Human Aspects of Nuclear Safety
NEA/CSNI/R(2016)17
June 2017
www.oecd-nea.org

Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications

Proceedings of the Expert Workshop
Charlotte, United States
19–21 February 2015

OECD
BETTER POLICIES FOR BETTER LIVES

NEA
NUCLEAR ENERGY AGENCY

For Official Use **NEA/CSNI/R(2016)17**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **02-Jun-2017**

English - Or. English

NUCLEAR ENERGY AGENCY
COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS

NEA/CSNI/R(2016)17
For Official Use

Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications

Proceedings of the Expert Workshop
Charlotte, United States
19-21 February 2015

JT03415205

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

English - Or. English

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 35 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Republic of Korea, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

NUCLEAR ENERGY AGENCY

The OECD Nuclear Energy Agency (NEA) was established on 1 February 1958. Current NEA membership consists of 31 countries: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, the Netherlands, Norway, Poland, Portugal, the Republic of Korea, the Russian Federation, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission also takes part in the work of the Agency.

The mission of the NEA is:

- to assist its member countries in maintaining and further developing, through international co-operation, the scientific, technological and legal bases required for a safe, environmentally friendly and economical use of nuclear energy for peaceful purposes;
- to provide authoritative assessments and to forge common understandings on key issues, as input to government decisions on nuclear energy policy and to broader OECD policy analyses in areas such as energy and sustainable development.

Specific areas of competence of the NEA include the safety and regulation of nuclear activities, radioactive waste management, radiological protection, nuclear science, economic and technical analyses of the nuclear fuel cycle, nuclear law and liability, and public information. The NEA Data Bank provides nuclear data and computer program services for participating countries.

This document, as well as any map or data included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Corrigenda to OECD publications may be found online at: *www.oecd.org/publishing/corrigenda*.

© OECD 2017

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of the OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to *rights@oecd.org.* Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at *info@copyright.com* or the Centre français d'exploitation du droit de copie (CFC) *contact@cfcopies.com.*

COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS

The Nuclear Energy Agency (NEA) Committee on the Safety of Nuclear Installations (CSNI) is an international committee made up of senior scientists and engineers with broad responsibilities for safety technology and research programmes, as well as representatives from regulatory authorities. It was created in 1973 to develop and co-ordinate the activities of the NEA concerning the technical aspects of the design, construction and operation of nuclear installations insofar as they affect the safety of such installations.

The committee's purpose is to foster international co-operation in nuclear safety among NEA member countries. The main tasks of the CSNI are to exchange technical information and to promote collaboration between research, development, engineering and regulatory organisations; to review operating experience and the state of knowledge on selected topics of nuclear safety technology and safety assessment; to initiate and conduct programmes to overcome discrepancies, develop improvements and reach consensus on technical issues; and to promote the co-ordination of work that serves to maintain competence in nuclear safety matters, including the establishment of joint undertakings.

The priority of the CSNI is on the safety of nuclear installations and the design and construction of new reactors and installations. For advanced reactor designs, the committee provides a forum for improving safety-related knowledge and a vehicle for joint research.

In implementing its programme, the CSNI establishes co-operative mechanisms with the NEA Committee on Nuclear Regulatory Activities (CNRA), which is responsible for issues concerning the regulation, licensing and inspection of nuclear installations with regard to safety. It also co-operates with other NEA Standing Technical Committees, as well as with key international organisations such as the International Atomic Energy Agency (IAEA), on matters of common interest.

ACKNOWLEDGEMENTS

The planning, organisation, implementation, and reporting of The Experts' Workshop on Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications was a collaborative effort by many individuals who contributed substantively to the success of the workshop and the report that follows. The Working Group on Human and Organisational Factors (WGHOF) task group that was responsible for all aspects of this workshop comprised the following individuals: Per Øivind Braarud, Halden Reactor Project (HRP), Amy D'Agostino, United States Nuclear Regulatory Commission (NRC), Cecilia De la Garza of Électricité de France (EdF), David Desaulniers (NRC), Stephen Fleger (NRC), Robert Hall (REH Technologies), Jari Laarni (VTT), Dina Notte (ERGODIN for AREVA), Daniel Tasset (Institut de Radioprotection et de Sûreté Nucléaire), and Cyril Rivere (AREVA), and the US Nuclear Regulatory Commission served as the lead organisation. In addition, the task group was fortunate to have the support of Leena Norros (VTT), John O'Hara, Brookhaven National Laboratory (BNL for NRC), and Douglas Hill (Hill Engineering Consultants for NuScale). As a former representative to the WGHOF, Leena Norros was the original advocate for the WGHOF addressing the challenges of control room validation. She was instrumental in the early workshop planning and contributed substantively to the workshop's technical programme as a challenge speaker. John O'Hara provided the workshop's opening address and was the principal author of the post-workshop analysis section of this report. Douglas Hill served as an external reviewer for this report and his efforts helped ensure that the report is clear and that our conclusions and recommendations appropriately reflect the expert input we gathered during the workshop. Finally, we note that this workshop would not have been possible without the many experts (see Appendix A) who donated their time and freely shared their expertise as workshop participants. The WGHOF extends its appreciation to all of these individuals and their organisations for their generosity and dedication in contributing their time and expertise to this effort.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Validation is a critical step in the design of new control rooms and major control room modifications. It is important, not just to designers, but to regulators as well. Integrated system validation (ISV) is a method of validation in which a design is evaluated using performance-based tests (e.g. assessments based upon the physical, cognitive, and functional performance of operating crews under simulated normal, off-normal, and accident conditions) to determine whether the integrated system design (i.e. hardware, software, environmental, and personnel elements) meets performance requirements and supports safe operation of the plant. Although ISV can provide important insights into how well system elements work together to support overall system performance, the complexity of nuclear power plants, the diversity of operational conditions, and practical resource constraints collectively challenge our ability to achieve reasonable confidence. As a result, achieving reasonable confidence in ISV test results is a matter of particular concern for ISV tests of nuclear power plant (NPP) main control room (MCR) designs. This report documents a workshop focused on addressing this important topic.

From 19-21 February 2015, the Nuclear Energy Agency (NEA) Committee for Safety of Nuclear Installations' Working Group on Human and Organisational Factors (WGHOF) hosted a 3-day international experts' workshop on the validation of nuclear power plant main control room system designs and modifications. The theme of the workshop was "Establishing Reasonable Confidence in the Human Factors Validation of Main Control Room Systems of Nuclear Power Plants" and it had three high level objectives:

- Critically examine preliminary and final (integrated) validation activities to better understand their strengths, limitations, and potential inter-relationships with respect to the technical and practical considerations for achieving reasonable confidence in nuclear power plant control room designs and modifications.
- Identify recommended practices, potential solutions, and available technical bases for addressing current limitations.
- Identify priority areas for future research.

The workshop participants included 28 individuals invited by the workshop task group based on their recognised expertise in planning or conducting control room validations or validations of complex systems. Their principal task was to answer four challenge questions:

1. What are the critical considerations in defining validation objectives and how do these impact achieving reasonable confidence?
2. What methods, approaches, resources, or rationales might be used for deriving performance requirements, selecting measures, and establishing acceptance criteria so as to support reasonable confidence?
3. What methods might be used to develop scenarios that maximise the amount and relevance of information in support of the validation conclusions and achieving reasonable confidence?

4. How should the validation results be aggregated and analysed to determine the final validation conclusions? Are inferential statistics meaningful in the context of MCR validation? If not, why not? What substitutes might be proposed as alternatives to traditional statistical modelling approaches?

The workshop task group provided these questions to the workshop participants several months in advance of the workshop to provide the experts an opportunity to consider their responses and develop white papers addressing these challenge areas. These white papers were posted to a website open to all workshop participants to facilitate sharing of views and concepts in advance of the meeting. During the workshop, each question was addressed in a topical session comprising challenge presentations followed by small-group brain-storming sessions. The workshop concluded with panel sessions reviewing the key concepts and recommendations for future practices and research in control room validation.

We prioritised the elicitation of recommendations and, considering time constraints, did not endeavor to vet them during the workshop in accordance with guidance development procedures, such as subjecting them to a consensus process or peer review. Nevertheless, the recommended practices proposed by the workshop participants and documented in this report (see Section 3 for details) will be valuable inputs to future guidance development efforts. Examples of these recommendations include:

- Validate modifications even if not related to safety systems because they can change operator tasks and human-system interfaces (HSIs) and degrade performance.
- Identify a representative test set to rigorously challenge the system.
- Pass or fail the design on safety criteria, which are objective, conservative, and independently established and confirmed.
- Establish independence of the validation team from the design team.
- Use pilot tests to ensure the simulator and the methodology perform acceptably before validation trials begin.
- Use a multidisciplinary approach to establishing performance criteria and the interpretation of validation data.

In addition to offering specific practices for the validation of control rooms and control room modifications, the participants addressed the broader themes and concepts of the workshop, including the notion of preliminary validation and its relationship to final, or integrated system, validation. Although the participants generally perceived a need to further define or clarify the concept of preliminary validation, there was also a common view that validation efforts early in the design process could be of value in establishing reasonable confidence in the final control room design or modification and that additional work to develop and formalise guidance for such practices could be fruitful.

With respect to the issue of reasonable confidence, the workshop papers and discussions indicated that establishing confidence in the validation results and determining what is reasonable (reasonableness) were separate but inter-related matters.

