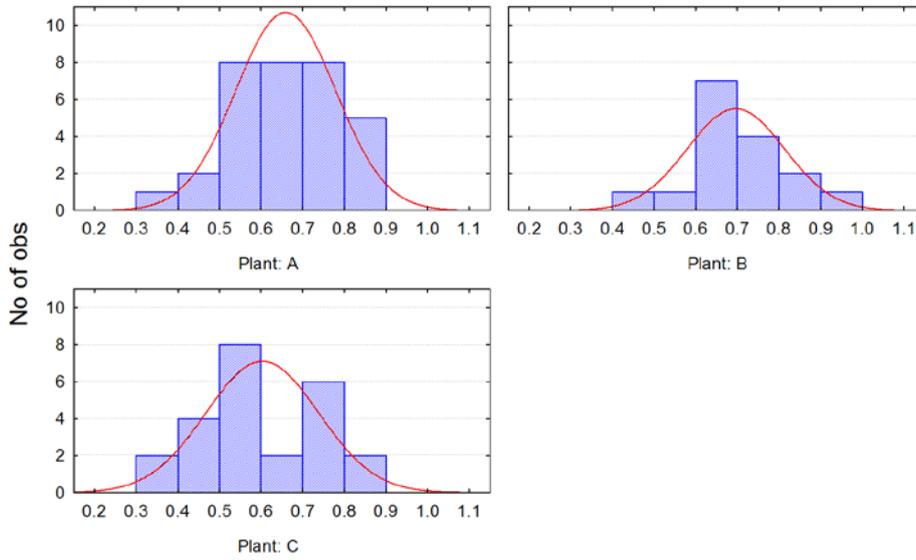


Figure 38: Descriptive statistics of Process Overview measurements (Context-sensitive only) in Study 2.

The Process Overview measurements (context-sensitive content only; Figure 38) is normally distributed ($W=.96$, $p=.59$). The distribution does not indicate any ceiling or floor effect of the measure. The intra-class correlation for Process Overview measurements amongst the participants is $ICC(3,1)=.07$, indicating limited within-subject effects across experimental conditions and suggesting statistical analysis based on CRF design as opposed to a RBF design (see section 7.1)

Histograms by Plant: Process Overview (context sensitive only)

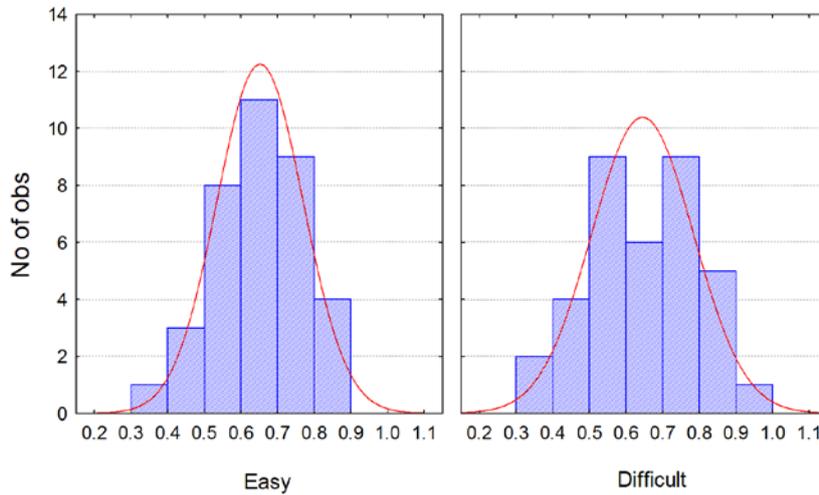


Process Overview (Proportion Correct)

Figure 39: Histograms by Plant for Process Overview (context-sensitive only) in Study 2.

Figure 39 presents the distributions of the Process Overview data collected from operators of three different nuclear power plants. The histograms appear similar (i.e., skewness and kurtosis in the same direction) and thus do not indicate any threats to the robustness of one-way ANOVA for comparing Process Overview of operators from different plants.

Histograms by Period: Process Overview (Context-sensitive)



Context-sensitive Process Overview (Proportion Correct)

Figure 40: Histograms by Period for Process Overview (context-sensitive only) in Study 2.

Histograms by Staff: Process Overview (Context-sensitive)

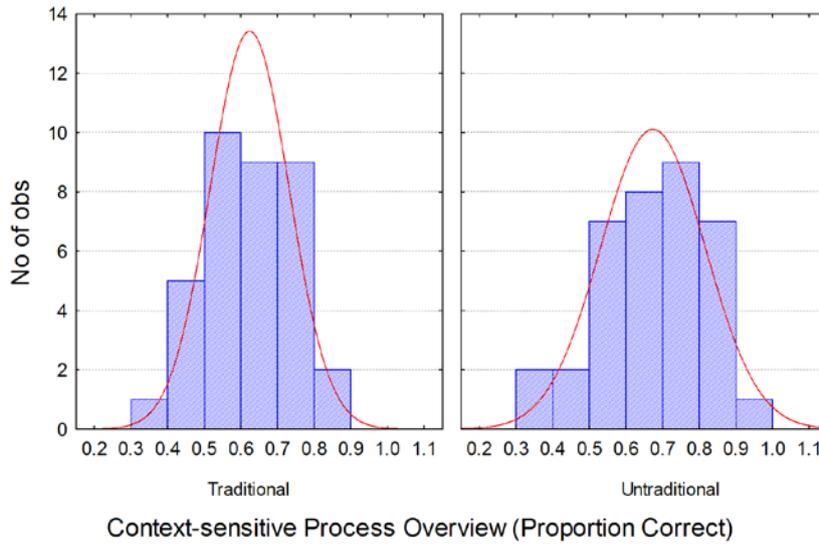


Figure 41: Histograms by Staff for Process Overview (context-sensitive only) in Study 2.

Histograms by Automation Display: Process Overview (Context-sensitive)

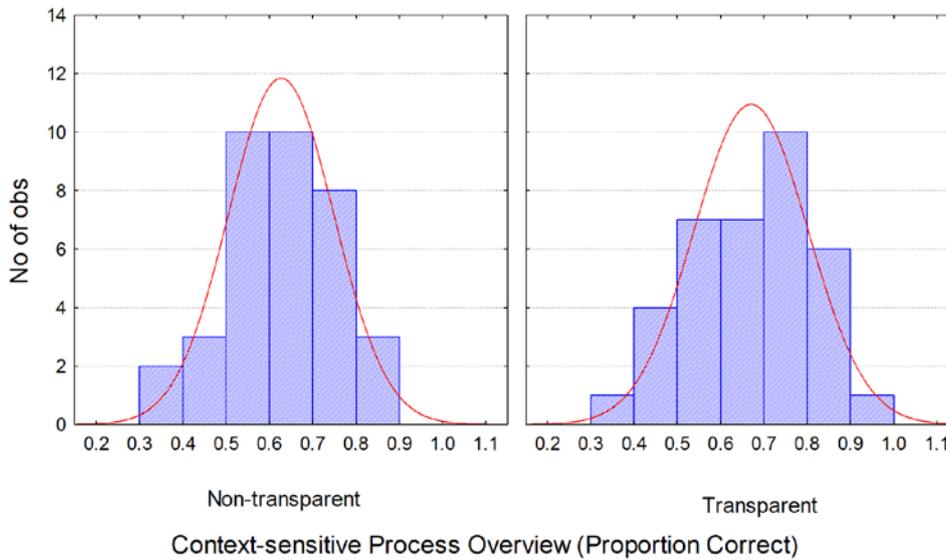


Figure 42: Histograms by Automation Display for Process Overview (Context-sensitive only) in Study 2.

Figure 40-Figure 45 present the distributions of the Process Overview data by individual treatments. The histograms within the treatments appear normally distributed and should not pose any threats to the robustness of ANOVA results.

7.3.3.2 Correlations

In contrast to Study 1, which focuses on the performance of the turbine operators, Study 2 investigates the performance of the individual roles and the team with different measures. Thus, the correlation statistics¹⁷ between measures are difficult to interpret.

Table 11: Correlation between Process Overview and other performance measure in Study 2.

	<i>Scenario Understanding (ASURS)</i>	<i>Task Performance – Detection (OPAS)</i>	<i>Task Performance – Operation (OPAS)</i>
Context-sensitive Process Overview	r(80)=.1377 p=.237	(r144)=.0669 p=.426	r(144) =.1789 p=.032

In Study 2, Process Overview does not correlate with Scenario Understanding as operationalized by the ASURS ($r(80)=.14$, $p=.24$). Note that the aggregation of Process Overview excludes data collected from the role of Shift-supervisor and Work Manager in the Traditional and Untraditional staffing solution, respectively, because ASURS is intended to measure the performance of the operators only. Further, ASURS data of the first four crews (all from Plant 1) is omitted from the analysis because of re-training of the process experts (i.e., raters) on the measurement protocol.

Process Overview does not correlate with detection task performance as operationalized by OPAS-Detection ($r(144)=.07$, $p=.43$), but does correlate slightly with OPAS-Operation ($r(144)=.18$, $p=.03$). Note that the aggregation of Process Overview includes all roles in both staffing solutions.

Process Overview generally showed negligible correlations with Scenario Understanding and task performance (also see results from Study 1 in section 7.2.3.2). The slight positive correlation between Process Overview and OPAS-Operation occurred in the expected direction. That is, Process Overview improved with operation task performance. However, this finding appears to be irrelevant because of the low shared variance ($r^2=.03$).

7.3.3.3 Sensitivity (ANOVA effects)

To test the influence of control experience on the HAMBO simulator process, the Process Overview data was analyzed in a one-way ANOVA using Type III/Unique sums of squares with a fixed, between-subjects factor of Plant (1, 2 and 3) in a CRF design. (For analysis in CRF design, confidence intervals are presented graphically. See section 7.1.7)

Table 12: ANOVA of Process Overview (context-sensitive items) on plant factor in Study 2.

<i>Fixed Effects</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
Plant	0.08913	2	0.04456	2.895	0.06	0.08
Error	1.06210	69	0.01539			

The one-way ANOVA indicates a difference in Process Overview between plants where operators work ($F(2,69)=2.90$, $p=.06$, $\eta^2=.08$). Levene's test indicates homogeneity of variance for plant ($F(2,69)=0.91$, $p=.41$).

¹⁷Correlations are calculated according to original data collection plan of 2x2x2x2 randomized block design with fixed factors of with treatments of staffing solution (Traditional and Untraditional), automation interface (Non-transparent and Transparent), overview interface (Future and Present), and scenario period (Easy and Difficult)

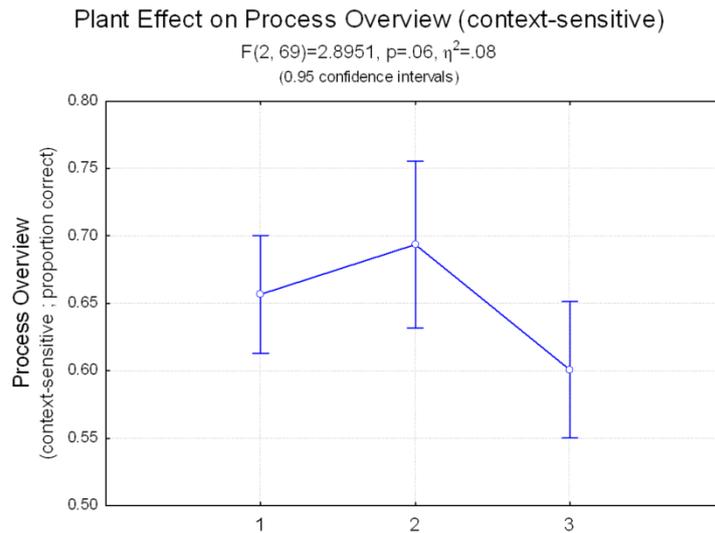


Figure 43: Plant effect on Process Overview (context-sensitive items) in Study 2.

Figure 43 illustrates that the crews from Plant 2 exhibit the best Process Overview. This finding is consistent with expectations because crews from Plant 2 have more control experience on the HAMBO simulated nuclear process than crews from Plant 1 and 3.

To assess the experimental manipulations, the Process Overview data was analyzed in an ANOVA using Type III/Unique sums of squares with fixed factors of scenario period (Easy and Difficult), staffing solution (Traditional and Untraditional), automation display (Non-transparent and Transparent) in a CRF design. Note that a CRF design was selected because of the low intra-class correlation within subjects as documented in section 7.3.3.1. (Refer to 7.1.3 for a general discussion.) In addition, confidence intervals for results analysed with a CRF design are presented. (Refer to 7.1.7 for a general discussion.)

Table 13: ANOVA of Process Overview (Context-sensitive items) for Study 2.

Fixed Effects	SS	df	MS	F	p	η^2
Period	0.00087	1	0.00087	0.056	0.81	0.00
Staff	0.04390	1	0.04390	2.840	0.10	0.04
Automation Display	0.03361	1	0.03361	2.174	0.15	0.03
Period*Staff	0.01239	1	0.01239	0.801	0.37	0.01
Period*Automation Display	0.01097	1	0.01097	0.710	0.40	0.01
Staff*Automation Display	0.05711	1	0.05711	3.695	0.06	0.05
Period*Staff*Automation Display	0.00310	1	0.00310	0.200	0.66	0.00
Error	0.98928	64	0.01546			

The ANOVA on the Process Overview data reveals: a) the main effect of staffing solution ($F(1,64)=2.84, p=.10, \eta^2=.04$); and b) the interaction effect of staffing solution and automation display ($F(1,64)=3.70, p=.05, \eta^2=.05$). Levene's tests indicates homogeneity of variance for the two-way interaction of staffing solution and automation display ($F(2,69)=0.91, p=.44$) but slight heterogeneity of variance for staffing solution ($F(1,70)=2.80, p=.10$). In addition, there appears to be a slight correlation between means and standard deviation for the effect of staffing solution.

A non-parametric, Mann-Whitney U test on staffing solution ($U(36)=498.5$, $z=1.68$, $p=.09$) provides some statistical confirmation¹⁸ on the ANOVA staffing solution main effect (indicating the robustness of ANOVA to deviations in homogeneity of variance).