The workshop results support the conclusion that although existing validation guidelines are valuable in achieving confidence in validations, a considerable amount of work remains to be done. The available guidance documents have limitations in both their level of detail (e.g. there is insufficient guidance to support the identification of specific scenarios to be used to support generalisation of results) and coverage (e.g. there is a lack of guidance addressing validation conclusions). On the positive side, the workshop participants offered a range of methodological considerations and proposals that may prove useful in addressing these limitations. Addressing these limitations will provide a clearer picture of what is needed

for results to be reasonable. Providing promising pathways to address these issues was the workshop's most significant contribution. Toward that end, we concluded that the following research and development areas would be among the most promising pathways for establishing reasonable confidence in NPP control room designs and modifications:

- multi-stage validation;
- unique considerations for validating plant modifications;
- validation testbeds and simulation capabilities;
- scenario design and selection;
- performance measures and acceptance criteria;
- methods and guidance for arriving at validation conclusions.

Another important conclusion to be drawn from the workshop is that the issue of reasonableness is not strictly a validation methodology consideration. There are many broad challenges that have to be addressed, such as the need to address cultural and operational differences when conducting validation for designs that will be used in different parts of the world; and the challenges posed by different regulatory approaches to the review and evaluation of design validations.

Defining and achieving reasonable confidence is a complex challenge for many reasons; not the least of which is that validation is a test of strength rather than a test of truth. This challenge, however, is not intractable. The workshop results provide good reason to believe that defining and achieving reasonable confidence in the validation of nuclear power plant control room designs can be substantially addressed by establishing guidance that is sufficiently detailed to support the development and execution of validation processes and that is agreed upon by the community of stakeholders as being technically sound and meeting standards of reasonableness. The workshop resulted in a better understanding as to what it will take to define and establish reasonable confidence.

LIST OF ABBREVIATIONS AND ACRONYMS

BDB	Beyond-design basis
BNL	Brookhaven National Laboratory
CNRA	NEA Committee on Nuclear Regulatory Activities
CR	Control room
CSNI	NEA Committee on the Safety of Nuclear Installations
EdF	*Électricité de France*
EHPG	Enlarged Halden Programme Group
EOS	Emergency operating system
FICRV	Final integrated control room validation
HA	Human actions
HF	Human factors
HFE	Human factors engineering
HRA	Human reliability analysis
HRP	Halden Reactor Project
HSI	Human-system interface
IAEA	International Atomic Energy Agency
I&C	Instrumentation and control
ISV	Integrated system validation
MCL	Mission capability level
MCR	Main control room
MMI	Man-machine interface
NEA	Nuclear Energy Agency

NPP	Nuclear power plant
NRC	Nuclear Regulatory Commission
OECD	Organisation for Economic Co-operation and Development
ONR	Office of Nuclear Regulation
PRA	Probabilistic reliability assessment
PSA	Probabilistic safety assessment
PV	Preliminary validation
PWRs	Pressurised water reactors
R&D	Research and development
RIHAs	Risk-important human actions
SA	Situation awareness
SEM	Structural equation modelling
V&V	Verification and validation
VTT	VTT Technical Research Centre of Finland Ltd.
WGHOF	Working Group on Human and Organisational Factors

1. MOTIVATION FOR AND DEVELOPMENT OF THE WORKSHOP

Design validation is a critical step in the design of new control rooms and major control room modifications. It is important, not just to designers and plant operators, but to regulators as well. The United States Nuclear Regulatory Commission (NRC), for example, evaluates integrated system validation (ISV) as part of their human factors engineering safety review process. ISV is an evaluation of the control room design using performance-based tests (e.g. assessments based upon the physical, cognitive, and functional performance of operating crews under simulated normal, off-normal, and accident conditions) to determine whether an integrated system design (i.e. hardware, software, environmental, and personnel elements) meets performance requirements and supports safe operation of the plant. Experience has shown that ISV presents significant technical challenges for both the conduct of the tests and the interpretation of the results.

ISV tests can provide critical information about the safety of a design that goes well beyond that which can be ascertained by a desktop analysis of the design or through testing of individual design elements. However, the extent to which one can draw inferences about the real-world performance of a design through validation tests, including ISV, is dependent upon a host of issues concerning the representativeness of the tests, not the least of which is the sampling of scenarios and operational conditions under which the system is tested. Achieving reasonable confidence in ISV test results is a matter of particular concern for ISV tests of nuclear power plant (NPP) main control room (MCR) designs.

A byproduct of the complexity of NPPs is the large number of conditions and scenarios that can present themselves during the life cycle of a plant and therefore, theoretically, should be tested. At the same time, the potential for significant adverse impacts on public safety and on the environment demands a strict and limited tolerance for failure. Yet tests of the integrated system involving trained crews of NPP operators and other personnel in full-scope simulations of MCR designs are highly resource intensive, with the availability of trained operating crews and simulator time constrained by practical limitations. These conditions challenge our ability to conduct the number of test trials that can support valid statistical analysis of the test data to achieve the desired levels of confidence in the test results; levels typically associated with rigorous analysis of large sample data.

The importance and challenge of achieving reasonable confidence in ISV tests of nuclear power plant control room designs has been previously identified and discussed as a matter of high priority for the nuclear power industry. In April 2009, the Halden Reactor Project (HRP) hosted a workshop entitled *Status of Current Approaches and R&D Needs for Integrated System Validation* (Braarud et. al, 2010). Workshop participants identified a range of ISV issues, including many which bear upon the matter of achieving reasonable confidence. HRP has been actively conducting work addressing some of these issues. The NRC also conducted assessments to identify high-priority research needs. In two separate studies, ISV was evaluated by subject matter experts as a high-priority topic (O'Hara et. al, 2008; Molino and O'Hara, 2010). Accordingly, the NRC has sponsored a research activity to document the state of the art in ISV, including practices in other industries, as well as alternatives to ISV testing (e.g. stepwise validation, equivalence testing) (O'Hara and Higgins, 2015). Thus while ISV has been identified as an important activity from both design and regulatory perspectives, experience gained from performing control room validation has found these evaluations to be challenging.

In March 2013, more than a dozen human and organisational factors professionals representing regulatory agencies, vendors, technical support organisations, and research laboratories gathered during the 2013 Enlarged Halden Programme Group (EHPG) Meeting to discuss recent ISV research, the challenges associated with achieving reasonable confidence in ISV results, and the potential for future information sharing and collaboration. Among the meeting participants were several representatives to the Nuclear Energy Agency (NEA), Committee on the Safety of Nuclear Installations (CSNI), Working Group on Human and Organisational Factors (WGHOF). Those attending generally agreed that given the complexity of nuclear power plants and the extensive resources required to conduct integrated testing, achieving reasonable confidence in validation test results was a matter of both technical and practical concern. The group agreed that there was a need to both examine and expand present day thinking about control room validation so that current methods could be improved and new methods could be identified to provide practical and technically sound means for achieving reasonable confidence. The discussions at the 2013 EHPG meeting gave birth to the notion of an experts workshop focused on best practices and research opportunities for further enhancing ISV and led to a consensus that WGHOF was a fitting organisation to lead such an initiative. By July 2013 WGHOF had requested and received authorisation from CSNI to conduct such a workshop and the planning began in earnest the following month.

Over the next 12 months, numerous task force meetings produced the broad framework of the meeting (i.e. solution oriented brain storming sessions focused on specific technical challenges) and then the detailed plans. These details included solicitation of the leading experts in validation to provide stimulating presentations, submit thought provoking white papers, and to engage in spirited discussions on each of the 4 challenge questions put forward. The result of these efforts was a 3-day international experts' workshop conducted in Charlotte, North Carolina from 19-21 February 2015.

This report provides a summary of the workshop. Section 2 describes the workshop participants and the approach to presenting information on key issues, fostering discussions among participants, and identifying promising approaches for improving ISV methods and practices. Section 2 also provides a summary of three plenary talks designed to set the stage for the workshop. Section 3 describes the challenge questions used to structure the workshop and summarises the discussions of those questions by workshop participants, including their recommendations for validation practices and future research. Section 4 provides our post-workshop analysis of the information presented and discussions of the issues. The report also contains appendices providing participant bios and contact information; white papers, and the presentation abstracts/texts and graphics.

2. THE WORKSHOP

2.1 Theme, format and focus

The theme of the workshop was "Establishing Reasonable Confidence in the Human Factors Validation of Main Control Room Systems of Nuclear Power Plants". This was guided by three specific objectives:

- Critically examine preliminary and final (integrated) validation activities to better understand their strengths, limitations, and potential inter-relationships with respect to the technical and practical considerations for achieving reasonable confidence in nuclear power plant control room designs and modifications.
- Identify recommended practices, potential solutions, and available technical bases for addressing current limitations.
- Identify priority areas for future research.

The workshop's participants included 28 individuals invited by the workshop task group based on their recognised expertise in planning or conducting control room validations or validations of complex systems. Biographical sketches of speakers are provided in Appendix A.

Prior to the workshop, many of the participants prepared white papers describing their validation experiences, identifying issues, and, where possible, suggesting promising technical approaches to address them. The organisers made these white papers available to the participants prior to the workshop. In addition, participants were asked to identify papers and other materials that contained valuable technical information about validation. This provided an excellent technical basis for the workshop itself. The white papers are provided, or their references listed, in Appendix B.

The workshop used a format that was a combination of focused presentations followed by breakout sessions to provide an opportunity for participants to discuss and brainstorm solutions to various technical and practical aspects of validation. The workshop opened with a plenary session of three presentations that provided participants with a broad overview of the approaches and challenges to control room validation. These papers are summarised in Section 2.2.

The plenary session was followed by four topical sessions each of which focused participants on specific aspects of validation and "challenged" them to develop recommendations for practices and research that will address known issues in conducting validations. The four challenge questions were:

1. What are the critical considerations in defining validation objectives and how do these impact achieving reasonable confidence?
2. What methods, approaches, resources, or rationales might be used for deriving performance requirements, selecting measures, and establishing acceptance criteria so as to support reasonable confidence?

3. What methods might be used to develop scenarios that maximise the amount and relevance of information in support of the validation conclusions and achieving reasonable confidence?
4. How should the validation results be aggregated and analysed to determine the final validation conclusions? Are inferential statistics meaningful in the context of MCR validation? If not, why not? What substitutes might be proposed as alternatives to traditional statistical modelling approaches?

Each topical session began with presentations from two challenge speakers. The workshop participants were then divided into four brainstorming groups; two focused on preliminary validation[1] and two on final validation.

The workshop concluded with two expert panels. The first expert panel addressed the topic of recommended practices. The panelists provide their perspective on the recommended practices identified during the workshop and the implications for current practices and future directions for control room validation. Each panelist provided a brief synopsis and commentary on a different one the four challenge issues and the proposed resolutions. The second expert panel addressed the topic of recommended research. The panelists provided their perspective on the recommended research identified during the workshop and discussed potential opportunities and prioritisation. Like the first expert panel, each panelist provided a brief synopsis and commentary on a different one the four challenge issues and the proposed resolutions.

2.2 Opening perspectives

Three opening presentations provided the perspectives and context for the subsequent topical sessions. The presentations addressed:

- Motivation for updating guidance, methods, and techniques for main control room validation and the ISV methodology issues found in the literature and industry experience (John O'Hara, Brookhaven National Laboratory – BNL).
- The challenges of conducting preliminary validation (Leena Norros, VTT Technical Research Centre of Finland Ltd. – VTT).
- The challenges of conducting final validation (Scott Malcolm, Candu Energy Inc.).

Each of these plenary papers will be summarised here. For additional information, abstracts and/or presentation materials for these papers are available in Appendix C.

The first plenary paper was given by John O'Hara, from BNL. The paper title was *Updating Guidance, Methods, and Techniques for Integrated System Validation*. Dr O'Hara began by noting that integrated system validation (ISV) is a key aspect of the US Nuclear Regulatory Commission's human factors engineering (HFE) safety reviews. The Nuclear Regulatory Commission (NRC) uses NUREG-0711(O'Hara et al., 2012) for review criteria for evaluating an applicant's ISV. The objective of the research O'Hara spoke about was to identify issues with ISV and suggest modifications and new guidance to address them. Specifically, BNL evaluated nuclear industry ISV experience, technical literature on test and evaluation of complex systems, and HFE standards and guidelines. They used this information to

1. The workshop organisers elected to use the term "preliminary validation" to refer generically to validation activities that might precede integrated system validation in the design process. Because the term and notion of preliminary validation had not been widely used in the technical literature, substantial discussion during the workshops was focused on gaining a common interpretation of this term (see the preface to Section 3 for more details).

identify issues and promising approaches to address them. O'Hara noted that most of the ISV issues involved knowledge gaps where additional guidance is needed in specific aspects of ISV methodology, such as scenario identification and performance measurement. In addition, BNL identified several high-level issues that have implications across many or all aspects of ISV methodology. These issues are alternative approaches to ISV, differences between new designs and plant modernisations, validating designs representing new concepts of operations, and grading the ISV effort.

The second plenary paper was given by Leena Norros, VTT Technical Research Centre of Finland. She discussed *Preliminary Validation – or a Life-Cycle Perspective to Validation of Complex Systems.* Preliminary validation (PV) is a concept and process that Dr Norros considers necessary for ensuring the quality of complex tools, e.g. the nuclear power plant control rooms, for their intended use. In contrast to ISV, the evaluation focuses on parts of the whole and guides design so that a "good" tool will be achieved. She noted that PV is independent from design, and it tests the achievements of the design against a conceptual reference of a good outcome. At the time of PV, the tool under evaluation is not yet ready, thus tests are needed that capture the potential of the *future* tool, and that are capable of anticipating the impact of proposed solutions on the whole system. Professional users, due to their insight into the requirements of the work to be performed, play a key role in foreseeing the appropriateness of the future tool and the PV methodology needs to accommodate this. Dr Norros gave workshop participants an example of a PV conducted for the modification of the instrumentation and control and control room systems of the Loviisa nuclear power plant (NPP). VTT's PV experience supports the adoption of a life-cycle perspective to validation and may be a realistic path for the future development of validation methodologies.

The third plenary paper was presented by Scott Malcolm from Candu Energy Inc. He discussed *The Challenges of Conducting Final Validation*. The author has over 20 years of experience in validating both new control room designs and modifications to existing facilities. That experience provides a basis for lessons learned for what Candu calls final integrated control room validation (FICRV). It is important to distinguish FICRV as it relates to design from validation of site operating procedures, training programmes, and staff licensing programmes. This design validation activity has a finite and clearly demonstrable end point, often associated with contractual and project milestones which shape its character. There cannot be ambiguity around the conclusion of this process for, in the end, the conclusion is acceptable or not. If not, identifying the remedies for making the design acceptable falls to the design team. The genesis of validation planning and hence the validation itself, is firmly rooted in the design requirements for the plant and its control rooms. It is also acknowledged that validation does contain an aspect of assuring that the design requirements are correct and complete, although this is tricky and must be navigated with careful consideration. Malcolm noted that there is much to be gained in advancing our confidence in validation outputs by improving the inputs – the requirements. He asserted that setting integrated system requirements for control room (CR) design has come a long way in the last 25 years and is producing designs that are demonstrably better than what went before. However, we can and must do better because the bar for acceptability of all aspects of plant design and operations is continually moving higher.

3. THE CHALLENGE QUESTIONS

In Section 3 we summarise the discussions that ensued in response to each of our four challenge questions. Although we did not transcribe the discussions verbatim, we captured the essence of these discussions in detailed notes. To faithfully represent the "brain storming" nature of these discussions, we have relied heavily on excerpts from these notes in the following summations, trying our best to use the original words of the speakers. As a result, the reader will undoubtedly find instances of our experts "thinking out loud" (i.e. putting forth new ideas for consideration with full understanding that upon further analysis their logic or concept may prove faulty or incomplete). The reader is therefore cautioned that our summations of these discussions are an attempt to capture the full range of discussion so that the ideas might be evaluated and further developed in subsequent efforts. Accordingly, one should not infer that the inclusion of a specific recommendation or position represents a consensus opinion of the workshop participants or endorsement by the Committee on the Safety of Nuclear Installations (CSNI)/Working Group on Human and Organisational Factors (WGHOF).

Sections 3.1 through 3.4 correspond to Challenge Questions 1-4, and each sub-section follows the same organisation: 1) introduction of the challenge question; 2) summation of the challenge presentations; 3) summation of the participant discussions; 4) proposed recommended practices and research; and 5) summaries of submitted white papers that we found relevant to the Challenge Question.

First, however, we provide the following prefatory information to provide context for understanding some of the diversity of views expressed by the workshop participants.

The notions of preliminary and final validation

During the early preparations for this workshop a consensus emerged among the organising committee that the scope of the workshop should not be limited to a focus solely on integrated system validation (ISV). Rather, review of the known challenges associated with conducting and interpreting ISV tests suggested that a broader examination of the control room design process and methods used to evaluate the design could be fruitful. To communicate this broader scope of interest to the workshop participants, the organising committee used the terms "preliminary validation" and "final validation". Our intent was for final validation to be typified by, though not necessarily limited to, ISV. ISV is a well-established term commonly referring to the acceptance testing during the final phase of a design project. A definition for ISV can be found, for example, in the NUREG-0711. Preliminary validation is a concept that has not been clearly defined through the technical literature. Given that the purpose of including "preliminary validation" in the workshop scope was to explore new paths to achieving reasonable confidence in control room validation, the organising committee did not define or bound this concept for the participants. Understandably, workshop participants sought to define this term during the workshop and potential definitions and concepts were discussed in each of the topical sessions, most notably in the session on

scope and objectives of validation. Here we report the highlights from the topical session discussions regarding the definition and role of preliminary validation. The technical matters regarding preliminary validation (e.g. how one might conduct a preliminary validation) are reported within each topical session.

Understanding and defining preliminary validation – discussion highlights

There seemed to be little disagreement among the workshop participants that with appropriate constraints, the results of activities conducted prior to ISV may be useful to achieving reasonable confidence in the validation of control room designs and modifications. However, as noted above, the participants discussed at length this notion of "preliminary validation" in an effort to reach a common understanding. As summarised below, general discussion of preliminary validation by the workshop participants touched upon the definition of preliminary validation (PV), the role of PV, its relationship to ISV and life cycle evaluations, and other practical considerations.