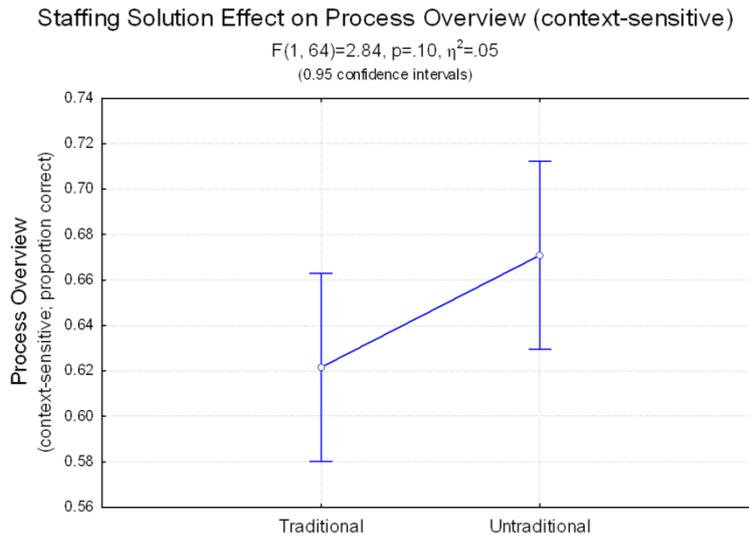


Figure 44: Staffing solution main effect on Process Overview (context sensitive items) in Study 2.

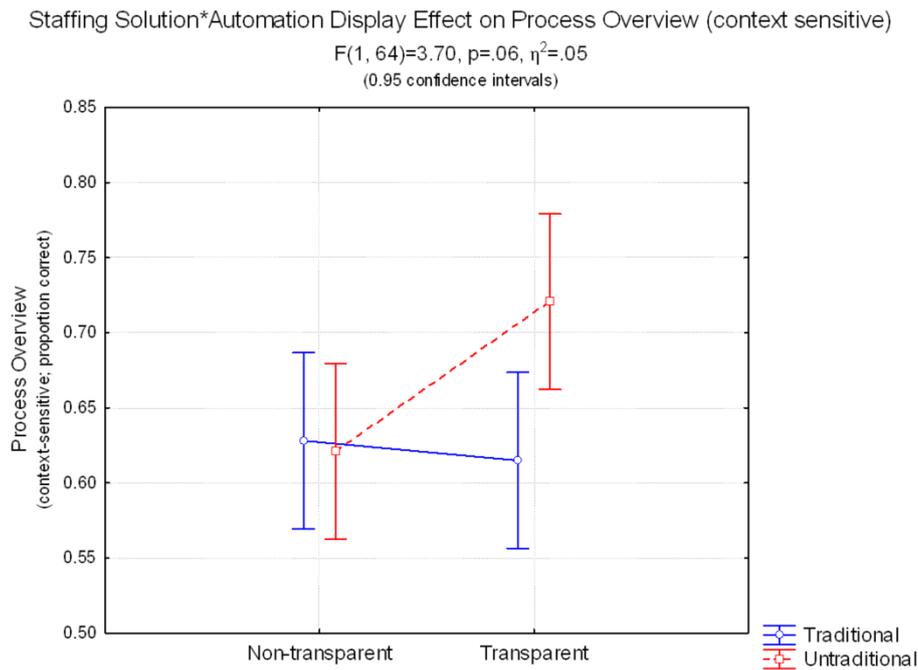


Figure 45: Staffing solution and automation display interaction effect on Process Overview (context sensitive items) in Study 2.

Figure 45 illustrates that Transparent automation displays produced a positive effect on the Process Overview for Untraditional staffing solution but a negligible effect for the Traditional staffing solution (which translates to a main effect of staffing solution as depicted by Figure 44).

¹⁸ The Mann-Whitney U test yielded a p-value of .09, which was too high to constitute full confirmation.

7.3.3.4 Other Empirical Evidence

Study 2 also provides empirical results on Scenario Understanding and task performance - as operationalized by ASURS and OPAS, respectively - for comparison to the Process Overview findings. (As mentioned, ASURS measurements from Plant 1 were omitted from all data analysis due to re-training of process expert on the measurement method.)

ASURS measurements correlate moderately with both OPAS-Detection ($r(80)=.47, p=.00$) and OPAS-Operation ($r(80)=.47, p=.00$). Note that these correlations must be interpreted with care because of the difference in the aggregation of the ASURS and OPAS scores. As mentioned, the aggregation of ASURS scores excluded data collected from the roles of Shift-supervisor and Work Manager in the Traditional and Untraditional staffing solution, respectively. On the other hand, the aggregation of both OPAS scores included data collected from the roles of Shift-supervisor and Work Manager. The moderate correlation suggests some overlap between the ASURS and OPAS measurements but not substantial enough to warrant concerns on discriminant properties between the measures or constructs.

OPAS-Detection and OPAS-Operation measurements demonstrate a modest to strong correlation ($r(144)=.62, p=.00$). This level of correlation is expected because operators perform control actions after detection of process faults. Further, the two scores follow the same measurement method.

To test the influence of process knowledge about the simulator based on the nuclear power plants where operators work, Scenario Understanding (ASURS) was analyzed in a one-way ANOVA using Type III/Unique sums of squares with fixed, between factor of Plant (2 and 3).

Table 14: ANOVA of Scenario Understanding (ASURS) on plant factor in Study 2.

Fixed Effects	SS	df	MS	F	p	η^2
Plant	17.672	1	17.672	3.925	0.05	0.09
Error	171.106	38	4.503			

The one-way ANOVA indicates a difference in Scenario Understanding between plants where operators work ($F(1,38)=2.76, p=.05, \eta^2=.09$).

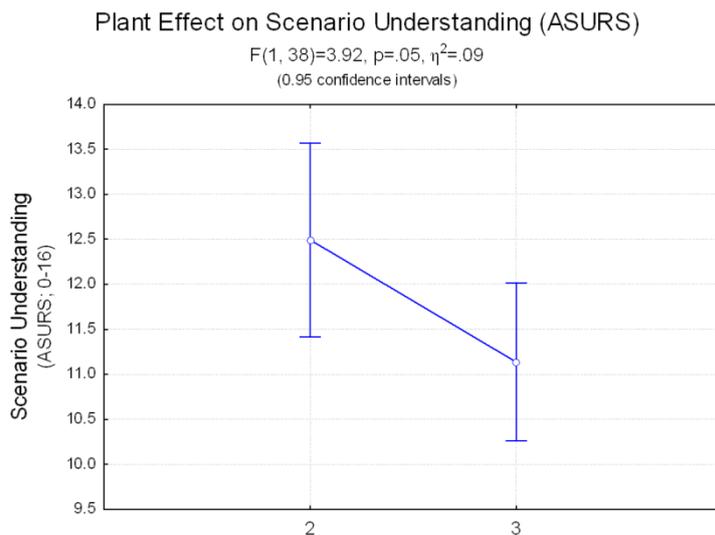


Figure 46: Plant effect on Scenario Understanding (ASURS) in Study 2.

Figure 46 illustrates that the crews from Plant 2 exhibit better Scenario Understanding than Plant 3. Due to omission of data from Plant A, this finding only affords an indirect comparison to the result of Process Overview. This finding is consistent with Process Overview and expectation because Plant 2 operates with the identical nuclear process as the HAMBO simulator, whereas Plants 3 operates with a merely similar nuclear process.

ASURS does not reveal any experimental effects relevant for this report. [100] documents the details and analysis of ASURS for Study 2.

To test the influence of process knowledge about the simulator based on the nuclear power plants where operators work, the detection task performance data (OPAS-Detection) was analyzed in a one-way ANOVA using Type III/Unique sums of squares with a fixed, between factor of Plant (1, 2 and 3) in a CRF design.

Table 15: ANOVA of detection task performance (OPAS-Detection) on plant factor in Study 2.

<i>Fixed Effects</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
Plant	0.24155	2	0.12078	2.7588	0.07	0.07
Error	3.02077	69	0.04378			

The one-way ANOVA indicates a difference in detection task performance between plants where operators work ($F(2,69)=2.76, p=.07, \eta^2=.07$).

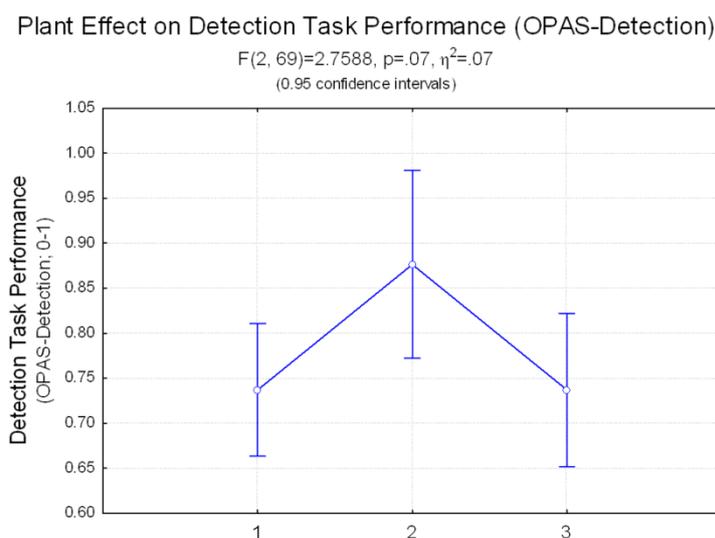


Figure 47: Plant effect on detection task performance (OPAS-Detection) in Study 2.

Figure 47 illustrates that the crews from Plant 2 exhibit the best detection task performance. This finding converges with the effect revealed by Process Overview. As mentioned, this finding is consistent with expectations because Plant 2 operators have more control experience of the simulated process than both Plant 1 and 3 operators.

To assess the experimental manipulations, the detection task performance data (OPAS-Detection) was analyzed in an ANOVA using Type III/Unique sums of squares with fixed factors of scenario period (Easy and Difficult), staffing solution (Traditional and Untraditional), automation display (Non-transparent and Transparent) in a RBF design. (Unlike Process Overview, the OPAS-Detection measurements demonstrate fair within-subject variance across experimental conditions ($ICC(3,1)=.17$) to conduct analysis in a RBF design.)

Table 16: ANOVA of Detection Task Performance (OPAS-Detection) for Study 2.

Fixed Effects	SS	Df	MS	SS-Err	Df-Err	MS-Err	F	p	η^2
Period	0.6019	1	0.6019	0.1600	8	0.0200	30.0960	0.00	0.79
Staff	0.5787	1	0.5787	0.2674	8	0.0334	17.3097	0.00	0.68
Auto	0.0969	1	0.0969	0.1753	8	0.0219	4.4241	0.07	0.36
Period*Staff	0.0903	1	0.0903	0.2100	8	0.0263	3.4404	0.10	0.30
Period*Auto	0.0213	1	0.0213	0.1307	8	0.0163	1.3010	0.29	0.14
Staff*Auto	0.0147	1	0.0147	0.2833	8	0.0354	0.4139	0.54	0.05
Period*Staff*Auto	0.0334	1	0.0334	0.1052	8	0.0132	2.5402	0.15	0.24

For the purpose of this report, the main effect of staffing solution ($F(1,8)=17.31, p=.00, \eta^2=.68$) and the main effect of automation display ($F(1,8)=4.42, p=.07, \eta^2=.36$).

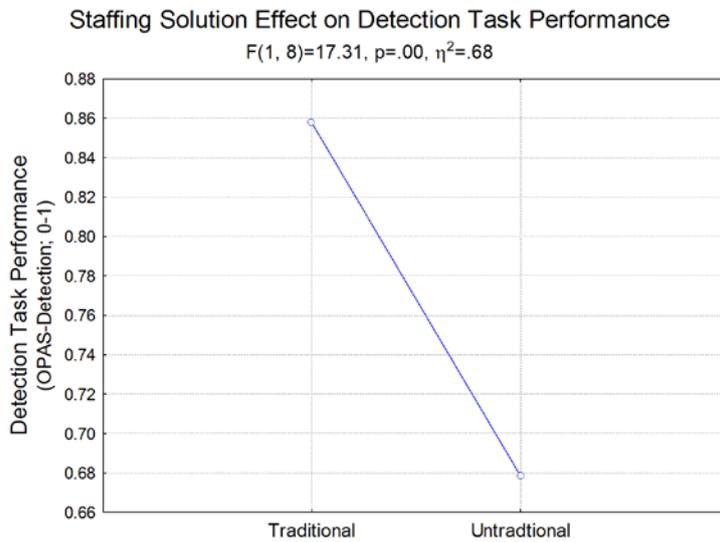


Figure 48: Staffing solution effect on Detection Task Performance (OPAS-Detection) in Study 2.

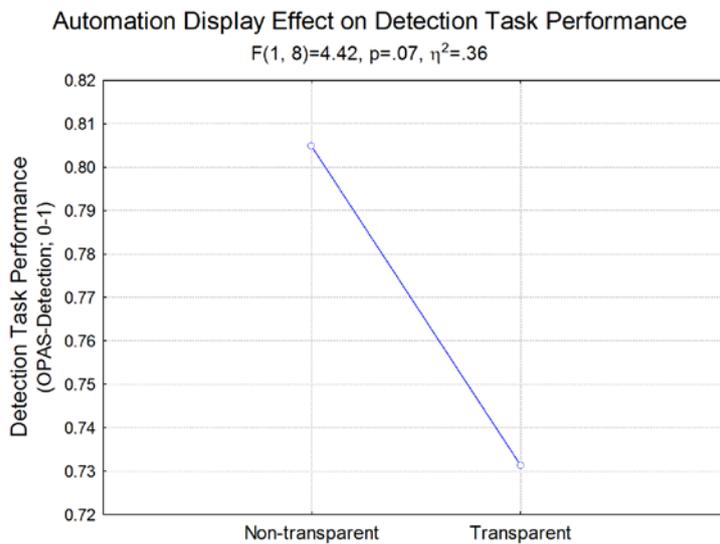


Figure 49: Automation display effect on Detection Task Performance (OPAS-Detection) in Study 2.

Figure 48 illustrates that detection task performance is superior for the Traditional compared to the Untraditional staffing solution. Figure 49 illustrates that the Transparent automation displays impede the

detection task performance in comparison to the Non-transparent automation displays. The automation display effect on detection task performance is generally opposite to the effect revealed by the Process Overview Measure (Figure 45).