Defining PV: Participants used several related terms when referring to preliminary validation such as "pre-validation" and early validation. Since PV is a relatively new term workshop participants likely held quite different views regarding its meaning. One discussion group stated that PV needs to be developed before many of the other questions can be adequately answered. The comments highlighted the fact that a widely agreed upon definition of preliminary validation has yet to emerge and that those engaging in discussions in this area should be mindful that this lack of definition can be challenge in communicating and developing the concept of preliminary validation. Additional participant comments:

- As the term suggests, it is validation that is performed before the final ISV. The question is whether "Pre-validation" is a pre-ISV or a set of human factors engineering (HFE)-tests conducted prior to the final ISV.

- There should be a clear distinction between PV from "every day" design tests.

Role of PV: A common theme of participant discussions was the belief that there is a need to define and develop the role of PV in the validation of main control room designs and modifications. Participants offered different visions of the role of PV, including the following:

- The purpose of PV is similar to final validation; to demonstrate acceptability, but PV has an additional purpose of evaluating design features.

- A characteristic feature of PV is that its scope is more limited than that of ISV: PV does not aim to validate the whole main control room (MCR) or all of the systems, but only selected parts of the whole (e.g. systems validated separately one from each other). As a consequence, there is a series of pre-validation test activities distributed over the design process, that is, PV activities support a process that includes periodic validation, throughout the design process.

Relationship between PV and ISV: Much of the discussion about defining PV was concentrated on clarifying the relationship between preliminary validation and final validation. This is understandable since preliminary and final validation of a given validation process likely need to be defined in relation to each other. Also, final (integrated) validation is a relatively well-established term that preliminary validation can be compared to and contrasted against. One line of discussion focused on a gap between PV and final ISV, while another line of discussion focused on an evolutionary process where the PV and ISV stages are defined by the completeness of the design. The following comment summations illustrate some of these views and concerns of the participants:

- Regarding the relationship between PV and ISV, there is a gap between pre-validation and final ISV which should be defined in detail and the way to bridge it has to be discussed.

- The completeness of the design determines the characteristics of the validation. At some point in the process of validation activities, a stage is reached where PV stops and ISV begins. This change of stages occurs when the design is complete. But it can be a challenge to define when the design is complete or when it is complete enough. Several considerations were suggested such as: when you can run all of your scenarios; when your procedures are ready; when you have trained crews.

- One objective of PV could be to evaluate whether the design is ready for final ISV. There is the possibility that systems are entering the final ISV stage too early; perhaps one objective of preliminary test activities is to provide a means for assessing whether a design is ready for ISV.
- Preliminary validation could be a series of independent tests dedicated to validate human-system interface components independently, one from each other but considering future ISV activities (e.g. by applying ISV scenarios, real procedures and so on).
- Preliminary validation could be a series of ISV pre-tests (e.g. tests done on a full scope simulator, refined along the way through simulator upgrades before entering the final ISV).
- The criteria for preliminary and final validation are different: for preliminary validation, the criteria could be, for example, improvement of performance over time and tests; and for ISV, pass/fail-type of acceptance criteria.
- It was asked whether preliminary validation can be viewed as part of ISV or should be viewed as a separate activity.

Relationship between PV and lifecycle evaluations: Some discussions related PV specifically to lifecycle evaluations and suggested a role for PV in longitudinal and evolutionary processes as captured in the following comment summary:

- The definition of PV should be considered as part of a longitudinal process and an evolutionary process; how the human performance characteristics are changing over time. It should be considered as part of the lifecycle of evaluation.

Practical considerations

Guidance: There seemed to be some controversy as to how much guidance is needed for PV testing.

- Some thought that more guidance, and some kind of regulatory guidance is needed, some thought that the designers can do whatever they like during the design process. According to this latter view, it was suggested that perhaps designers could benefit from industrial guidance, but that there is no place for regulatory guidance for pre-validation.
- It was asked whether a regulatory focus on its guidance on ISV could cause a vendor to focus only on passing ISV.

Managing Independence of the Validation Team: Several comments touched upon team independence as a potential practical challenge to conducting PV activities. Providing an independent validation team can be a resource challenge and establishing an expectation for independence early in the design process would add additional burden to the validation process.

- It was thought that perhaps the independence issue requires rethinking and modification.
- It was asked whether we should grade level of independence and if so, how we can do that.
- It was asked whether the independence changes over the design process.
- It was asked whether it could be enough that if a person is not defending "her own design", then she is able to do the validation.

Considered collectively, the comments and questions concerning the notion of preliminary validation suggest that whereas many see preliminary validation as a promising avenue towards achieving reasonable confidence, much work may need to be done to better define the concept and its relationship to final validation.

3.1 Challenge 1: Critical considerations in defining validation scope and objectives

Possible disagreements may arise among individuals about specific validation issues caused by different interpretations of the concept itself and of the objectives of validation. Thus, it is important first to discuss how the concept of validation is to be defined in the context of nuclear power plant control room design and how validation scope and objectives have to be determined. This question may even be one of the most challenging ones in that it is often difficult to put into words the meanings and objectives associated with a particular concept.

As the working group was developing plans for the workshop, it identified questions that are related to the determination of validation scope and objectives. These questions included:

- What is the main objective of the ISV and what are the general requirements for ISV?
- What are the critical considerations in defining the objectives and how do these impact achieving reasonable confidence?
- How should one develop and define objectives regarding the evaluation of the integrated performance and for the parts of the MCR?
- Should ISV be mainly an "acceptance" test, evaluating if minimum acceptance has been achieved or should it also include identification of areas for improvement beyond minimum acceptance? Does it include pointing to improvements when acceptance criteria are not met?
- Should the scope of main control room validation be expanded to explicitly address maintenance tasks?
- How can probabilistic safety assessment (PSA), including human reliability analysis (HRA), be used to better achieve reasonable confidence in the human factors validation of the MCR system?

Considering these and similar questions related to control room validation, the working group elected to pose the following challenge question to the workshop participants.

What are the critical considerations in defining the scope and objectives of a control room validation and how do these impacts achieving reasonable confidence?

Workshop participant discussion of challenge question 1 was preceded by presentations from two invited speakers who provided additional background and perspectives for consideration during the session. These

speakers were Robert Hall, REH Technologies, and Julie Reed, Westinghouse Electric Corp, who gave a joint presentation on the challenge topic.[2]

3.1.1 Summaries of challenge presentations

Note to reader: Mr Hall's and Ms Reed's presentation materials can be found in Appendix C.

Challenge presentation for Topical Session 1 (Robert Hall, REH Technologies, and Julie Reed, Westinghouse Electric Corp)

Mr Hall's and Ms Reed's presentation critically considered preliminary and final validation activities in order to identify and understand their strengths, limitations and relationships between technical and practical considerations. They also considered that it is important to discuss variations in international practices and present lessons learned in validation cases.

They set several practical goals for validation testing activities which emphasise the need for clear understanding of the validation objectives, validation must be: 1) focused in scope; 2) based on realistic resource commitments; 3) meet client and regulatory expectations; 4) support achieving reasonable confidence; and 5) include future marketplace demands. They noted that many times validation testing activities do not fulfill the hopes placed on them. The scope of testing may be too narrow in many respects, e.g. it is limited to either testing specific design elements of assessing final acceptance, it is limited to the main control room activities and operations, or does not utilise past testing results as a starting point in the design of validation tests. The objectives also often have several limitations regarding, e.g. the goals of achieving reasonable confidence, integration of preliminary and final validation, and the consideration of limitations of technology, and of costs, resource and schedule restraints.

The presenters had selected one specific challenge topic for a more in-depth consideration, that is, how to successfully introduce a design solution across international borders. This is a topical issue, since, on the one hand, many new plant designs and control room platforms have their country of origin outside the US, and on the other hand, US based designs and platforms have been exported to other countries. In both cases, the key question is how to take credit for country of origin validations in another country. For example, the presenters have seen difficulty in using earlier validation results to support design acceptance in the US market. According to them, the difficulties are mainly caused by the lack of a consistent set of objectives and standards for validation testing and the adequate use of preliminary testing results.

They wanted to reject the conception that there could be a standard plant providing a basis for control room validations in all countries, because regulatory expectations and practices vary from one country to another. However, they see the need for a debate about the possibility of having a universally accepted validation process and optimisation of the validation process for cross-country migration.

Finally, they raised several topics for consideration that were in part based on the list that the working group prepared for the workshop (see above). The new topics addressed the following questions: 1) what should be considered in setting the validation objectives for designs that may enter the world market; 2) in what ways can validation test results be used and reused for different purposes; 3) what are the differences between validation of a "first of its kind" design and validation of a design based on a predecessor/reference design; 4) how should tool availability issues be considered in validation planning; 5) what credit could be given to other test activities (e.g. pre-operation start up testing, factory acceptance tests) and how could they contribute to the achievement of reasonable confidence; 6) how do the objectives differ for validation of a new plant and validation of a modernised plant.

2. Biographical sketches of the speakers are available in Appendix A.

3.1.2 Discussion highlights

Challenge 1. What are the critical considerations in defining the scope and objectives of a control room validation and how do these impact achieving reasonable confidence?

The discussions of Challenge 1 are summarised and categorised into two major groups; characterisation of validation scope and characterisation of validation objectives. Within each of these two areas we grouped the workshop participants' comments into more specific themes. It should be noted, however, that individuals expressed the view that the concepts of validation scope and objectives were closely related and that it is difficult to discuss one concept independent of the other. Others noted that they were not clear on the distinction between objectives and scope. The reader will likely see these perceptions reflected in the discussion summaries.