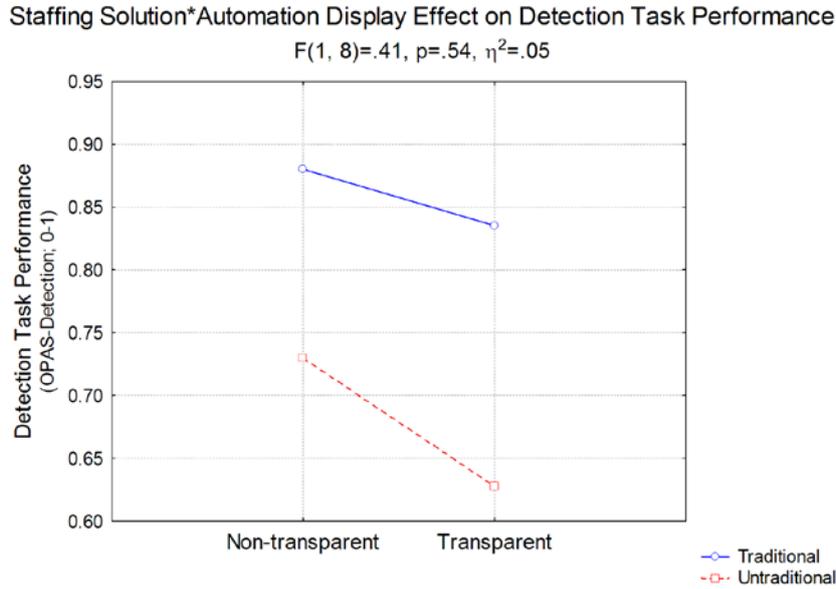


Figure 50: Scenario period, staffing solution, and home plant interaction effect on Detection Task Performance (OPAS-Detection) in Study 2.

In contrast to Process Overview, detection task performance clearly does **not** demonstrate any interaction between staffing solution and automation display ($F(1,8)=0.41, p=.54, \eta^2=.05$; Figure 50).

To test the influence of control experience about the simulator based on the nuclear power plants where operators work, the operation task performance data (OPAS-Operation) was analyzed in a one-way ANOVA using Type III/Unique sums of squares with a fixed, between factor of Plant (1, 2 and 3) in a CRF design.

Table 17: ANOVA of operation task performance (OPAS-Operation) on plant factor in Study 2.

Fixed Effects	SS	df	MS	F	p	η^2
Plant	0.56328	2	0.28164	5.6583	0.01	0.14
Error	3.43446	69	0.04977			

The one-way ANOVA indicates a difference in operation task performance between plants where operators work ($F(2,69)=5.66, p=.01, \eta^2=.14$).

Plant Effect on Operation Task Performance (OPAS-Operation)

$F(2, 69)=5.6583, p=.01, \eta^2=.14$
(0.95 confidence intervals)

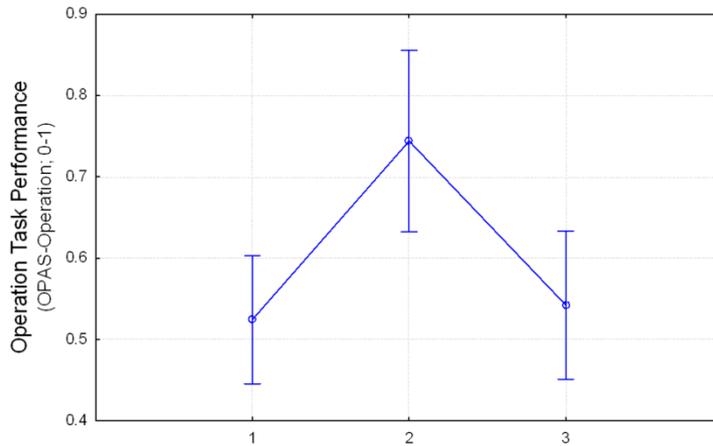


Figure 51: Plant effect on operation task performance (OPAS-Operation) in Study 2.

Figure 51 illustrates that the crews from Plant 2 exhibit the best operation task performance. This finding converges with the effect revealed by Process Overview. As mentioned, this finding is consistent with expectations because of greater control experience for operators from Plant 2 than Plant 1 and 3.

To assess the experimental manipulations, the detection task performance data (OPAS-Detection) was analyzed in an ANOVA using Type III/Unique sums of squares with fixed factors of scenario period (Easy and Difficult), staffing solution (Traditional and Untraditional), automation display (Non-transparent and Transparent) in a RBF design. (The OPAS-Operation measurements demonstrate substantial within-subject variance across experimental conditions ($ICC(3,1)=.28$) to conduct analysis in a RBF design.)

Table 18: ANOVA of Operation Task Performance (OPAS-Operation) for Study 2.

Fixed Effects	SS	Df	MS	SS-Err	Df-Err	MS-Err	F	p	η^2
Period	0.75603	1	0.75603	0.41065	8	0.05133	14.7284	0.00	0.65
Staff	0.16332	1	0.16332	0.24291	8	0.03036	5.3786	0.05	0.40
Auto	0.01287	1	0.01287	0.294	8	0.03675	0.3503	0.57	0.04
Period*Staff	0.10727	1	0.10727	0.22957	8	0.0287	3.7382	0.09	0.32
Period*Auto	0.00276	1	0.00276	0.31192	8	0.03899	0.0709	0.80	0.01
Staff*Auto	0.00011	1	0.00011	0.23136	8	0.02892	0.0039	0.95	0.00
Period*Staff*Auto	0.0083	1	0.0083	0.14869	8	0.01859	0.4466	0.52	0.05

For the purpose of this report, the main effect of staffing solution ($F(1,8)=5.38, p=.05, \eta^2=.40$).

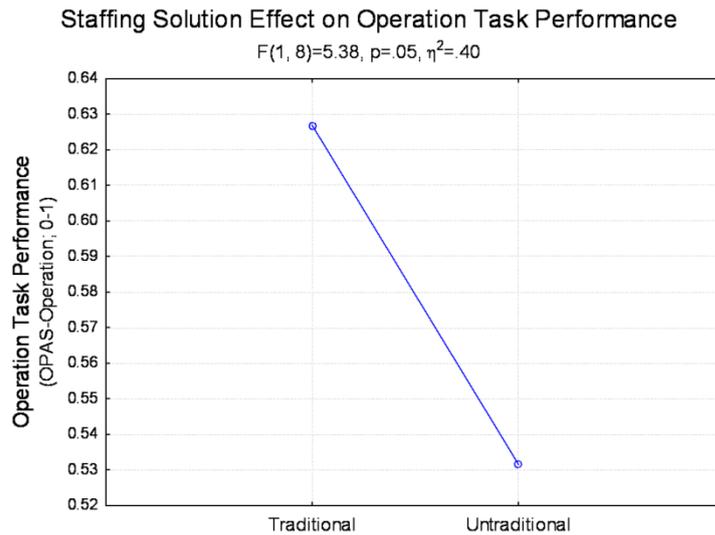


Figure 52: Staffing solution effect on operation task performance (OPAS-Operation) in Study 2.

Figure 51 illustrates that operation task performance is superior for the Traditional compared to the Untraditional staffing solution. This finding is very similar to the main effect of staffing solution on detection task performance.

[93] and [94] document in detail the debriefing data related to the staffing solution and automation displays, respectively. The Untraditional staffing solution appears unrealistic to the operators given the current operating organization, practice and culture of existing plants. Though possible during normal operations with only minor upsets, the operators would not feel comfortable operating in the Untraditional staffing solution during complex situations. The entire plant would need to re-organize such that control room operators become much more focused on monitoring the process. This qualitative feedback is generally consistent with quantitative findings in Study 2. Process Overview demonstrates a slightly more complex pattern. Specifically, there is better Process Overview with the Untraditional staffing solution when paired with Transparent automation displays than other conditions with either staffing solutions (i.e., Figure 45).

The transparent automation displays were very well-received by the participants. They all consider such displays necessary to develop trust, track automation activities, and observe operating states with high levels of automation. However, the detailed display of the Transparent automation displays (see section 7.3.2.5.2 Automation Interface) needs improvement to support problem solving. This qualitative feedback is consistent with many quantitative findings in Study 2. For instance, trust in automation increases with Transparent automation displays (see [94]). Process Overview improves with Transparent automation displays for the Untraditional staffing solution. However, Transparent automation displays provide inferior support for detection task performance, which relies heavily on problem solving ability of the operators.

7.3.4 Discussion

Study 2 provides empirical data that test the properties of the Process Overview Measure initially demonstrated in Study 1. The descriptive statistics do not indicate any ceiling or floor effects, despite the fact that this implementation of the Process Overview Measure only contains context-sensitive items (see section 7.2.4 for the discussion of Study 1 results). The noticeable ceiling effect of context-sensitive Process Overview observed in Study 1 does not appear in Study 2. This is likely because (i) the experimental environment was intentionally differentiated from conventional operating concepts to test futuristic plant concepts, and (ii) many participants work at nuclear power plants that do not have an

identical process to the HAMBO simulator. Consequently, operator routines established for sampling context-sensitive process parameters are not fully applicable. In summary, the statistics indicate that the Process Overview Measure generally provides adequate differentiation of operator monitoring performance.

Study 1 and 2 together highlight three important considerations for selecting context- and fault- sensitive parameters for the Process Overview Measure. First, for content validity, the Process Overview Measure should ideally include both context- and fault- sensitive parameters as each appears to be sensitive to a different aspect of monitoring. Note that the evidence indicating the difference between context- and fault sensitive items is preliminary due to the absence of fault-sensitive parameters in Study 2. Second, when practical constraints limit selection to one set or the other (as in Study 2), the choice between context- and fault- sensitive parameters should be driven by the purpose of experiments and use of other performance indicators. Context-sensitive parameters may be sufficient to assess the performance on routine and familiar tasks. On the other hand, fault-sensitive parameters are pertinent to assess operator response to unanticipated events. In some cases, the studies may include other measures that permit inference on either context- or fault-sensitive Process Overview, allowing the study to omit either set of parameters. Third, context-sensitive parameters, which are indicative of performance of well-established monitoring practices, can be prone to ceiling effects. Thus, researchers should consider the challenges in the scenarios and operator knowledge when employing the Process Overview Measure queries which only pertain to context-sensitive content. While further research is necessary, researchers should be aware of the relative merits of context- and fault-sensitive Process Overview in the interpretation of the Process Overview results.

The intra-class correlation, which depicts the amount of shared variance within subjects, is as low as in Study 1 (i.e., $r^2 < .01$) for Process Overview. That is, the operators were as similar to themselves responding to Process Overview queries in another scenario trial as they were to any other operators. Other measures do not demonstrate nearly as low intra-class correlations (also similar to Study 1, see section 7.2.4). The low response tendency can be interpreted in two ways¹⁹. Although field studies indicate that monitoring demands a top-down approach and judgment, Process Overview is still a product of information acquisition (as opposed to processing) from the environment that is inherently dictated by the situational characteristics (including the experimental manipulations). From this perspective, the lower response tendency compared to measurements of Scenario Understanding and task performance can be interpreted as an indication of discriminant validity. However, the exceptionally low response tendency within subjects (i.e., $ICC(3,1) < .1$) could indicate that the Process Overview Measure is not sensitive to the operator/person characteristics. That is, the Process Overview Measure is unable to capture the individual differences in monitoring behaviours between operators although it is able to capture the differences in monitoring performance between experimental manipulations. Further research is necessary to clarify the cause of the low response tendency for Process Overview. Both interpretations can contribute to the explanation of low response tendency. Because information acquisition is intrinsically more situation-driven than other types of situation assessment activities, a measure of monitoring performance needs to be especially sensitive to capture response tendency within subjects. This prompts for further development of the Process Overview Measure (also see section 8.2).

Process Overview generally does not correlate with either Scenario Understanding or task performance. On the other hand, Scenario Understanding and task performance tend to correlate highly. Both findings are consistent with Study 1. Together, the results suggest that the mechanisms for monitoring may differ substantially from those of diagnosis or control.

¹⁹ The interpretation that Process Overview is **not** driven by operator/person characteristics is very doubtful. For instance, it is unlikely that someone without any nuclear process control experience would be able to achieve the same level of Process Overview as an experienced operator.

The sensitivity of Process Overview is re-tested in Study 2 using ANOVA. The assessment on the sensitivity of Process Overview is more complex in Study 2 than Study 1. Taking all the empirical evidence together, Study 2 confirms that the Process Overview Measure is indeed a sensitive measure. The sensitivity assessment may be divided into three parts. First, the effects of plant factor in the one-way ANOVAs revealed by Process Overview, Scenario Understanding, detection and operation task performance all converge, indicating both sensitivity and validity. The plant factor provides a particularly valid test because the expected results are not related to the experimental manipulations developed by researchers.

Second, the Process Overview Measure reveals experimental effects but those effects are inconsistent with those of detection and operation task performance. In particular, Transparent automation displays provide superior Process Overview for the Untraditional but not for the Traditional staffing solution; whereas, detection and operation task performance decline with the Untraditional staffing solution. Further, detection task performance declines with Transparent automation displays (without interaction with staffing solutions). While Process Overview demonstrates sensitivity by revealing experimental effects, the inconsistency with detection and operation task performance is generally unexpected. Thus, this evidence only indicates sensitivity, and not validity, of the Process Overview Measure.

Third, the debriefing data (as well as subjective rating scales) suggests that operators in the Untraditional staffing solution were overwhelmed as the scenarios became challenging and they relied heavily on the Transparent automation displays when available. This feedback gives insight into the quantitative data. The interaction effect of Process Overview may be a consequence of operators tending to fixate on monitoring with the Transparent automation displays when they are overwhelmed by scenario challenges operating in the Untraditional staffing solution (due to lower than conventional staff level). [93, 94] discuss the ANOVA effects and debriefing data with respect to the experimental manipulations in detail. For the purpose of this report, the debriefing data and subjective ratings of the operators provide sufficient indication that the interaction effect revealed by Process Overview is not a result of chance but of experimental manipulations. In sum, Study 2 yields valid, empirical indications of sensitivity for the Process Overview Measure.

A pattern appears to be emerging between the measurements of Process Overview, Scenario Understanding and task performance based on the examination of correlations and sensitivities in Study 1 and 2. Process Overview appears to be discriminant from Scenario Understanding and task performance given the low correlations. From one perspective, measures discriminant from other measures can provide additional information about the experimental manipulations of interests. From another perspective, the virtually complete lack of relationship with task performance raises concerns about criterion validity. On criterion validity, the investigation into sensitivity of the Process Overview Measure provides confidence that the Process Overview Measure does yield valid results, alleviating the concerns on criterion validity. The literature has also been elusive and equivocal about empirical relationship between SA measurements and task performance in general (i.e., independent of domain), arguing that the lack of clear relationships between SA and task performance is expected given the number of other moderating factors related to task performance (e.g., [91, 101]). A key remaining issue appears to be establishing the precise empirical relationship between Process Overview and task performance amongst other factors. The empirical modelling work is beyond the scope of this report. Nevertheless, the evidence thus far suggests that the Process Overview Measure yields valid results and could contribute to the empirical modelling between SA and task performance in nuclear process control.