Many participants provided comments to indicate that NUREG-0711 provides a firm basis for control room validation and accomplishment of ISV. However, it was pointed out that there are important issues (identified later in this report) that are not covered in NUREG-0711, and in order to reach reasonable confidence, a broader perspective on validation may be necessary than is presented in NUREG-0711.

Many discussions were centered on the objective of PV and its relation to ISV; this is understandable because to many of the participants, PV was seen as "the new baby in the house" and there was an urge to determine its status in the validation process. Discussions of PV are presented in this section to the extent that they are related to defining the scope and objectives of validation. Discussions focused on defining the concept of PV are summarised as the introductory material to Section 3.

General considerations

Participants provided the following "general" comments which we understood to be considerations for establishing both the scope and objectives of validation:

- Consider type of project and extent of change. Is the validation for a new plant or upgrade? If an upgrade, what is the extent of the upgrade? The scope and objectives are affected by the extent of change.
- Identify the novel/unique design elements.
- Consider the operating philosophy/concept of operations when determining what you are trying to validate. Examine the tests that have been done previously:
  - Was the design evaluated earlier in the design process, or validated in another context?
  - Prior evaluations increase confidence and may impact the objectives, may be able to take credit for past evidence, even if it will not be sufficient by itself.
- One cannot necessarily take a validation plan from one country to another. Even with the same design, the validation team may need to do specific validations in different countries.

Validation scope

<u>Elements/breadth</u> – The following comments addressed the matter of scope in terms of what should be included in the scope or how the scope might be determined:

- Focus on the overall socio-technical system, including maintenance, minimum manning, etc.

- Consider including out of the control room activities (e.g. initiating events that start outside the control room, activities that are carried out in secondary control areas such as the remote shutdown station, and in technical support and outage control centers).
- Link the scope to a realistic example; establish check list of most frequent tasks linked with normal operation, difficult tasks with abnormal operation… use this database as a common base to be used by different users.
- Consider how lessons learned from past operating experience can inform the scope.
- Include risk-significant actions.
- Use probabilistic reliability assessment (PRA)/HRA to inform preliminary validation.
- Consider testing the operating philosophy/concept of operations (discussed further on pp. 23-24).[3]
- Identify a standard set of sub-systems for testing in PV (see also *Variance in scope of validation* on p. 24).

In discussing validation scope it was proposed that perhaps with a wide range of performance testing, covering what is expected during ISV, one would only need to perform limited testing during ISV. This would appear to presume that it is possible to identify tests of sub-systems that would be unaffected by the subsequent integration of the sub-systems and that the ISV tests would focus on areas of performance likely to be impacted by the integration. Related to this notion, it was suggested that to achieve reasonable confidence in conclusions, the focus of ISV should be on key systems and interactions so that sufficient information can generated concerning those aspects critical to the integration of systems.

Although as noted above some individuals advocated expanding the scope of control room validations to include their interactions with activities outside the main control room, such as maintenance activities, others questioned how this would be done as it would seem to require an entirely new process to establish the requirements, mock-ups, etc. to conduct the validation.

Testing the operating philosophy/concept of operations – As noted above, it was recommended the operating philosophy or concept of operations be addressed in the scope of the validation because it was seen as useful for providing an overall framework, in a systemic approach, that describes the roles and responsibilities of the operating staff in operating the plant and can include what operating staff will do and how they will perform their activities within the control rooms for operating the plant. There was substantial discussion on this topic and the following comments were noted:

3. During the workshop some participants advocated that the validation scope include consideration of the "philosophy of operations" or "concept of operations". The choice of term used by individuals appeared to be more dependent on which term was more common or familiar to them than on suggesting one over the other, as it appeared the participants considered the two terms to be nearly equivalent in meaning. In this report we have elected to use "operating philosophy/concept of operations" as the default term in summarising comments. As these terms may be new to some readers, we note that NUREG-0711 provided the following definition of "concept of operations": *A concept of operations (ConOps) defines the goals and expectations for the new system from the perspective of users and other stakeholders and defines the high level considerations to address as the detailed design evolves. An HFE-focused ConOps addresses the following six dimensions: 1) Plant Goals (or Missions); 2) Agents' Roles and Responsibilities; 3) Staffing, Qualifications, and Training; 4) Management of Normal Operations; 5) Management of Off-normal Conditions and Emergencies; and 6) Management of Maintenance and Modifications.*

- It is important to recognise that ISV is not just for a design, it is also for a particular operating philosophy/concept of operations. An operating philosophy/concept of operations can be specific to a country, or reflect particular constraints that are required by the operating organisation (e.g. level of acceptable automation, crew member versatility and interaction, etc.).
- Operating Philosophy/Concept of operations should be tested as early as possible, but sometimes it may be difficult (e.g. in case of revolutionarily new design concepts). For example, novel operational concepts (e.g. staffing level) should be tested early on, before it is too late to change them.

Variance in scope of validation – Although some participants suggested that the scope of validation include a standard set of sub-systems, the following comments reflected a belief that the validation scope is variable both between and within projects:

- Scope is project-specific (especially in modernisation projects).
- Scope also changes over the project.
- Although it was suggested that the scope of validation testing should become smaller as testing progressed from the early design to ISV, others expressed opposing views.
- At early stages, pre-validation activities typically focus on individual systems, procedures etc.; at later stages a more integrated set of systems are tested in an operational context.

Practical considerations – In addition to the factors to be considered for ensuring that the scope is appropriate for providing confidence in the validation, the workshop participants also identified practical considerations that were more focused on what is reasonable, including the following:

- cost;
- schedule;
- resource availability (people and equipment).

In the theme of practical considerations, some noted that it is important to pay attention to costs and concentrate on important systems and issues in validation tests. However, it was also noted that it is not the goal to make validation cheap and simple if reasonable confidence is compromised and questioned whether it should it be simple to demonstrate the safety of a complex system.

Validation objectives

General points of discussion – Regarding the objectives of control room validations, one line of discussion was consideration of whether the objectives of design validation can be distinguished from those of design testing. It was proposed that the objective of validation is on a determination of acceptability whereas design tests explore alternatives/strengths and weaknesses in design. Some individuals suggested that validation is a test of the completeness of a design, noting that it has to be defined when the design can be said to be complete. It was also noted that to the extent that an objective of validation is attaining reasonable confidence, it is necessary to define what "reasonable" means.

Specifying the purpose of validation – Views concerning the objectives of validation and the critical considerations in determining the objectives were varied. Many individuals expressed the belief that we cannot determine a fixed set of objectives for validation. Consistent with this position individuals commented that different types of projects have different kinds of objectives. The following comments

reflect both that validations can have multiple objectives and that there are diverse views on the objectives of validation:

- One individual suggested that beyond making acceptability determinations, validation includes learning about how to improve conduct of operations.
- Establish ISV as an integrated final design test, not just an HFE test. Establish critical points in the design process to include more integrated testing.
- Objective should be to include unexpected situations as a form of stress test.
- Objective should be to evaluate at the outcome level and the elements should support the overall performance. Sometimes a sub-system (e.g. a new alarm system design) may not support performance to the extent anticipated.
- The overall objective should be a human performance centered safety case for which ISV is only one piece of evidence.
- Being able to demonstrate that the testing context generalises to real world performance.

It was asserted that clarifying the objectives of validation is important because sometimes there are conflicting motivations (e.g. schedule pressures).

3.1.3 Recommendations

Recommended practices

- To enhance confidence in the validation, establish coherence between early and final validation efforts.
- Allow sufficient time for validation such that identification and resolution of design issues are not challenged by schedule pressures.
- Include the operating philosophy/concept of operations in the scope of validation.

Recommended research

- Evaluate which activities conducted outside of the control room should be included in the validation and at what level of fidelity?

3.1.4 Related white papers

Many of the white papers submitted by workshop participants touched upon the question of validation scope and objectives. Two papers are summarised briefly here. Except where noted otherwise, the full texts of these papers are available in Appendix B.

Tecnatom perspective in relation to validation and ISV from the HFE viewpoint for nuclear power plants by Mr Pedro Trueba Alonso. Mr Alonso's white paper presented several considerations regarding validation and ISV from the perspective of Tecnatom. Based on international standards, three kinds of validation cases can be identified: 1) validations occurring as part of research and development (R&D) activities; 2) validations as part of *generic* licensing activities; and 3) validations as part of *specific* licensing activities. The validations included in the third group can be further categorised into partial validations and ISV. As Mr Alonso emphasised, both of these approaches are consistent with an iterative process advocated by several standards. Since most of Tecnatom's experience is from ISV, the rest of Mr Alonso's paper focused on validation matters from the ISV perspective.

With regard to scope and objectives of ISV, the objectives have to comply with the main standards and guidelines in the nuclear domain. Regarding ISV applicability, ISV is a necessary task for new designs, but its applicability to plant modifications depends on the type of modification, the way the plant is modified and the safety class of the systems involved in the modification. Also ISV objectives differ between new plants and design modifications as the scope and technical approach may be different. However, in all ISV tests the main objective is to achieve a safe plant condition.

Mr Alonso pointed out that it has to be taken into account that ISV has to be in a schedule with other design activities that influence each other. A graded approach is needed, because it is impossible to evaluate everything with enough detail in a limited time frame. At least those activities and equipment which have a significant contribution to Core Damage Frequency should be considered in ISV. For the same reason, the focus must be on the main control room and remote shutdown stations, and Mr Alonso also suggested that maintenance tasks would be out of scope of ISV, and instead a separate verification test for maintenance activities and equipment could be arranged during training activities. The issues raised during the design process should be addressed before the execution of the ISV so that there is no longer a need to consider these issues. In a similar way, it is important that all the findings are addressed, and all Human Engineering Discrepancies are solved – otherwise ISV would be meaningless.

What do we mean by reasonable confidence in human factors-related validations of integrated system performance for NPPs by Dr Alice Salway. Dr Salway described in her paper the evaluation topic of a new Canadian Standards Association standard "CSA N290.12 Human Factors in Design for Nuclear Power Plants". According to Dr Salway's paper, human factors (HF) evaluations are needed because of complexity and variability of several technical systems. Human and system performance cannot be entirely discovered through analytical techniques. HF evaluation activities can identify critical aspects of emerging system behavior throughout the design process.

The paper presented HF activities as active processes that are carried out throughout the design process and even beyond. Early evaluation activities enable HF issues to be identified while the design process is ongoing and changes can still be made. Dr Salway pointed out that it is important to start HF evaluations as early as possible, because it is more difficult and expensive to resolve the HF issues later on when the design has been further developed. Human Factors evaluations as well as HF design activities should proceed in an iterative fashion. In this way cumulative and converging evidence can be gathered of the safety and functionality of the new systems. An iterative series of evaluation activities also provide useful insights concerning operator training and procedure design.

Dr Salway presented an overview of different types of HF evaluation activities ranging from inspection-based methods to usability testing and validation. Two types of validations were identified: validation of equipment and integrated system validation. Validation of equipment was characterised as a validation where equipment and a set of system components are evaluated against relevant system requirements. ISV, in turn, has a larger scope and scale than validation of equipment; high fidelity is required in ISV; and it is focused on operator performance.

At the end of the paper, Dr Salway described the role of HF evaluation at different design phases, and she emphasised the importance of follow-up evaluation activities after commissioning as an effective way to identify HF issues. Finally, some additional considerations were made regarding importance of a risk-informed graded approach, and some examples of HF evaluations beyond engineering design (i.e. evaluation of minimum staff complement and validation of procedures) were described.

3.2 Challenge 2: Rationale for selecting measures and acceptance criteria

Approaches to the human factors validation of MCR systems may differ in their requirements or emphasis regarding measurement during test scenarios and what constitutes acceptable performance. However, since most MCR validation projects evaluate quite similar systems, there may be substantial overlap in these areas. Nevertheless, there is currently no universal agreement regarding the measurements, performance requirements, and associated acceptance criteria that should be used for establishing reasonable confidence. In developing its plans for this workshop, the working group identified a number of questions pertaining to measurement, performance requirements, and acceptance criteria during validation testing. These questions included:

- Is it possible to specify a minimum set of observations and measurements that should be conducted as part of the human factors validation of the MCR? If so, what is the minimum set and what is the rationale (e.g. model or framework) that explains why they are comprehensive or sufficient?
- What is the appropriate role, if any, of measures that are not performance-based (e.g. usability evaluations) in the human factors validation of the MCR?
- Do measures used in the validation of a MCR differ in their value for achieving reasonable confidence? If so, what are these differences and what are the implications for their use in MCR validations?
- When using a reference system to define the performance requirements (i.e. as used in a benchmark approach) for the human factors validation of an MCR, what are the implications for achieving reasonable confidence?
- Is the benchmark approach to human factors validation of MCRs limited to certain objectives and types of projects (e.g. MCR modifications)? If so, what are the boundaries for appropriate application of the benchmark approach?
- What methods, approaches, resources, or rationales might be used for defining performance requirements for the MCR validation?
- What material from the design activities and safety analyses can be used to inform the requirements? And how?
- Is it possible to establish human factors validation criteria in absolute terms? How might it be done and what would be the challenges?

Considering these and similar questions related to control room validation, the working group elected to pose the following challenge question to the workshop participants.

What methods, approaches, resources, or rationales might be used for deriving performance requirements, selecting measures, and establishing acceptance criteria so as to support reasonable confidence?

Consistent with the format for all Topical Sessions, workshop participant discussion of the challenge question 2 was preceded by presentations from two invited speakers who provided additional background and perspectives for consideration during the session. These speakers were Cecilia De la Garza, of *Électricité de France* (EdF), and Per Øivind Braarud, from the Halden Reactor Project (HRP).

3.2.1 Summaries of challenge presentations

Note to reader: Dr De la Garza's and Mr Braarud's presentation materials can be found in Appendix C.

Challenge presentation for Topical Session 2 (Cecilia De la Garza, EdF)

Dr De la Garza (De la Garza et al., 2015) presented a case study for participants to consider in which EdF conducted a control room validation for a new reactor. The goal of the validation effort was to confirm that the MCR design supports the required control room functions. She explained that validation is carried out as an iterative process at EdF; it is linked to design, which continues until the commissioning and the first phases of operational feedback.

EdF validation activities included three preliminary validations and four final validations. Preliminary and final validations are delineated in the following way:

- Preliminary validations use mockups to test and validate design principles and specify engineering rules for the detailed design. These validations are used to reduce project risk associated with design features.

- Final validations use full-scale simulators to validate design choices. Multiple final validations allow for more robust testing by increasing the number of scenarios and crews tested across multiple and complementary evaluations.

The human factors measurement and analysis done as part of the validation effort are carried out by a multidisciplinary team consisting of psychologists/ergonomists and human reliability analysts. The focus of the ergonomic analysis is on the man-machine interface (MMI) and the interactions between the operator/crew and the MMI. The analysis is conducted from the point of view of the individual and team (De la Garza et al., 2013). It utilises observations, note taking, video/audio recordings, simulator logs, psychological instruments, interviews and crew debriefs for data collection. The ergonomic analysis results in recommendations that are discussed with the design teams and analysed in terms of the feasibility. The focus of the HRA is on emergency operating system (EOS) activity that is linked to safety. The goal of the HRA is to point out weakness and strengths of the EOS in terms of safety. The HRA gives priority to examination of teamwork behavior and uses observers, note taking, simulator logs, dynamic process data (e.g. curves related to the scenario), interviews and crew debriefs for data collection. The HRA results in recommendations which are assessed with regard to the spectrum of situations (as in probabilistic safety assessment).

Preliminary validations were supported only by ergonomics, while final validations by both ergonomics and HRA. EdF intends to explore how this dual approach could be developed for preliminary validation.

Challenge presentation for Topical Session 2 (Per Øivind Braarud, Halden Reactor Project)

Mr Braarud challenged participants to consider what is most important with regard to establishing human performance requirements for an ISV. During actual operation human performance "requirements" consist of the behaviors/actions that allow for safe operation. We must be careful not to divorce the ISV context from the operational environment.

Mr Braarud presented the following viewpoint to workshop participants, "The most important issue regarding human factors evaluation is what should be measured and why. For ISV, the link between human performance requirements and system requirements is central – what human performance requirements relate to the plant's safe operation?" ISV test performance criteria should be viewed such that if criteria are satisfied, we believe the real life requirement can be fulfilled. If criteria are not met, performance issues have been identified. There must be a clear linkage between "technical" requirements (e.g. reactor vessel pressure, core temperature, readiness of safety trains and safety functions) and human performance

requirements (e.g. process control actions and plant monitoring, requirements of how work should be performed, performance support tools that should be provided in the MCR).

Mr Braarud proposed several sources/activities that can help identify and clarify human performance requirements and specify criteria including: task analysis, event analysis, function analysis, training, procedures, HRA, and the operating philosophy/concept of operations. Mr Braarud explained, through a graphic, his proposed overall measurement framework that requirement based criteria would represent a top down approach while empirically-based criteria would be a bottom up approach. He suggests that a good approach could be to combine different types of criteria, but that ultimately, the criteria applied should relate to the human performance requirements for the actual plant (Braarud, 2015). Further, he conveyed that it must be recognised that each plant is a unique case and the plant specific operational and safety concept are important. While literature provides generic information about human performance measures, the final measures for a given ISV need to be defined/adjusted, with plant specific content in mind which Mr Braarud notes is a non-trivial task.

3.2.2 Discussion highlights

Challenge 2: What methods, approaches, resources, or rationales might be used for deriving performance requirements, selecting measures, and establishing acceptance criteria so as to support reasonable confidence?

The discussion of Challenge 2 is summarised under the three major elements of the challenge question: methods for deriving performance requirements, selecting measures, and establishing acceptance criteria. Within each of these elements the workshop participants' comments are further organised into more specific themes that the working group identified as it reviewed the input received from the workshop participants.

Performance requirements

The workshop participants suggested two approaches for deriving performance requirements and identified additional information that they thought should be considered when deriving performance requirements. The reader should note that the approaches summarised were not necessarily presented as alternatives to each other and in fact could be used in conjunction. The following is a summary of participant comments concerning these approaches.

Multidisciplinary approach

- A multidisciplinary team (e.g. systems, operations, and HF experts) should be used to conduct a system functional review and task analysis as a means to define human performance requirements.
- A multidisciplinary approach should be used for the analysis of empirical data (both quantitative and qualitative) from observational studies.

Functional decomposition model approach (see Figure 4.1)

- There should be multiple levels of performance requirements including:
  - mission requirements;
  - human performance requirements;

  - cognitive and process requirements;
  - Human-system interface (HSI) support requirements.