7.3.5 Corollary issues

Study 1 and 2 examine the practicality and sensitivity of the Process Overview Measure for HAMMLAB experiments. The reliability of the Process Overview Measure has not been discussed. In particular, the

inter-rater reliability of the responses to the queries (i.e., the reference answers) between process experts requires evaluation as Process Overview indicates that monitoring is associated with a substantial amount of judgment. Study 3 turns to the reliability of the Process Overview Measure.

7.4 Study 3

Study 3 is a controlled study that investigates the inter-rater reliability of three measures:

- Process Overview Measure;
- Operator Performance Assessment System (OPAS; [13, 52, 94]); and
- Automation and Scenario Understanding Rating Scale (ASURS; [100]).

7.4.1 Objective

For the purpose of this report, the objective of Study 3 is to assess the inter-rater reliability of the Process Overview Measure. Specifically, the Process Overview Measure should demonstrate, at minimum, adequate inter-rater reliability in a representative process control environment (see section 7.1 and e.g., [72]).

7.4.2 Method

Study 3 was conducted in two parts and relied on Study 2 as source observations and data. In this report, the first and second parts are referred to as Study 3a and 3b, respectively. Study 3a collected data from two raters who answered Process Overview queries in real time during the scenario trials of Study 2, whereas, Study 3b collected data from one rater who answered queries after the experiment based on graphs of the parameters generated from simulator log data from the same trials in the first part of Study 3. (That is, raters in Study 3a followed a different procedure than the rater in Study 3b.)

7.4.2.1 Participants

Three raters participated in this study. Due to the small number, the characteristics of participants may be particularly important for interpretation of results. The raters are coded as A, B, and C with characteristics described below.

Rater A was a process expert employed at the HRP (for two years before Study 2). He worked as a control room operator in multiple nuclear plants for fifteen years and participated as a process expert to develop of the HAMBO simulator in late 1990's. He designed all of the scenarios for Study 2 and advised other HRP scientists on the processes and operations of the physical and simulator plant. He also identified all of the process parameters for the Process Overview queries. During data collection, Rater A played the role of field operator and electrician as required by the scenarios and participant interventions. He was also the designated process expert for completing questionnaires related to participant performance.

Rater B was a recently retired (within two months) shift supervisor at the nuclear plant that the HAMBO simulator replicates. He worked as a control room shift supervisor for 32 years and participated in about 10 HRP studies prior to Study 2. He was a participant for the pilot trials of Study 2.

Rater C was a process expert employed at the HRP for eleven years. He worked as a control room operator in multiple nuclear plants for fifteen years and participated as a process expert in the development of the HAMBO simulator. Prior to Study 2, he had similar responsibility for numerous HAMMLAB simulator studies. For Study 2 preparation, he developed the scripts that automatically start-up the simulator plant, implemented the scenarios, co-designed the transparent automation interface and

advised other HRP scientists on the processes and operations of the physical and simulator plant. During data collection, he was the designated technical support person for simulator operations.

All raters demonstrated substantial experience working in NPP as control room operators. Therefore, they could be the process experts implementing the Process Overview Measure to assess quality of technology and training in representative simulator settings. The raters did show differences in their experience and knowledge of the HAMBO simulator and details of the experiment (i.e., Study 2) that may influence their judgments and thus the final results.

7.4.2.2 Experimental Environment

Study 3a was conducted in the observation gallery during the data collection of the last two participating crews in Study 2. That is, Study 3a occurred simultaneously with the last two crews of Study 2. This section only describes the adaptations made to the observation gallery to permit Study 3a and 3b. (For details on the original setup of the observation gallery for Study 2, refer to section 7.3.2.)

The gallery setup for Study 3a (Figure 53) included an extra workstation to support the additional rater (compared to Study 2). The workstation (Figure 54) provided the rater with (i) two displays for process information, (ii) a pair of earbuds for audio feed from the control room, (iii) a laptop for inputting scores/ratings, (iv) a pair of binoculars for observing participants close up, (v) descriptions of all the scenarios, (vi) pen and paper for note taking, and (vii) a pair of ear muffs for muting any discussion on operator performance in the observation gallery. Furthermore, an opaque, black curtain was installed to block the raters' lines of sight to (i) each other's monitors used for collecting their responses to questionnaires and (ii) any body language expressed from judging operator performance.

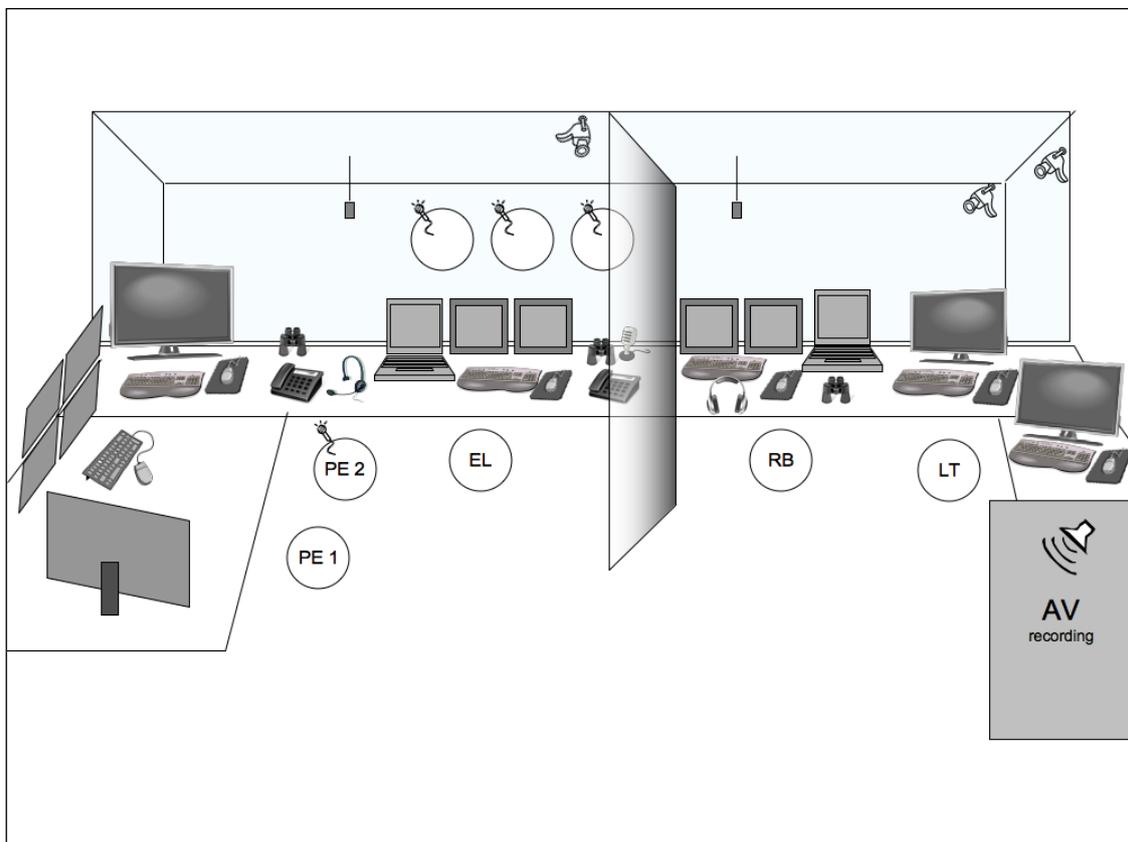


Figure 53: Observation gallery for Study 3a (and for last two participating crews of Study 2). PE 1= process expert 1; PE 2 = process expert 2; EL= experimental leader; RB = Rater B (in Study 3); LT = laboratory technician



Figure 54: Workstation for Rater B (recruited shift supervisor) in Study 3b.

Study 3b was conducted in an office after the data collection of Study 2. The office was equipped with a computer connected to five large LCD monitors and a projector. The key reason for selecting this room for data collection was space and privacy for the Rater C. Only the computer and two LCD monitors were used to collect data in Study 3b.

7.4.2.3 Procedure

7.4.2.3.1 Study 3a

Rater A (in Study 3a) was given a verbal introduction about the study and instructed to withhold all discussion about ratings and participants' performance from Rater B. He had received instruction on the Process Overview Measure in the preparation phase of Study 2. Therefore, he was not given any further instruction as he was the process expert responsible for all the ratings in Study 2. (Refer to section 7.3.2.4 for a description of the process expert's responsibilities.) In other words, his responsibilities for Study 3a were a subset of Study 2.

Rater B (in Study 3a) was responsible for answering the Process Overview queries (and rating other performance indicators). Table 19 presents an overview of his activities in Study 3a.

Table 19: Key activities of Rater B in Study 3.

Day	Time	Activity of Study 3a Participants (raters)	Activity of Study 2 Participants (crews)
Tuesday	08:30-9:30	Introduction & Overview of Study 2	Crew 8 Training
	09:30-11:00	Training on measures, experimental protocols and apparatus	Crew 8 Training
	11:00-12:00	Lunch	
	12:00-13:30	Training Trial	
	13:30-14:30	Practice Trial #1 (data not-analyzed)	Crew 8 Trial #1
	14:30-16:00	Practice Trial #2 (data not-analyzed)	Crew 8 Trial #2
Wednesday	08:00-09:30	Trial #1	Crew 8 Trial #3
	09:30-11:00	Trial #2	Crew 8 Trial #4
	11:00-12:00	Lunch	
	12:00-13:30	Trial #3	Crew 8 Trial #5
	13:30-15:00	Trial #4	Crew 8 Trial #6
Thursday	08:00-09:30	Trial #5	Crew 8 Trial #7
	09:30-11:00	Trial #6	Crew 8 Trial #8
2 Weeks of Separation			
Tuesday	09:30-11:00	Review of Training Materials	Crew 9 Training
	11:00-12:00	Lunch	
	12:00-13:30	Training Trial	
	13:30-14:30	Trial #7	Crew 9 Trial #1
	14:30-16:00	Trial #8	Crew 9 Trial #2
Wednesday	08:00-09:30	Trial #9	Crew 9 Trial #3
	09:30-11:00	Trial #10	Crew 9 Trial #4
	11:00-12:00	Lunch	
	12:00-13:30	Trial #11	Crew 9 Trial #5
	13:30-15:00	Trial #12	Crew 9 Trial #6
Thursday	08:00-09:30	Trial #13	Crew 9 Trial #7
	09:30-11:00	Trial #14	Crew 9 Trial #8
	11:00-12:00	Lunch	
	12:00-13:30	Debriefing	Debriefing (separate from Study 3)

Instruction for Rater B was given in two sessions. The first session consisted of a Powerpoint presentation providing an overview of the experiment. It also included a review of the inter-rater reliability concept and explicit instruction to prevent any discussion about measurements with anyone (except the experimentalist).

The second session began by reviewing the ratings (i.e., the experimental task) to be performed concurrent with Study 2. Theoretical descriptions of the performance constructs were omitted. Rater B first received samples of the Process Overview queries, followed by a short Powerpoint presentation describing the general structure of the queries. Specifically, "recently" in the queries was described as the last meaningful change of the parameter according to his expert judgment. In addition, the parameter

change should account for noise in the process. The second session ended with familiarizing the Rater with the apparatus and data collection tools in the observation gallery.

During the experiment in Study 2, Rater B was seated at his workstation and performed ratings by entering his response with mouse and keyboard into the Halden Questionnaire System (HQS; in the same manner as Rater A). Because Rater B was new to the role of rater of operator performance in experimental settings, the first two trials of Crew 8 were treated as practice sessions, with scores omitted from the data analysis.

7.4.2.3.2 Study 3b

Study 3b was conducted four weeks after the data collection of Study 2. That is, the data collection from Rater C was not concurrent with Study 2 and other raters. Rater C was given a brief introduction of study. He was verbally instructed to review trend graphs of process parameters (e.g., Figure 55) and answer the Process Overview queries²⁰. To support his ratings with contextual information, Rater C was provided with the scenario descriptions and OPAS scores (i.e, the actions performed by the participants with respect to the solutions prescribed by the experts; see section 7.3.2.7.3). In total, Rater C scored 290 queries (i.e., determined the reference keys) in two sessions on the same day, with a lunch break in between. Study 3b contained fewer queries than Study 3a because several parameters were not logged.

The graphs (e.g., Figure 55) were generated by a java applet – Flot – and compiled into PowerPoint files for presentation on one LCD monitor. The graphs were presented in the order of Crew, Scenario, and Period. The HQS presented the queries and collected the responses on a separate LCD monitor.

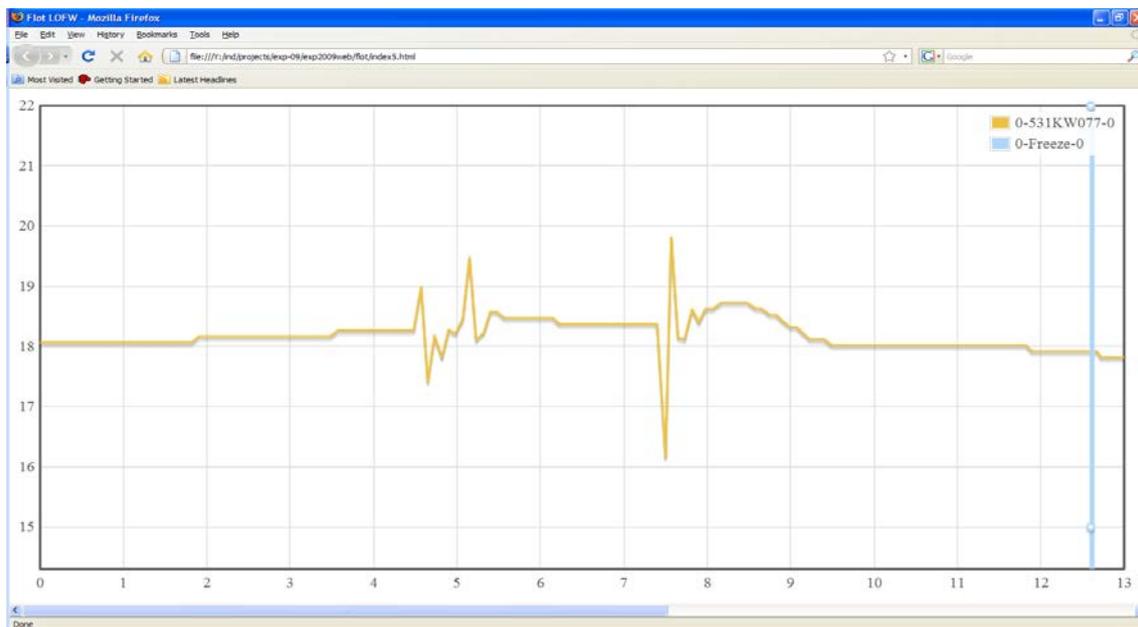


Figure 55: Example of Flot-generated graph based on simulator log data.

²⁰ Rater C was very familiar with the Process Overview queries. He was the designated process expert for answering Process Overview queries in Study 1 and SACRI queries in prior HRP experiments. Hence, he was not given detailed descriptions/instruction about the Process Overview Measure.

7.4.2.4 Measure of Inter-rater Reliability

Cohen's Kappa, κ , [102] is applied to assess the agreement between raters on a nominal scale²¹. Equation [3] and [4] express κ and its standard error, σ_{κ} , respectively.

$$\kappa = \frac{p_o - p_c}{1 - p_c} = \frac{f_o - f_c}{N - f_c}, \tag{3}$$

$$\sigma_{\kappa} = \sqrt{\frac{p_o(1 - p_o)}{N(1 - p_c)^2}} = \sqrt{\frac{f_o(N - f_o)}{N(N - f_c)^2}}, \tag{4}$$

where (for both Equation 3 and 4) p_o =proportion of units agreed; f_o = frequency of units agreed; p_c =proportion of units agreed expected by chance; f_c =frequency of units agreed expected by chance; N denotes total sample size.

κ ranges from -1 to 1, for which the lower bound expresses complete disagreement and upper bound expresses complete agreement. The statistic is similar to basic correlation statistics except that κ is adjusted for the agreement by chance. Table 20 (identical to Table 4) outlines the common interpretation of the Kappa statistic [72].

Table 20: Interpretation of the Cohen's Kappa, κ .

Kappa coefficient	Strength of agreement
≤ 0	Poor
$0 < \kappa \leq .2$	Slight
$.2 < \kappa \leq .4$	Fair
$.4 < \kappa \leq .6$	Moderate
$.6 < \kappa \leq .8$	Substantial
$.8 < \kappa \leq 1$	Almost perfect

Cohen also provides Equation [1] for test of significance between two κ 's [102].

$$z = \frac{\kappa_1 - \kappa_2}{\sqrt{\sigma_{\kappa_1}^2 + \sigma_{\kappa_2}^2}} \tag{5}$$

7.4.2.5 Methodological Note

For assessing inter-rater reliability, all raters would typically be given exactly the same tasks and situated in the same environment. Unfortunately, such ideal conditions are simply unrealistic in highly representative experiments involving process experts that are always scarcely available. It is impractical to recruit two or more process experts to perform ratings for the sole purpose of evaluating a single measure. It is also impractical to have two or more process experts to conduct the same tasks during data collection for the sake of studying reliability. We are fully aware that this experiment involves some atypical properties relatively to more controlled studies but studies in highly representative environments most likely entail similar practical constraints in which different process experts would be responsible for

²¹ Intra-class correlation, a statistic of inter-rater reliability for interval scales, may be calculated for the proportion correct of the queries for each scenario period. However, the aggregation may lead to misrepresented intra-class correlation values. Appendix A illustrates the potential problems of applying intra-class correlation for assessing inter-rater reliability for the Process Overview measure.

different tasks. Process experts may also be responsible for different tasks from one experiment to another. Therefore, to be practically useful, a measure should be sufficiently robust in reliability to these uncontrolled properties of this study. The reliability results should also be representative of the research circumstances in applied settings, where a substantial increase in the controls of process expert work and training on the measures is often impractical.

7.4.2.6 Evaluation Criteria (Hypothesis)

In experiments representative of nuclear process control settings, measurements on operator knowledge derived from significant judgment should exhibit “moderate” to “substantial” inter-rater reliability (also see section 7.1). [57] indicates that typical rating scales generally have “moderate” reliability and [72] reports that medical diagnosis (which also involves significant cognitive judgment) would typically demonstrate a κ range from moderate to substantial. Therefore, the Process Overview Measure should demonstrate a κ of above .4.

7.4.3 Data Analysis and Results

Table 21 presents the κ 's and the standard errors between the raters. (Appendix B documents the frequency distributions between the three raters with respect to the participant responses in Study 2²².)

Table 21: Kappa statistics between raters in Study 3.

Raters	B		C	
	κ	σ	κ	σ
A	0.473	0.045	0.620	0.042
B	-		0.341	0.049

The κ statistics range from “fair” to “substantial” agreement (see Table 4) between the raters, indicating that the Process Overview Measure is acceptable in terms of reliability for measuring monitoring performance.

Another finding is that the differences between the κ 's appear unexpectedly large (Table 21 and Figure 56), indicating a rater effect. ANOVA is not available for testing multiple κ 's. Thus, significance of rater effect is tested with multiple comparisons following Holm's procedure for controlling type I error rate [61].

Table 22 indicates significant differences between all the κ 's: $\kappa_{A,C} > \kappa_{A,B}$ ($Z=2.39, p<0.05$); $\kappa_{A,C} > \kappa_{B,C}$ ($Z=4.32, p<0.05$); and $\kappa_{A,B} > \kappa_{B,C}$ ($Z=1.99, p<0.05$) suggesting a significant effect of rater.

Table 22: Multiple comparisons of κ 's following the Holm's multistage procedure.

Raters	AC		BC	
	Z	p	Z	p
AB	2.39	.017*	1.99	.047*
AC	-		4.32	.000*

*Significant effect following Holm's procedure to control for type I error-rate

²² The frequency distributions incorporating the responses of the participants provide details on the scoring of the Process Overview measure in addition to rater responses. Appendix C documents the κ results purely based on the responses of the rater (without any reference to the responses of the participants). The difference between the two sets is negligible.

Figure 56 graphs the κ 's between raters in Table 21, illustrating that the highest and lowest reliabilities are both associated with collecting data from one rater in real time during the experiment and another rater after the experiment.

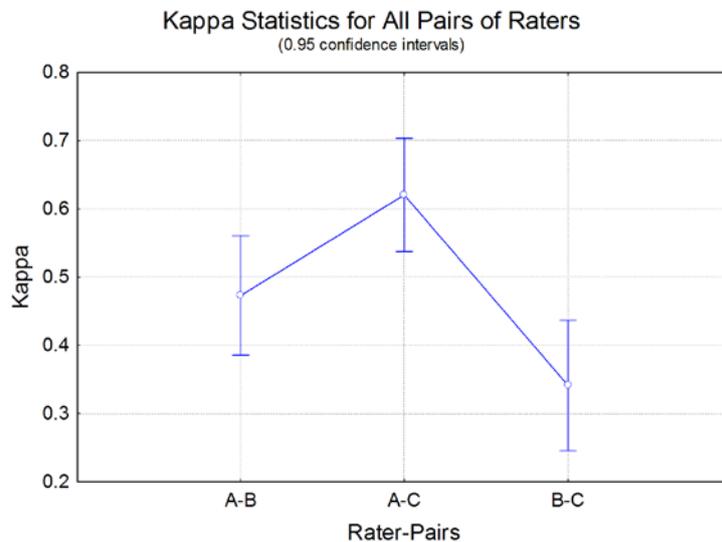


Figure 56: Mean and 95% confidence intervals of Kappas in Study 3.

During the data collection of Study 3b, Rater C often referred to the scenario descriptions to support his judgment about the parameter behaviours. He seldom reviewed the OPAS scores of the operators in Study 2 for guidance. The contextual information provided by the scenario descriptions appeared to be particularly important for monitoring and judging parameter behaviours.

7.4.4 Discussion

The κ results generally indicate “moderate” inter-rater reliability, suggesting that measurement errors related to inconsistencies between raters should pose only a small effect on data analysis (see [57, 72, 103]). For representative setting in nuclear process control, Study 3 indicates that the Process Overview Measure is sufficiently reliable for assessing monitoring performance in representative nuclear process control settings.

Moderate, as opposed to ideal (i.e., $\kappa > .8$), inter-rater reliability of the Process Overview Measure is an indication that the queries demand substantial inferences or cognitive judgement to answer. In other words, operator detection of parameter changes (i.e., monitoring tasks in NPPs and representative experiments) is likely a result of not only sensory registration of some signals but also information processing of the signals in relations to the operating context. The demand on information processing and importance of context is also indicated by the observation that Rater C frequently referred to the scenario descriptions when inspecting graphs of various parameters and responding to the queries (Study 3b). The statistics and observation together reflect the characteristics captured by Process Overview.

The κ results also illustrate the different influences that rater characteristics and data collection procedures have on inter-rater agreement. Rater A and Rater C demonstrated the greatest agreement; Rater B and Rater C demonstrated the least agreement; Rater A and B demonstrated an intermediate level of agreement. We believe the most likely explanation relates to the degree of involvement in the preparation of the experiment (Study 2) and data collection procedures (i.e., Study 3a vs. Study 3b). Rater A and C, both HRP process experts, were heavily involved in preparing Study 2 and, hence, possessed deep knowledge of the simulator and scenarios. In contrast, Rater B was only exposed to

Study 2 as a participant in the pilot trials. Thus, even though Rater A answered the queries in real time (Study 3a) while Rater C answered them after the experiment (Study 3b), they demonstrated the greatest agreement. Raters A and B, on the other hand, shared the same data collection procedure, resulting in the second highest agreement. Raters B and C demonstrated the lowest level of agreement as they shared neither preparation of Study 2 nor data collection procedure. Given the circumstances, this finding is not that surprising. The clearest implication of these results may be that knowledge of the experiment is probably more important than the overall data collection procedure or environment.

The κ results on Process Overview also direct attention to inter-rater reliability of query- and probe-based techniques in general, including the commonly used SAGAT. The discussion in the SA literature on the methodology of query- or probe- based techniques generally overlooks the issue of inter-rater reliability, focusing instead on other issues such as intrusiveness and query content in relation to the construct. The SA literature does not provide any guidance for constructing queries or probes that explicitly consider reliability of the measure. When queries and probes are well-defined, inter-rater reliability may be inherently high and thus irrelevant. However, as queries or probes become complex and ill-defined (e.g., comprehension of a complex set of signals), most techniques rely on subject-matter-experts for reference answers without explicit consideration of inter-rater reliability (e.g., [46]). Study 3 provides the first explicit examination of inter-rater reliability of query-based SA measures. The Process Overview queries, though seemingly direct, demand substantial information processing from process experts to answer and inter-rater reliability of the answers between the process experts is not perfect. The empirical results alert researchers to the issue of reliability of query- and probe- based techniques for measuring SA.

PART III: GENERAL DISCUSSION

8. DISCUSSION

The notion of Situation Awareness is increasingly useful to discuss operator work as nuclear process control is increasingly knowledge-based. The literature contains a substantial amount of SA research; however, much of the focus has been on either abstract theories and models of SA, or on domains other than process control. In SA research thus far, abstract theories or models do not clearly reflect the nature of monitoring in the nuclear domain, while adaptation of theories and domain-specific characterizations are absent. As a result, the nuclear industry cannot fully capitalize on the SA notion because current research neglects characteristics unique to the domain. To advance SA for in nuclear process control, we develop Process Overview specifically to characterize SA acquired through monitoring of NPPs and apply this domain-specific concept for operationalizing SA and designing control room technology (Part I). We further apply Process Overview to guide the modification of the SAGAT and SACRI measures into the Process Overview Measure and evaluate the Process Overview Measure on practicality, sensitivity and reliability in three empirical studies (Part II). We now discuss the concept, measure and empirical data together to illustrate the overall contribution of this work to research in nuclear process control and SA.

8.1 Nuclear Process Control

The Process Overview Measure is a substantial modification of the SAGAT and SACRI measures to operationalize Process Overview - the SA acquired through monitoring of NPPs. SAGAT offers a useful measurement framework to elicit operator knowledge for assessment but omits many methodological details necessary to fully operationalize SA in specific contexts. In other words, SAGAT leaves many decisions to the investigators without much theoretical and methodological guidance. SACRI adapts SAGAT and Signal Detection Theory (SDT) to operationalize SA. The SACRI measure introduces methodological details relevant to assessing SA in process control. However, aspects of SACRI are incompatible with field research on process control (as captured by Process Overview) and SDT. Hence, the Process Overview Measure is developed in adaptation of both SAGAT and SACRI in accordance to the research on monitoring process plants.

Chapter 5 and 6 together present the details and comparisons between SAGAT, SACRI and the Process Overview Measure (also see Table 1). Briefly, compared to SAGAT and SACRI, the Process Overview Measure is different (to various extents) in:

- employing Process Overview as the conceptual foundation,
- specifying a single query structure and corresponding response format,
- determining relevant process parameters for queries according to scenario characteristics,
- selecting timing of query administration according to scenario characteristics, and
- calculating the final scores as proportion correct.

Through these methodological adaptations, the Process Overview Measure is expected to collect psychometrically sound SA measurements.

The three empirical studies indicate that the Process Overview Measure can collect SA measurements that are both sensitive and reliable. The Process Overview Measure reveals significant effects of experimental manipulations that corroborate with other empirical results in two full-scope simulator experiments investigating dramatically different operational concepts. The Process Overview Measure also demonstrates adequate inter-rater reliability in a controlled experiment. Furthermore, the researchers and process experts are convinced of the practicality of employing the measure in full-scope

NPP simulator experiments. Therefore, the Process Overview Measure is a practical, reliable and sensitive method to measure Process Overview.

The Process Overview Measure offers two general advantages over SAGAT and SACRI. First, the Process Overview Measure provides conceptual and methodological details leading to improved consistency in measuring SA across studies for nuclear process control. For instance, it specifies a single query structure and response format that are evaluated on both sensitivity and inter-rater reliability. Second, the Process Overview Measure collects data in a manner that is compatible with operator cognitive work. For instance, the queries do not elicit knowledge on prediction and irrelevant process parameters, which are practically unnecessary and conceptually incompatible, respectively, for nuclear process control. In essence, the Process Overview Measure prescribes sufficient methodological details to measure SA effectively. In effect, full-scope simulator experiments can employ the Process Overview Measure to reveal the impact of new control room technology and operational concepts on monitoring.

Besides assessment on practicality, sensitivity and reliability of the measure, the empirical data also speak to the concept of Process Overview. The empirical findings of the inter-rater reliability study reflect on the foundation of Process Overview (i.e., field research on process operator work). The inter-rater reliability (i.e., κ) statistics are not trivially high, suggesting expert judgment as well as sensory registration of signals (i.e., parameter values) is necessary to respond to Process Overview queries. This finding is consistent with field studies that identify substantial information processing demand during monitoring. The difference in inter-rater reliability between different rater-pairs suggests that intimate knowledge about the scenarios (i.e., the problem and context), as well as the physical process, appears critical for being the “experts” or raters in the representative experiments. Furthermore, the frequent reference to scenario descriptions by one rater when responding to the queries post-experimental trails (Study 3b) confirms the importance of contextual information during monitoring. The statistics and observations in the inter-rater reliability study provide additional empirical support of the Process Overview concept.