- Criteria at the mission requirements levels should be pass/fail.
- Lower level requirements should be used as diagnostic indicators to improve design.

With respect to the functional decomposition model, there was discussion but no consensus as to whether measures related to human performance requirements should be pass/fail or diagnostic.

In the course of discussing considerations for establishing performance requirements, the workshop participants identified operational experience/previous testing, culture, training requirements, and novel aspects of the design, as potentially relevant.

Selecting measures

At a general level, it was commented that the current knowledge base on human performance and measuring needs to be properly incorporated into industry and regulatory guidance for ISV. Workshop participants discussed several areas where there is debate with regard to the type of measures that should be used in ISV and identified a potential approach to selecting combinations of measures so as to enhance confidence in test results. The following is a summary of participant comments:

Tailored measures – It was proposed that measures should be tailored to the evaluation phase being conducted as some measures may be more important in pre-validation than final validation or vice versa.

Practical vs. theoretically-based measures – The question was raised of whether it is more important to have measures with "practical value" or measures that have been shown to be valid for the construct being measured. In general there was no consensus on when to use practical/expert input to select measures and when to use a theoretical/academic basis for measurement selection. It was noted that it should be possible to have measures that have both practical value and a sound technical basis.

Types of measures to consider – Although the participants did not attempt to identify an exhaustive list of the types of measures to consider when conducting validation testing, discussions touched upon the possibility of including: 1) resilience measures (i.e. indicators that should be able to predict good performance in scenarios for which the crews have not been specifically trained); 2) physiological, eye-tracking and neuro-ergonomic measures; 3) usability measures; and 4) observer-based measures.

Triangulation approach – It was suggested that to improve confidence in measurement results one should use three types of measures to gain convergent validity. Such measures could include:

- A task performance measure (human or plant performance).
- A primary measure of a cognitive or physical construct (e.g. workload).
- A secondary measure of another construct that helps provide insight when the task performance measure and the primary measure do not agree (e.g. situation awareness).

Acceptance criteria

On the matter of establishing acceptance criteria, participants expressed a need for guidance. Aligned with this view, there was a perceived need for research to support the development of acceptance criteria in the areas of teamwork, decision-making, command and control, and workload. At a more general level, others

saw a need for guidance on when to use different approaches to establishing acceptance criteria, such as determining acceptance based on ability to manage worst case conditions, use of convergent measures, and margins/tolerance. It was suggested that different strategies may be needed to establish acceptance criteria for different dimensions of performance. (e.g. worst case approach for workload, tolerance/boundary approach for performance time).

It was also suggested that the operating philosophy/concept of operations for a design could provide a framework to guide the selection of acceptance criteria throughout the design process. It was not discussed how this might be accomplished. Others noted that it may be necessary to tailor acceptance criteria to the phase of validation in which one is conducting the tests.

3.2.3 Recommendations

Recommended practices

- Use a multidisciplinary approach to establish performance requirements.
- Vendors should share measures and measurement data along with good practices to enable cost-effective validation.
- Tailor acceptance criteria to the phase of validation testing being conducted.
- Include usability measures in validation.

Recommended research

- Identify measures that predict resilience.
- Determine how simulation can play a role in selecting measures and establishing acceptance criteria.
- Establish recommended measurement technologies (e.g. eye-tracker, electroencephalogram, etc.).
- Determine how to combine data from different constructs in a meaningful way.
- Determine the predictive value of data collected during pre-validation.
- Develop the technical basis for establishing acceptance criteria in the areas of teamwork, decision-making, command and control, and workload.

3.2.4 Related white papers

ISV – Reduction of requirements, clear and simple measures, hypotheses and success criteria by Thomas Elsner. Mr Elsner's white paper briefly describes his perspective on ISV testing requirements. He notes that the purpose of ISV is to demonstrate that the integrated control-room system meets the intended purpose of enabling control room staff to effectively (and efficiently) and, thus safely, operate the plant. He suggests that ISV should have testable performance requirements with strong and unambiguous relationships to the purpose. He explains that supplemental measures (e.g. situation awareness) may provide further understanding about good or lacking performance. Mr Elsner concludes that when task performance can be successfully validated (i.e. no violation of effectiveness/efficiency in any ISV scenario), the integrated control room system can be seen as validated.

Envy in V&V: Preliminary thoughts in support of the WGHOF Workshop on verification and validation by Ronald Laurids Boring, PhD. Dr Boring's white paper is an opinion piece addressing verification and validation (V&V) in the nuclear community. He suggests that V&V researchers envy

fields that provide highly conclusive findings, and, subsequently, compensate by increasing the amount and types of measures used. However, he maintains that, regardless of these measurement efforts, findings from V&V will rarely be conclusive. He suggests that the V&V should be considered not just an integrated system validation striving for finality in the results but as part of an iterative user centered design process. He suggests, to be successful, iterative V&V requires early involvement of human factors practitioners in the design process and including only those measures that have practical utility and provide insights into operator performance. Dr Boring resolves that it is the process of demonstrating improvements in the design – not the conclusiveness of the findings – that determines if the system is successful.

ISV – Human performance evaluation: Needs, considerations and constraints, criteria, and rationale by Jun Su Ha and Poong Hyun Seong. Dr Seong's paper discusses considerations to take into account when conducting an ISV. Regarding human performance evaluation, Dr Seong suggests that particular attention should be paid to the changes in the operational environment when selecting human performance measures for the ISV. He also suggests that the evaluation of human performance should be practical and cost-effective and that evaluation criteria for the performance measures should be clear. Dr Seong proposes that the following questions must be asked and satisfactorily answered before selecting human performance measures for ISV:

- Do human performance measures adopted and/or developed encompass the changed operating environments?
- Are they practical and cost-effective?
- Do they have clear performance criteria?

If all the questions are satisfied, the measure can be included.

3.3 Challenge 3: Methods for scenario development

Additional guidance pertaining to the selection and design of scenarios is a need that is frequently identified for ISV. The questions surrounding scenario development tend to revolve around the circumstances, types, and numbers of scenarios that should be used. This includes guidance for the selection and use of operational conditions, as well as guidance for the identification of what constitutes a sufficient set of test scenarios (i.e. what scenarios should be used and how many of them?). In developing its plans for this workshop, the working group identified a number of questions pertaining to the construction of validation test scenarios, the test design, and their relationship to the validation framework. These questions included:

- How might it be possible to define a set of scenarios that is sufficient for achieving reasonable confidence?
- How is a comprehensive set of scenarios constructed and how should these scenarios be modeled and analysed?
- What methods might be used to develop scenarios that maximise the amount and relevance of information in support of the validation conclusions and achieving reasonable confidence?
- What methods can be used to optimise the predictive value of validation scenarios?
- How might methods such as functional requirements analysis, function allocation, task analysis, and human reliability analysis be used to develop scenarios or define a set of scenarios that support achieving reasonable confidence?

- What methods can be used to classify, divide into elements, or otherwise process scenarios in a manner that allows for meaningful aggregation (collapsing) of data across scenarios for purposes of analysis?
- What are the limitations of using scenarios of high complexity and difficulty as a means to compensate for practical restrictions on the number of validation scenarios that can be conducted?
- How might one characterise sub-system validation or other multi-phase methods in terms of a test design that supports human factors validation of integrated MCR system performance?

Considering these and similar questions related to control room validation, the working group elected to pose the following challenge question to the workshop participants.

What methods might be used to develop scenarios that maximise the amount and relevance of information in support of the validation conclusions and achieving reasonable confidence?

Consistent with the format for all Topical Sessions, workshop participant discussion of the Challenge Question 3 was preceded by presentations from two invited speakers who provided additional background and perspectives for consideration during the session. These speakers were Cecilia De la Garza, of EdF and Emilie M. Roth, from Roth Cognitive Engineering.

3.3.1 Summaries of challenge presentations

Note to reader: Dr De la Garza's and Dr Roth's presentation materials can be found in Appendix C.

Challenge presentation for Topical Session 3 (Cecilia De la Garza[4], EdF)

Dr De la Garza presented an approach that was used by EDF for developing scenarios for the ISV programme of a new reactor. The primary focus of the presentation was a three-step methodology that involved an ISV programme definition, scenario definition, and participant sampling (number of crews).

Because the ISV programme definition impacts the construction of the validation test scenarios, the first step involves reviewing the principles of an ISV programme to ensure that the combinations of different control room components (e.g. procedures, human-machine interfaces, personnel) help achieve the performance objectives of the plant (e.g. safety, production). Dr De la Garza explained that the evaluation campaigns (e.g. simulations, tests) performed throughout the ISV should cover all of the themes to be validated (e.g. procedures, human-machine interface, team organisation) as well as the different possible operating situations (e.g. normal operations, emergency operations). She emphasised that the scope and level of representativeness of the evaluation campaigns must be consistent with the schedule milestones (e.g. design phases, availability of operating procedures). Toward this end EdF carries out its validation in an iterative fashion using a series of complimentary tests (Labarthe et al., 2011). By working with the regulatory agency and following this iterative process a reasonable level of confidence is achieved in the operating systems performance. Dr De la Garza explained that as a result of this approach the ISV programme is agreed upon by the regulatory authority before it is implemented by the vendor, thereby substantially reducing the downstream risk.