The empirical results across the two full-scope simulator studies suggest the relationships between situation awareness and task performance are more complex than many models of SA (e.g., Endsley’s three levels of SA) intuitively suggest. There appears to be no correlation between Process Overview and either Scenario Understanding or task performance, but high correlation between Process Overview and Scenario Understanding. Furthermore, the response tendency of Process Overview is much lower than Scenario Understanding and task performance measurements. Based on these results, monitoring tasks appear to be substantially different from diagnosis and control tasks. Even at general or abstract level (e.g., knowledge from monitoring and diagnosis), the relationships between different aspects of SA seems to be beyond what correlation statistics could depict. The literature has also been equivocal and elusive on the empirical relationships between different aspects of SA and task performance (e.g., [91]). Determining the relationships between Process Overview, Scenario Understanding and task performance most likely involves other moderating factors (or constructs) and requires sophisticated modelling techniques.

8.2 Situation Awareness

While the primary research aim of this work was to advance nuclear process control, the conceptual and empirical investigation have some direct implications on SA research in general. Study 1 and 2 offer the first discussion on the relative contribution of situational and person characteristics to the variance captured by an SA measure. In Study 1 and 2, the measurements of Process Overview demonstrate substantially lower response tendency (as measured by intra-class correlations) than measurements of Scenario Understanding and task performance. That is, operators were as similar to themselves

responding to Process Overview queries in another scenario trial as they were to any other operators. This suggests that variance in the Process Overview measurements appears to be driven by situational characteristics with only negligible contribution from operator characteristics. The issue of relative contribution of situational and person characteristics in SA measurements is unexplored in the literature and may be pertinent to the development of both SA theory and measurement. Research on SA measurements should thus examine response tendencies within subjects (for repeated measures experimental designs) using intra-class correlations for different measurement methodologies and aspects of SA (e.g., Process Overview, levels of SA). The ICC can shed light on the relative performance impacts of situation and person characteristics that are being captured by the measures. Though preliminary, this research presents a new technique (i.e., ICC) and topic (i.e., the relative dependence on situation and person) to study SA concepts and measures.

Study 3 exposes a widely neglected issue of SA measurement – inter-rater reliability. Process Overview Measure demonstrates acceptable, but not ideal, reliability for seemingly straightforward queries, prompting researchers to be aware of reliability concerns as well as content and construct validity in formulating queries to elicit SA in complex situation. The empirical results indicate that many queries in representative settings involve substantial judgment (i.e., cognition), and subject-matter-experts may **not** necessarily agree with each other even for seemingly straightforward queries and probes (as a natural outcome of expertise and complexity). Therefore, query- and probe- based measurement techniques, which are most commonly applied in SA research and assessment (e.g., SAGAT [46]), cannot assume inter-rater reliability. However, probe- and query-based techniques generally assume that process experts can provide the “correct” answers to queries that cannot be produced by the simulators (e.g., level 2 SA/comprehension queries). Furthermore, the SA literature contains neither discussion on measurement reliability nor guidance on developing queries on operator knowledge about complex situations. Study 3 provides the first empirical evidence and discussion that direct researchers to attend to the potential inter-rater reliability concerns of query- and probe- based measurement techniques.

The development and evaluation of the concept and measure of Process Overview altogether exemplify the value of adopting a domain-specific approach to study SA. As argued, the approach advocates more precise descriptions of the SA dimensions than general SA models, guiding development and interpretation of measurements in a manner pertinent to the domain. Not only do the empirical results following this approach indicate success for studying and measuring SA in process control, but the results also offer insights into SA research independent of domains. As mentioned, the finding on low response tendency for Process Overview prompts investigation of other SA measures and concepts into the relative impact of situation and person characteristics. Similarly, empirical findings on inter-rater reliability highlight the research gap in query- and probe- based measures. While the approach inherently compromises some generalization for precision, the domain-specific SA research can still contribute to both theory and measurement of SA.

9. CONCLUSION

This report presents a domain-specific characterization of SA for monitoring NPPs – Process Overview. Process Overview is built upon technical and operational properties of NPPs and field research on monitoring activities of control room operators. The concept captures domain-specific details supporting design of control room technology and operationalization of SA.

The utility of the Process Overview concept for operationalization is exemplified in transforming the SACRI measure into the Process Overview Measure. The Process Overview Measure is evaluated in three empirical studies demonstrating both sensitivity and reliability as a SA measurement tool in

representative nuclear process control environment. This human performance assessment tool is valuable for evaluation and validation of control room technology and operational concepts.

The concept and measure of Process Overview represents an explicit attempt to apply a domain specific approach for studying and applying SA in the nuclear domain. The approach demonstrates promise for SA research through capturing details of operator work that support designs and guiding sensitive and reliable measurements that reveal impacts of new operational concepts and technology.

10. REFERENCES

- [1] K. J. Vicente, *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, N.J.: Lawrence Erlbaum Associates, 1999.
- [2] N. Moray, "Où sont les neiges d'antan?," in *Human Performance, Situation Awareness and Automation*, Daytona Beach, FL, 2004.
- [3] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, pp. 32-64, 1995.
- [4] P. A. Hancock, *et al.*, "The Real Deal: Essential Lessons from Human Factors Leaders," presented at the at the 51st Annual Meeting of the Human Factors and Ergonomics Society, Baltimore, MD, USA, 2007.
- [5] N. A. Stanton, "Situation awareness: where have we been, where are we now and where are we going?," *Theoretical Issues in Ergonomics Science*, vol. 11, pp. 1 - 6, 2010.
- [6] S. P. Banbury and S. Tremblay, Eds., *A Cognitive Approach to Situation Awareness: Theory and Application*. Hampshire, UK: Ashgate, 2004.
- [7] M. R. Endsley, "Theoretical Underpinnings of Situation Awareness: A Critical Review," in *Situation awareness: analysis and measurement*, M. R. Endsley and D. J. Garland, Eds., Mahwah, NJ: Lawrence Erlbaum Associates, 2000, pp. 3-32.
- [8] J. M. Flach and J. Rasmussen, "Cognitive engineering: Designing for situation awareness," in *Cognitive Engineering in the Aviation Domain*, N. B. Sarter and R. Amalberti, Eds., Mahwah, NJ: Erlbaum, 1999.
- [9] J. M. Flach, *et al.*, "The concept of the situation in psychology," in *A Cognitive Approach to Situation Awareness: Theory and Application*, S. P. Banbury and S. Tremblay, Eds., Hampshire, UK: Ashgate, 2004, pp. 42-60.
- [10] J. Patrick and N. James, "A task-oriented perspective of situation awareness," in *A Cognitive Approach to Situation Awareness: Theory and Application*, S. P. Banbury and S. Tremblay, Eds., Hampshire, UK: Ashgate, 2004.
- [11] D. J. Bryant, *et al.*, "Modeling Situation Awareness in an Organizational Context: Military Command and Control," in *A Cognitive Approach to Situation Awareness: Theory and Application*, S. P. Banbury and S. Tremblay, Eds., Hampshire, UK: Ashgate, 2004.
- [12] S. Freedman and J. A. Adams, "The Inherent Components of Unmanned Vehicle Situation Awareness," *Proc. of the 2007 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 973-977, 2007.
- [13] G. Skraaning Jr., *et al.*, "The Ecological Interface Design Experiment (2005)," OECD Halden Reactor Project, Halden, Norway HWR-833, 2007.
- [14] D. N. Hogg, *et al.*, "Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms," *Ergonomics*, vol. 38, pp. 2394-2413, 1995.

- [15] D. N. Hogg, *et al.*, "Measurement of the Operator's Situation Awareness for Use Within Process Control Research: Four Methodological Studies," OECD Halden Reactor Project, Halden, Norway HWR-377, 1994.
- [16] T. B. Sheridan and G. Johannsen, Eds., *Monitoring Behaviour and Supervisory Control*. New York, NY: Plenum Press, 1976.
- [17] E. Edwards and F. P. Lees, *The human operator in process control*. New York: Halsted Press, 1974.
- [18] R. J. Mumaw, *et al.*, "There Is More to Monitoring a Nuclear Power Plant than Meets the Eye," *Human Factors*, vol. 42, pp. 36-55, 2000.
- [19] J. Rasmussen, *et al.*, Eds., *New Technology and Human Error* (New Technologies and Work. New York, NY: John Wiley & Sons, 1987.
- [20] C. D. Wickens and J. G. Hollands, *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice Hall, 2000.
- [21] F. Cara and V. Lagrange, "Emerging expertise in process control," *Ergonomics*, vol. 42, pp. 1418 - 1430, 1999.
- [22] J. Reason, *Human Error*. Cambridge, England Cambridge University Press, 1990.
- [23] C. Perrow, *Normal accidents: Living with high-risk technologies*. New York, USA: Basic Books, 1995.
- [24] D. A. Lucas, "Mental Models and New Technology," in *New Technology and Human Error*, J. Rasmussen, *et al.*, Eds., Chichester, UK.: John Wiley & Sons Ltd., 1987.
- [25] K. J. Vicente, *et al.*, "Operator monitoring in a complex dynamic work environment: a qualitative cognitive model based on field observations," *Theoretical Issues in Ergonomics Science*, vol. 5, pp. 359-384, 2004.
- [26] L. Bainbridge, *et al.*, "A study of real-time human decision making using a plant simulator," in *The Human Operator in Process Control*, E. Edwards and F. P. Lees, Eds., New York, NY: Taylor & Francis, 1974, pp. 91-104.
- [27] N. Moray, "Process Control," in *International Encyclopedia of Ergonomics and Human Factors*, vol. 2, W. Karwowski, Ed., Boca Raton, FL, USA: Taylor and Francis Inc., 2006, pp. 1999-2002.
- [28] K. D. Duncan, "Reflections on Fault Diagnostic Expertise," in *New Technology and Human Error*, J. Rasmussen, *et al.*, Eds., Chichester, UK: John Wiley & Sons Ltd., 1987.
- [29] D. V. Reising and P. M. Sanderson, "Work domain analysis and sensors I: principles and simple example," *International Journal of Human-Computer Studies*, vol. 56, pp. 569-596, 2002.
- [30] O. St-Cyr, "Sensor Noise and Ecological Interface Design: Effects of Noise Magnitude on Operators' Performance and Control Strategies," PhD dissertation doctoral dissertation, Dept. of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada, 2006.
- [31] N. Moray and S. Haudegond, "An Absence of Vigilance Decrement in a Complex Dynamic Task," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 42, pp. 234-238, 1998.
- [32] V. de Keyser, "Structuring of Knowledge of Operators in Continuous Processes: Case Study of a Continuous Casting Plant Start-up," in *New Technology and Human Error*, J. Rasmussen, *et al.*, Eds., Chichester, UK: John Wiley & Sons Ltd., 1987, pp. 247-259.
- [33] D. McNicol, *A Primer of Signal Detection Theory*. Mahwah, N.J: Lawrence Erlbaum Associates, Inc., 2005.

- [34] M. N. Lees and J. D. Lee, "The influence of distraction and driving context driver response to imperfect collision warning systems," *Ergonomics*, vol. 50, pp. 1264-1286, 2007.
- [35] Y. Xiao and F. J. Seagull, "An analysis of problems with auditory alarms: defining the roles of alarms in process monitoring tasks," in *Proceedings of the Human Factors and Ergonomics Society 43th Annual Meeting*, 1999, pp. 256-260.
- [36] S. Guerlain and P. Bullemer, "User-initiated notification: A concept for aiding the monitoring activities or process control strategies," *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 283-287, 1996.
- [37] W. B. Rouse, "Measures of Complexity of Fault Diagnosis Tasks," *IEEE Trans. Systems, Man and Cybernetics*, vol. 9, pp. 720-727, 1979.
- [38] W. B. Rouse, "Problem Solving Performance of Maintenance Trainees in a Fault Diagnosis Task," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 21, pp. 195-203, 1979.
- [39] K. B. Bennett and J. M. Flach, "Graphical displays: implications for divided attention, focused attention, and problem solving," *Human Factors*, vol. 34, pp. 513-533, 1992.
- [40] K. J. Vicente and J. Rasmussen, "Ecological interface design: theoretical foundations," *IEEE Trans. Systems, Man and Cybernetics*, vol. 22, pp. 589-606, 1992.
- [41] C. M. Burns and J. R. Hajdukiewicz, *Ecological interface design*. Boca Raton, FL: CRC Press, 2004.
- [42] P. F. Fantoni, *et al.*, "PEANO, a Toolbox for Real-time Process Signal Validation and Estimation.," OECD Halden Reactor Project, Halden, Norway HWR-515, 1998.
- [43] G. A. Jamieson and K. J. Vicente, "Designing Effective Human-Automation-Plant Interfaces: A Control-Theoretic Perspective," *Human Factors*, vol. 47, pp. 12-34, 2005.
- [44] S. Massaiu, *et al.*, "Studying Human-Automation Interactions: Methodological Lessons Learned from the Human-Centred Autation Experiments 1997-2001," OECD Halden Reactor Project, Halden, Norway HWR-760, 2004.
- [45] A. B. M. Skjerve, *et al.*, "The influence of automation malfunctions and interface design on operator performance. The HCA-2001 experiment.," OECD Halden Reactor Project, Halden, Norway HWR-686, 2002.
- [46] M. R. Endsley, "Measurement of situation awareness in dynamic systems," *Human Factors*, vol. 37, pp. 65-84, 1995.
- [47] M. R. Endsley, "Direct Measurement of Situation Awareness: Validity and Use of SAGAT," in *Situation awareness: analysis and measurement*, M. R. Endsley and D. J. Garland, Eds., Mahwah, NJ: Lawrence Erlbaum Associates, 2000, pp. 147-174.
- [48] E. Jeannot, "Situation Awareness, Synthesis of Literature Research," EUROCONTROL, Brussels, Belgium ECC Note No. 16/00, 2000.
- [49] E. Jeannot, *et al.*, "The Development of Situation Awareness Measures in ATM Systems," Eurocontrol, Brussels, Belgium HRS/HSP-005-REP-01, 2003.
- [50] R. Breton and R. Rousseau, "Situation Awareness: A review of the Concept and its Measurement," Defence Research and Development Canada, Valcartier, QC, Canada TR 2001-220, 2001.
- [51] M. R. Endsley, "Situation Awareness: Progress and Directions," in *A cognitive approach to situation awareness : theory and application*, S. P. Banbury and S. Tremblay, Eds., Hampshire, UK: Ashgate, 2004, pp. 342-351.

- [52] G. Skraaning Jr., "Experimental Control versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments.," OECD Halden Reactor Project, Halden, Norway HPR-361, 2003.
- [53] N. Lau, *et al.*, "Situation Awareness in Monitoring Nuclear Power Plants: The Process Overview Concept and Measure," OECD Halden Reactor Project, Halden, Norway HWR-954, 2010.
- [54] R. Rousseau, *et al.*, "The role of metacognition in the relationship between objective and subjective measures of situation awareness," *Theoretical Issues in Ergonomics Science*, vol. 11, pp. 119 - 130, 2010.
- [55] B. McGuinness, "Quantitative Analysis of Situational Awareness (QUASA): applying signal detection theory to true/false probes and self-ratings," in *Proceedings of the ninth International Command and Control Research and Technology Symposium*, 2004.
- [56] E. J. Pedhazur and L. P. Schmelkin, *Measurement, design, and analysis : an integrated approach*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1991.
- [57] K. R. Murphy and C. O. Davidshofer, *Psychological testing : principles and applications*, 4th ed. Upper Saddle River, NJ, USA: Prentice Hall, 1998.
- [58] P. Kline, *Handbook of Psychological Testing*, 2nd ed. London, UK: Routledge, 2000.
- [59] J. W. Tukey, "Analyzing data: Sanctification or detective work?," *American Psychologist*, vol. 24, pp. 83-91, 1969.
- [60] D. G. Mayo, *Error and the growth of experimental knowledge* Chicago, IL, USA: University of Chicago Press, 1996.
- [61] D. Howell, *Statistical Methods for Psychology*, 5th ed. Pacific Grove, CA: Duxbury/Thomson Learning, 2002.
- [62] R. E. Kirk, *Experimental design: procedures for the behavioral sciences*, 3rd ed. Pacific Grove, CA: Brooks/Cole, 1995.
- [63] Statsoft. (2010, May). ANOVA/MANOVA. Available: <http://www.statsoft.com/textbook/anova-manova/#homogeneity>
- [64] S. E. Maxwell and H. D. Delaney, *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, 1990.
- [65] E. F. Vonesh and M. A. Schork, "Sample Sizes in the Multivariate Analysis of Repeated Measurements," *Biometrics*, vol. 42, pp. 601-610, 1986.
- [66] L. Wilkinson, "Statistical Methods in Psychology Journals," *American Psychologist*, vol. 54, pp. 594-604, 1999.
- [67] H. D. Kimmel, "Three criteria for the use of one-tailed tests," *Psychological Bulletin*, vol. 54, pp. 351-353, 1957.
- [68] E. F. Fern and M. K. B., "Effect size estimates: Issues and problems in interpretation," *Journal of Consumer Research*, vol. 23, pp. 89-105, 1996.
- [69] J. Cohen, "Eta-Squared and Partial Eta-Squared in Fixed Factor Anova Designs," *Educational and Psychological Measurement*, vol. 33, pp. 107-112, April 1, 1973 1973.
- [70] J. Cohen, "A Power Primer," *Psychological Bulletin*, vol. 112, pp. 155-159, 1992.
- [71] R. Rosenthal, *et al.*, *Contrasts and effect sizes in behavioral research : a correlational approach*. Cambridge, U.K: Cambridge University Press, 2000.
- [72] J. Sims and C. C. Wright, "The Kappa Statistics in Reliability Studies: Use, Interpretation, and Sample Size Requirements," *Physical Therapy*, vol. 85, pp. 257-268, 2005.

- [73] M. E. J. Masson and G. R. Loftus, "Using Confidence Intervals for Graphically Based Data Interpretation," *Canadian Journal of Experimental Psychology*, vol. 57, pp. 203-220, 2003.
- [74] G. R. Loftus and M. E. J. Masson, "Using confidence intervals in within-subject designs," *Psychonomic Bulletin and Review*, vol. 1, pp. 476-490, 1994.
- [75] G. Cumming and S. Finch, "Inference by Eye: Confidence Intervals and How to Read Pictures of Data," *American Psychologist*, vol. 60, pp. 170-180, 2005.
- [76] S. Belia, *et al.*, "Researchers Misunderstand Confidence Intervals and Standard Error Bars," *Psychological Methods*, vol. 10, pp. 389-396, 2005.
- [77] D. Cosineau, "Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method," *Tutorials in Quantitative Methods for Psychology*, vol. 1, pp. 42-45, 2007.
- [78] N. Lau, *et al.*, "Ecological Interface Design in the nuclear domain: An empirical evaluation of ecological displays for the secondary subsystems of a boiling water reactor plant simulator," *IEEE Trans. Nuclear Science*, vol. 55, pp. 3597-3610, 2008.
- [79] C. M. Burns, *et al.*, "Evaluation of ecological interface design for nuclear process control: situation awareness effects," *Human Factors*, vol. 50, pp. 663-679, 2008.
- [80] N. Lau, *et al.*, "The Impact of Ecological Displays on Operator Task Performance and Workload," OECD Halden Reactor Project, Halden, Norway HWR-888, 2008.
- [81] R. Welch, *et al.*, "The 2005 Ecological Interface Design Process and The Resulting Displays," in *CEL Technical Reports*. Toronto, ON, Canada: University of Toronto, 2007.
- [82] N. Lau, *et al.*, "Ecological Interface Design in the nuclear domain: An application to the secondary subsystems of a boiling water reactor plant simulator," *IEEE Trans. Nuclear Science*, vol. 55, pp. 3579-3596, 2008.
- [83] G. Skraaning Jr. and C. Nilhwing, "The Impact of Graphically Enhanced Mimic-Displays on Operator Task Performance and Situation Awareness," OECD Halden Reactor Project, Halden, Norway HWR-871, 2008.
- [84] N. Lau, *et al.*, "Metacognition in Nuclear Process Control," *Proc. of the 17th Triennial World Congress on Ergonomics [CD-Rom]*, 2009.
- [85] J. M. O'Hara and W. S. Brown, "Human-System Interface Design Review Guidelines," U.S. Nuclear Regulatory Commission, Washington, D.C., USA Tech. Rep. NUREG-700, Rev. 2, May 2002.
- [86] K. B. Bennett, *et al.*, "Visual displays," in *Handbook of Human Factors and Ergonomics*, G. Salvendy, Ed., 3rd ed Hoboken, NJ: John Wiley & Sons, Inc., 2006.
- [87] K. B. Bennett, *et al.*, "Emergent Features and Graphical Elements: Designing More Effective Configural Displays," *Human Factors*, vol. 35, pp. 71-97, 1993.
- [88] R. Welch, *et al.*, "The 2005 Ecological Interface Design Process and The Resulting Displays," OECD Halden Reactor Project, Halden, Norway HWR-847, 2007.
- [89] G. Skraaning Jr., "The Operator Performance Assessment System (OPAS)," OECD Halden Reactor Project, Halden, Norway HWR-538, 1998.
- [90] G. A. Jamieson, "Ecological Interface Design for Petrochemical Process Control: An Empirical Assessment," *IEEE Trans. Systems, Man and Cybernetics*, vol. 37, pp. 906-920, 2007.
- [91] K. S. O'Brien and D. O'Hare, "Situational awareness ability and cognitive skills training in a complex real-world task," *Ergonomics*, vol. 50, pp. 1064-1091, 2007.

- [92] A. W. Lang, *et al.*, "Using a Benchmark-Referenced Approach for Validating a Power Plant Control Room: Results of the Baseline Study," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 46, pp. 1878-1882, 2002.
- [93] M. H. R. Eitheim, *et al.*, "Staffing Strategies in Highly Automated Future Plants Results from the 2009 HAMMLAB Experiment," OECD Halden Reactor Project, Halden, Norway HWR-938, 2010.
- [94] G. Skraaning Jr, *et al.*, "Coping with Automation in Future Plants: Results from the 2009 HAMMLAB Experiment," OECD Halden Reactor Project, Halden, Norway HWR-937, 2010.
- [95] L. Hurlen, *et al.*, "Transparent automation displays - design and evaluation.," OECD Halden Reactor Project, Halden, Norway HWR-935, 2010.
- [96] C. Nihlwing and J. Kvaem, "Multi-layered Human System Interface Design.," OECD Halden Reactor Project, Halden, Norway HWR-936, in preparation.
- [97] J. Heimdal and G. Skraaning Jr, "National and Organizational Culture in an Operational Nuclear Environment - Analysis of Cultural Measures and HAMMLAB 2009 Experimental Data," OECD Halden Reactor Project, Halden, Norway HWR-956, in preparation.
- [98] N. Lau, "Situation Awareness in Process Control," PhD dissertation PhD dissertation, Dept. of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada, in preparation.
- [99] T. Karlsson, *et al.*, "OECD Halden Reactor Project: The HAMBO BWR Simulator of HAMMLAB," Institutt for Energiteknikk, Halden, Norway Tech. Rep. HWR-663, Feb 2001.
- [100] N. Lau, *et al.*, "Situation Awareness in the Diagnosis of Process Disturbances - the Scenario Understanding Concept and Measures," OECD Halden Reactor Project, Halden, Norway HWR-970, 2010.
- [101] D. G. Jones and M. R. Endsley, "Use of real-time probes for measuring Situation Awareness," *The International Journal of Aviation Psychology*, vol. 14, p. 4, 2004.
- [102] J. Cohen, "The coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37-46, 1960.
- [103] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, pp. 159-174, 1977.
- [104] R. D. Luce, R. R. Bush, E. Galanter, "Detection and recognition of human observers," in *Handbook of mathematical psychology*. Vol. 1, Anonymous, Ed., New York, USA: Wiley, 1963.
- [105] R. D. Luce, "The Choice Axiom after Twenty Years," *Journal of Mathematical Psychology*, vol. 15, pp. 215-233, 1977.
- [106] P. Kline, *Intelligence: The Psychometric View*. London, UK: Routledge, 1991.

APPENDIX A. BIAS IN APPLYING INTRA-CLASS CORRELATION TO ASSESS INTER-RATER RELIABILITY OF THE PROCESS OVERVIEW MEASURE

To apply intra-class correlations to assess inter-rater reliability across experts for the Process Overview Measure, scores must first be calculated by comparing the responses of the operators to the process experts/judges. This conversion could lead to errors that do not occur in the Kappa procedures. Table 23 illustrates this problem. For Item 3 in the table, the scores of Judge A and B would match even though they disagree with each other.

Table 23: Illustration of the problem in applying intra-class correlation to assess inter-rater agreement of the Process Overview Measure.

Item	Operator	Judge A		Judge B		Note
	Response	Response	Score	Response	Score	
1	Increased	Increased	1	Increased	1	A & B judgements and scores of A and B match.
2	Increased	Increased	1	Stayed the same	0	A & B judgements and scores do not match.
3	Increased	Stayed the same	0	Decreased	0	A & B judgements do not match but scores do match.

APPENDIX B. FREQUENCY DISTRIBUTIONS BETWEEN THE THREE RATERS WITH RESPECT TO THE PARTICIPANT RESPONSES IN STUDY 2

Table 24: Frequency distribution of responses to Process Overview queries for Raters A and B in Study 3.

Rater A's Responses with respect to Participant's Responses	Rater B's Responses respect to Participant's Responses					Total
	Responded UPWARD relative to participant's by 2 options	Responded UPWARD relative to participant's by 1 option	AGREE	Responded DOWNWARD relative to participant's by 1 option	Responded DOWNWARD relative to participant's by 2 options	
Responded UPWARD relative to participant's by 2 options (increased vs decreased)	5	3	3			11
Responded UPWARD relative to participant's by 1 option	4	36	13	3		56
AGREE	1	19	123	20	2	165
Responded DOWNWARD relative to participant's by 1 option		1	19	29	5	54
Responded DOWNWARD relative to participant's by 2 options (decreased vs increased)				1	3	4
Total	10	59	158	53	10	290

Table 25: Frequency distribution of responses to Process Overview queries for Raters A and C in Study 3.

Rater A's Responses with respect to Participant's Responses	Rater C's Responses respect to Participant's Responses					Total
	Responded UPWARD relative to participant's by 2 options	Responded UPWARD relative to participant's by 1 option	AGREE	Responded DOWNWARD relative to participant's by 1 option	Responded DOWNWARD relative to participant's by 2 options	
Responded UPWARD relative to participant's by 2 options (increased vs decreased)	6	3	2			11
Responded UPWARD relative to participant's by 1 option	1	40	12	3		56
AGREE	1	9	148	7		165
Responded DOWNWARD relative to participant's by 1 option			23	31		54
Responded DOWNWARD relative to participant's by 2 options (decreased vs increased)			1	1	2	4
Total	8	52	186	42	2	290

Table 26: Frequency distribution of responses to Process Overview queries for Raters B and C in Study 3.

Rater B's Responses with respect to Participant's Responses	Rater C's Responses respect to Participant's Responses					Total
	Responded UPWARD relative to participant's by 2 options	Responded UPWARD relative to participant's by 1 option	AGREE	Responded DOWNWARD relative to participant's by 1 option	Responded DOWNWARD relative to participant's by 2 options	
Responded UPWARD relative to participant's by 2 options (increased vs decreased)	3	7				10
Responded UPWARD relative to participant's by 1 option	4	27	23	5		59
AGREE	1	16	127	14		158
Responded DOWNWARD relative to participant's by 1 option		2	32	19		53
Responded DOWNWARD relative to participant's by 2 options (decreased vs increased)			4	4	2	10
Total	8	52	186	42	2	290

APPENDIX C KAPPA STATISTICS WITHOUT REFERENCE TO PARTICIPANT RESPONSES

For the Process Overview measure, two sets of Kappa are calculated to assess inter-rater reliability. The first set of Kappas investigates the agreement of their response between the raters without any reference to the responses of the participants in Study 2.

Table 27: Frequency distribution of responses to Process Overview queries for Rater A and B in Study 3.

Rater A's Responses	Rater B's Responses			Grand Total
	Increased	Stayed the same	Decreased	
Decreased	26	5	2	33
Stayed the same	20	116	38	174
Increased	8	21	54	83
Grand Total	54	142	94	290

Table 28: Frequency distribution of responses to Process Overview queries for Rater A and C in Study 3.

Rater A's Responses	Rater C's Responses			Grand Total
	Increased	Stayed the same	Decreased	
Decreased	25	6	2	33
Stayed the same	4	142	28	174
Increased	5	18	60	83
Grand Total	34	166	90	290

Table 29: Frequency distribution of responses to Process Overview queries for Rater B and C in Study 3.

Rater B's Responses	Rater C's Responses			Grand Total
	Increased	Stayed the same	Decreased	
Decreased	22	25	7	54
Stayed the same	7	104	31	142
Increased	5	37	52	94
Grand Total	34	166	90	290

Table 30: Kappas between raters *without* references to responses of participants.

Raters	B		C	
	κ	σ	κ	σ
A	0.452	0.046	0.608	0.044
B	-		0.353	0.048



CEL TECHNICAL REPORT SERIES

CEL 93-01	<p>“Egg-sucking, Mousetraps, and the Tower of Babel: Making Human Factors Guidance More Accessible to Designers”</p> <ul style="list-style-type: none"> • Kim J. Vicente, Catherine M. Burns, & William S. Pawlak 	CEL-SP	<p>“Strategic Plan”</p> <ul style="list-style-type: none"> • Cognitive Engineering Laboratory
		CEL-LP	<p>“Cognitive Engineering Laboratory Profile”</p> <ul style="list-style-type: none"> • Cognitive Engineering Laboratory
CEL 93-02	<p>“Effects of Expertise on Reasoning Trajectories in an Abstraction Hierarchy: Fault Diagnosis in a Process Control System”</p> <ul style="list-style-type: none"> • Klaus Christoffersen, Alex Pereklita, & Kim J. Vicente 	CEL 95-04	<p>“A Field Study of Operator Cognitive Monitoring at Pickering Nuclear Generating Station-B”</p> <ul style="list-style-type: none"> • Kim J. Vicente & Catherine M. Burns
CEL 94-01	<p>“Cognitive ‘Dipsticks’: Knowledge Elicitation Techniques for Cognitive Engineering Research”</p> <ul style="list-style-type: none"> • Klaus Christoffersen, Christopher N. Hunter, & Kim J. Vicente 	CEL 95-05	<p>“An Empirical Investigation of the Effects of Training and Interface Design on the Control of Complex Systems”</p> <ul style="list-style-type: none"> • Christopher N. Hunter
CEL 94-02	<p>“Muddling Through Wicked Problems: Exploring the Role of Human Factors Information in Design”</p> <ul style="list-style-type: none"> • Catherine M. Burns 	CEL 95-06	<p>“Applying Human Factors to the Design of Medical Equipment: Patient-Controlled Analgesia”</p> <ul style="list-style-type: none"> • Laura Lin, Racquel Isla, Karine Doniz, Heather Harkness, Kim J. Vicente, & D. John Doyle
CEL 94-03	<p>“Cognitive Work Analysis for the DURESS II System”</p> <ul style="list-style-type: none"> • Kim J. Vicente & William S. Pawlak 	CEL 95-07	<p>“An Experimental Evaluation of Transparent Menu Usage”</p> <ul style="list-style-type: none"> • Beverly L. Harrison & Kim J. Vicente
CEL 94-04	<p>“Inducing Effective Control Strategies Through Ecological Interface Design”</p> <ul style="list-style-type: none"> • William S. Pawlak 	CEL 95-08	<p>“Research on Factors Influencing Human Cognitive Behaviour (II)”</p> <ul style="list-style-type: none"> • Christopher N. Hunter, Michael E. Janzen, & Kim J. Vicente
CEL 94-05	<p>“Research on Factors Influencing Human Cognitive Behaviour (I)”</p> <ul style="list-style-type: none"> • Klaus Christoffersen, Christopher N. Hunter, & Kim J. Vicente 	CEL 95-09	<p>“To the Beat of a Different Drummer: The Role of Individual Differences in Ecological Interface Design”</p> <ul style="list-style-type: none"> • Dianne Howie
CEL 94-06	<p>“Ecological Interfaces for Complex Industrial Plants”</p> <ul style="list-style-type: none"> • Nick Dinadis & Kim J. Vicente 	CEL 95-10	<p>“Emergent Features and Temporal Information: Shall the Twain Ever Meet?”</p> <ul style="list-style-type: none"> • JoAnne H. Wang
CEL 94-07	<p>“Evaluation of a Display Design Space: Transparent Layered User Interfaces”</p> <ul style="list-style-type: none"> • Beverly L. Harrison, Hiroshi Ishii, Kim J. Vicente, & Bill Buxton 	CEL 95-11	<p>“Physical and Functional Displays in Process Supervision and Control”</p> <ul style="list-style-type: none"> • Catherine M. Burns & Kim J. Vicente
CEL 94-08	<p>“Designing and Evaluating Semi-Transparent ‘Silk’ User Interface Objects: Supporting Focused and Divided Attention”</p> <ul style="list-style-type: none"> • Beverly L. Harrison, Shumin Zhai, Kim J. Vicente, & Bill Buxton 	CEL 96-01	<p>“Shaping Expertise Through Ecological Interface Design: Strategies, Metacognition, and Individual Differences”</p> <ul style="list-style-type: none"> • Dianne E. Howie
CEL 95-01	<p>“An Ecological Theory of Expertise Effects in Memory Recall”</p> <ul style="list-style-type: none"> • Kim J. Vicente & JoAnne H. Wang 	CEL 96-02	<p>“Skill, Participation, and Competence: Implications of Ecological Interface Design for Working Life”</p> <ul style="list-style-type: none"> • Peter Benda, Giuseppe Cioffi, & Kim J. Vicente
		CEL 96-03	<p>“Practical Problem Solving in a Design Microworld: An Exploratory Study”</p> <ul style="list-style-type: none"> • Klaus Christoffersen

CEL 96-04	<p>“Review of Alarm Systems for Nuclear Power Plants”</p> <ul style="list-style-type: none"> • Kim J. Vicente 	CEL 98-01	<p>“Applying Human Factors Engineering to Medical Device Design: An Empirical Evaluation of Patient-Controlled Analgesia Machine Interfaces”</p> <ul style="list-style-type: none"> • Laura Lin
CEL 96-05	<p>“DURESS II User’s Manual: A Thermal-hydraulic Process Simulator for Research and Teaching”</p> <ul style="list-style-type: none"> • Lisa C. Orchanian, Thomas P. Smahel, Dianne E. Howie, & Kim J. Vicente 	CEL 98-02	<p>“Building an Ecological Foundation for Experimental Psychology: Beyond the Lens Model and Direct Perception”</p> <ul style="list-style-type: none"> • Kim J. Vicente
CEL 96-06	<p>“Research on Factors Influencing Human Cognitive Behaviour (III)”</p> <ul style="list-style-type: none"> • Dianne E. Howie, Michael E. Janzen, & Kim J. Vicente 	CEL 98-03	<p>“Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-based Work”</p> <ul style="list-style-type: none"> • Kim J. Vicente
CEL 96-07	<p>“Application of Ecological Interface Design to Aviation”</p> <ul style="list-style-type: none"> • Nick Dinadis & Kim J. Vicente 	CEL 98-04	<p>“Ecological Interface Design for Petrochemical Processing Applications”</p> <ul style="list-style-type: none"> • Greg. A. Jamieson & Kim J. Vicente
CEL 96-08	<p>“Distributed Cognition Demands a Second Metaphor for Cognitive Science”</p> <ul style="list-style-type: none"> • Kim J. Vicente 	CEL 98-05	<p>“The Effects of Spatial and Temporal Proximity of Means-end Information in Ecological Display Design for an Industrial Simulation”</p> <ul style="list-style-type: none"> • Catherine M. Burns
CEL 96-09	<p>“An Experimental Evaluation of Functional Displays in Process Supervision and Control”</p> <ul style="list-style-type: none"> • Catherine M. Burns and Kim J. Vicente 	CEL 98-06	<p>“Research on Characteristics of Long-term Adaptation (II)”</p> <ul style="list-style-type: none"> • Xinyao Yu, Gerard L. Torenvliet, & Kim J. Vicente
CEL 96-10	<p>“The Design and Evaluation of Transparent User Interfaces: From Theory to Practice”</p> <ul style="list-style-type: none"> • Beverly L. Harrison 	CEL 98-07	<p>"Integrated Abstraction Hierarchy and Plan-Goal Graph Model for the DURESS II System: A Test Case for Unified System- and Task-based Modeling and Interface Design"</p> <ul style="list-style-type: none"> • Christopher A. Miller & Kim J. Vicente
CEL 97-01	<p>“Cognitive Functioning of Control Room Operators: Final Phase”</p> <ul style="list-style-type: none"> • Kim J. Vicente, Randall J. Mumaw, & Emilie M. Roth 	CEL 98-08	<p>"Comparative Analysis of Display Requirements Generated via Task-Based and Work Domain-based Analyses: A Test Case Using DURESS II"</p> <ul style="list-style-type: none"> • Christopher A. Miller & Kim J. Vicente
CEL 97-02	<p>"Applying Human Factors Engineering to Medical Device Design: An Empirical Evaluation of Two Patient-Controlled Analgesia Machine Interfaces"</p> <ul style="list-style-type: none"> • Laura Lin 	CEL 98-09	<p>"Abstraction Decomposition Space Analysis for NOVA's E1 Acetylene Hydrogenation Reactor"</p> <ul style="list-style-type: none"> • Christopher A. Miller & Kim J. Vicente
CEL 97-03	<p>“ADAPT User’s Manual: A Data Analysis Tool for Human Performance Evaluation in Dynamic Systems”</p> <ul style="list-style-type: none"> • Xinyao Yu, Farzad S. Khan, Elfreda Lau, Kim J. Vicente, & Michael W. Carter 	CEL 99-01	<p>“Development of an Analytical Framework and Measurement Tools to Assess Adaptive Performance of Individual and Teams in Military Work Domains”</p> <ul style="list-style-type: none"> • John R. Hajdukiewicz, Kim J. Vicente, & Robert G. Eggleston
CEL 97-04	<p>“Research on the Characteristics of Long-Term Adaptation”</p> <ul style="list-style-type: none"> • Xinyao Yu, Renée Chow, Greg A. Jamieson, Rasha Khayat, Elfreda Lau, Gerard Torenvliet, Kim J. Vicente, & Michael W. Carter 	CEL 99-02	<p>“Applying Perceptual Control Theory and Ecological Interface Design to the Control Display Unit”</p> <ul style="list-style-type: none"> • Sandra Chéry
CEL 97-05	<p>“A Comprehensive Experimental Evaluation of Functional Displays in Process Supervision and Control”</p> <ul style="list-style-type: none"> • Catherine M. Burns and Kim J. Vicente 	CEL 99-03	<p>“Research on the Characteristics of Long-Term Adaptation (III)”</p> <ul style="list-style-type: none"> • Gerard L. Torenvliet & Kim J. Vicente
		CEL 99-04	<p>"Comparative Analysis of Display Requirements Generated via Task-Based and Work Domain-based Analyses in a Real World Domain: NOVA's Acetylene Hydrogenation Reactor"</p>

CEL 99-05	<ul style="list-style-type: none"> • Christopher A. Miller & Kim J. Vicente <p>“A Cognitive Engineering Approach for Measuring Adaptive Behavior”</p> <ul style="list-style-type: none"> • John R. Hajdukiewicz & Kim J. Vicente 	CEL 07-03	<p>“Factors Influencing the Reliance on Combat Identification Systems”</p> <ul style="list-style-type: none"> • Lu Wang
CEL 00-01	<p>“Differences Between the Eye-fixation Patterns of Novice and Expert Operators of the DURESS II Physical Interface”</p> <ul style="list-style-type: none"> • Madhava Enros & Kim J. Vicente 	CEL 07-04	<p>“Applying a Formative Ecological Framework to Simulator Design Challenges”</p> <ul style="list-style-type: none"> • Antony Hilliard
CEL 00-02	<p>“If Technology Doesn’t Work for People, then It Doesn’t Work”</p> <ul style="list-style-type: none"> • Kim J. Vicente 	CEL 08-01	<p>“Cognitive Work Analysis of the City of Toronto Winter Maintenance Program”</p> <ul style="list-style-type: none"> • Laura Thompson, Antony Hilliard, & Cam Ngo
CEL 01-01	<p>“A Field Study of Collaborative Work in Network Management: Implications for Interface Design and Evaluation”</p> <ul style="list-style-type: none"> • Renée Chow & Kim J. Vicente 	CEL 08-02	<p>“The Impact of Ecological Displays on Operator Task Performance and Workload”</p> <ul style="list-style-type: none"> • Nathan Lau, Gyrd Skraaning jr., Greg A. Jamieson, & Catherine M. Burns
CEL 01-02	<p>“A Prototype Ecological Interface for a Simulated Petrochemical Process”</p> <ul style="list-style-type: none"> • Greg A. Jamieson & Wayne H. Ho 	CEL 11-01	<p>“Situation Awareness in Monitoring Nuclear Power Plants – The Process Overview Concept and Measure”</p> <ul style="list-style-type: none"> • Nathan Lau, Gyrd Skraaning jr., Maren H. R. Eitheim, Tommy Karlsson, Christer Nihlwing, & Greg A. Jamieson
CEL 01-03	<p>“EID Design Rationale Project: Case Study Report”</p> <ul style="list-style-type: none"> • Greg A. Jamieson, Dal Vernon C. Reising & John Hajdukiewicz 		
CEL 02-01	<p>“Ecological Interface Design for Petrochemical Process Control: Integrating Task-and System-Based Approaches”</p> <ul style="list-style-type: none"> • Greg A. Jamieson 		
CEL 06-01	<p>“Canada Foundation for Innovation (CFI) Emerson DeltaV / MiMiC Industrial Process Control Simulator”</p> <ul style="list-style-type: none"> • Antony Hilliard & Laura Thompson 		
CEL 06-02	<p>“Developing Human-Machine Interfaces to Support Monitoring of UAV Automation”</p> <ul style="list-style-type: none"> • Greg A. Jamieson, Lu Wang, Jamy Li 		
CEL 06-03	<p>“Sensor Noise and Ecological Interface Design: Effects of Noise Magnitude on Operators’ Performance and Control Strategies”</p> <ul style="list-style-type: none"> • Olivier St-Cyr 		
CEL 07-01	<p>“The 2005 Ecological Interface Design Process and the Resulting Displays”</p> <ul style="list-style-type: none"> • Robin Welch, Alf Ove Braseth, Christer Nihlwing, Gyrd Skraaning, Arild Teigen, Øystein Veland, Nathan Lau, Greg A. Jamieson, Jordana Kwok, Catherine M. Burns 		
CEL 07-02	<p>“The Ecological Interface Design experiment (2005)”</p> <ul style="list-style-type: none"> • Gyrd Skraaning, Nathan Lau, Robin Welch, Christer Nihlwing, Gisle Andresen, Liv Hanne Brevig, Øystein Veland, Greg A. Jamieson, Catherine M. Burns, Jordanna Kwok 		