EdF's evaluation campaigns are based on scenarios that consist of real-life situations (i.e. operating conditions or activities). The principal criteria used for selecting the activities used in the campaigns are frequency of use, operating issues, importance, level of innovation, level of complexity, and variety of

4. Presented on behalf of Jean-Paul Labarthe and Louis Graglia.

resources used. To ensure the activities remain as plausible as possible subject matter expertise is solicited from the facilities in operation. Dr De la Garza said that an evaluation campaign generally consists of 5 to 10 scenarios, with each scenario typically lasting between 3 to 4 hours during which time the team addresses a variety of operating activities.

With regard to the number of crews, Dr De la Garza opined that a minimum of 3 teams should be used per campaign, with all teams acting out the same scenarios, but preferably in a different order to limit learning biases. During preliminary validations the teams are not known, so a crew will generally consist of experienced operators recruited from other plants. For final validation however, the teams recruited and trained for plant start-up are used whenever possible.

Dr De la Garza concluded her presentation by offering a number of measures that can be used to safeguard the running of a campaign. Recommended protective measures include early involvement of the designers in selecting and defining the scenarios; the conduct of pilot tests to refine the evaluation process; technical validation of the simulator, including all scenarios, to ensure the simulator is adapted to the operating resources; the development of extra scenarios as backup to safeguard against unforeseen technical problems associated with the simulator; and finally ensuring that there are sufficient reserves built in to the schedule so scenarios can be replayed in the event of a problem with the simulator or the crews.

Challenge presentation for Topical Session 3 (Emilie M. Roth, Roth Cognitive Engineering)

Dr Roth's presentation challenged participants to construct validation scenarios that are designed to elicit human performance responses similar to those that one would expect in real-world situations. According to Dr Roth, human performance in real-world work settings is a function of three interacting influences: 1) individual and team cognitive and collaborative factors; 2) situational complexities in the unfolding events; and 3) attributes of the available support artifacts (e.g. displays, procedures). In order to design validation scenarios that are authentically accurate, the scenarios must be sensitive to and representative of the convergence and interplay among the three elements of the cognitive triad. Dr Roth explained that in actual accidents that involve complex systems (e.g. Fukushima Daiichi), it is the confluence and interplay of these three complicating factors that could challenge the operators and lead to performance vulnerabilities. She explained that the challenge to the validation team is to identify operating conditions and to design scenarios that go beyond the routine textbook cases and sample the "edge cases" – "those demanding decision-making situations that constitute the 'edge' of human performance that could lead to performance degradations".

Dr Roth explained that using scenarios with realistically demanding conditions that challenge the crew's cognitive and collaborative processes is particularly important when evaluating the operator's ability to perform critical human actions, such as those identified in the PRA. While such actions may be fairly straightforward to accomplish in a typical 'textbook' scenario, she explained they become significantly more challenging under conditions that stress or push the edge of the human-machine system.

In order to achieve reasonable confidence that the system will operate resiliently in the face of the complicating situational factors that challenge crew performance in real life ways, Dr Roth posited several challenges that will need to be addressed in addition to the development of edge case scenarios. These involve the pragmatic challenges associated with the design and use of edge cases, including the need to:

- develop better analysis and modelling tools to define the edge cases;
- expand beyond the control room to the larger emergency response organisation;
- extend the operating conditions under evaluation to include severe accidents; and

- overcome the limitations in existing simulators for testing edge cases, including the practical limitation of time availability for evaluating edge case events.

Dr Roth concluded her presentation by suggesting some approaches for identifying edge cases, which can be found at the intersection of the crew/situation/technology cognitive triad. One approach is to start with the critical human actions, similar to the approach used in HRA, and then use experts to decide what could challenge the actions (e.g. failure of automation). Additional methods include the conduct of operational experience reviews, lessons learned analysis (e.g. from industries outside of nuclear power), cognitive task analysis, and defining generic 'complicating situational factors' that can be used as seeds for identifying domain complexities to incorporate in test scenarios, a list of which she provided in her presentation.

3.3.2 Discussion highlights

Challenge 3: What methods might be used to develop scenarios that maximise the amount and relevance of information in support of the validation conclusions and achieving reasonable confidence?

The breakout discussions from Challenge Session 3 are summarised under the following three topic areas: sources of scenario content, content recommendations, and methods and frameworks for developing scenarios/sets. For some of the topic areas, the workshop participants' comments are further organised into more specific themes that the working group identified as it reviewed the input it received from the workshop participants. The majority of the discussion focused on approaches that are equally applicable to both preliminary and final validation. Exceptions to this general principle are noted, where applicable.

Sources of scenario content

The participants identified a number of sources that could serve as the basis of input to scenario development to ensure coverage of the important operational conditions that comprise a test scenario. To this end, the most frequently suggested sources of scenario content included:

- consideration of the "objectives" of the ISV (e.g. modification vs. design);
- regulations and statutory requirements;
- guidance documents (e.g. NUREG-0711) and standards (e.g. those of the Institute of Electrical and Electronics Engineers and the International Electrotechnical Commission);
- expert input, such as from simulator instructors:
  - Using trainers could be challenging, however, since they are accustomed to testing operator performance not the performance of the integrated system. It will therefore be necessary to train the trainers for the evaluation.
- operating experience;
- training data (e.g. difficult tasks);
- lessons learned databases;
- credible worst case operational conditions;
- tasks checklists (normal and abnormal operation);
- PSA/PRA/HRA results to show what is important from the plant hardware perspective (e.g. what systems and operator actions drive core damage frequency).

Content recommendations

In terms of scenario content recommendations, the participants offered a variety of suggestions when it came to the dimensions that should be addressed and the content that should be covered by the individual test scenarios. Some believed that the scenarios that are worth testing are the ones that:

- have an absence of key information;
- exhibit conflict between automation and operator intent or desire; and
- cause operators to have to plan and then re-plan.

Participants stated that NUREG-0711 Revision 3 is quite adequate and should be used as the selection basis for building scenarios as it provides an already established and widely accepted sampling basis. The guidance document lists a total of 15 attributes or considerations that are grouped across three primary dimensions (i.e. plant conditions, personnel tasks, situational factors), which can serve as the foundation for identifying and selecting the operational conditions for integration into scenarios.

Regarding scenario content, the workshop participants provided several specific recommendations, including:

- range of operations that are representative and realistic (e.g. normal, emergency, severe accident situations);
- novel/revolutionary design features;
- remote shutdown;
- hazards (e.g. fire);
- links to areas/work outside the control room;
- degraded instrumentation and control (I&C) and consequent impact on HSI;
- important human actions, including both risk-important human actions and deterministically identified important human actions;
- edge cases (i.e. those demanding decision making events that constitute the edge of a human-machine system that may represent potential weaknesses).

The proposal to include edge cases stimulated considerable discussion among the participants. One participant identified the H.B. Robinson fire of 2010 as a potential example of an edge case, noting that it was not anticipated and asserting that if crews can be shown to handle such events it gives confidence that the design is robust. However, a clarification was put forth that unexpected events – e.g. beyond-design-basis (BDB) events – are not necessarily edge cases. For example, there can be edge events from a plant perspective, or from a human response perspective. It was proposed that in selecting edge cases it is important to understand the cognitive and collaborative challenges the impairments pose to the crew members. This is where the ability to translate between cognitive and collaborative challenges and plant scenarios, and vice versa, is important – to ensure the scenario selection provides adequate coverage for the potential challenges. Thus, it was recommended that for testing purposes, edge cases should be defined in terms of operator challenges (e.g. cognitive/workload/team-collaborative/planning).

Some participants raised the concern that it is very difficult to ask a vendor to perform validation for BDB edge cases, as it is by definition boundless (i.e. outside the design box). Moreover, it is not clear what to do with the results; perhaps the design changes would be prohibitively expensive. In addition, it was

observed that whereas one could argue that being able to handle edge cases implies the plant will perform well for the design basis, but that is not clear. It was pointed out that in the US, "station blackout" is an example of an edge scenario, but current guidance does not require validation testing. Alternatively it was noted that impairments in plant systems in conjunction with staffing challenges are included as part of the testing, and some of these scenarios perhaps constitute edge cases. As part of these discussions it was debated whether scenarios should be representative cases (i.e. realistic) or edge cases similar to a 'stress test' that push the limits of the system? One argument was for selecting scenarios that are representative because, noting that ultimately we need to generalise. Another argument was that safety is the ultimate goal; therefore, it is most important for the scenario to be informative, not necessarily representative. There was a view that edge events may be valuable for generating a lot of information through mistakes.

However, the ability to capture the unexpected using our current deterministic and analytic tools was questioned, with a participant noting, "It's a paradox that we want to engineer the unexpected." Many participants believed that we likely need methods to help develop scenarios that are a compromise between representative and realistic scenarios and edge cases. Similarly, it was a common view that we need to establish a framework for defining and selecting edge events.

Methods and frameworks for developing scenario sets

The majority of the discussions related to the methods and frameworks for developing ISV scenario sets can be categorised as falling into one of the three following areas: ensuring coverage of the important elements, methodological considerations, or a general framework for planning.

Ensure coverage of important elements

- Identify, prioritise, and track the interactions (between people, and between people and machines) that are most important. The ISV plan should test those interactions more frequently.
- Establish a methodology and selection criteria for determining which system, organisational element, or human failures require validation testing.
- Assess the cognitive and collaborative challenges posed by plant activities (e.g. operational conditions).
- Translate the cognitive and collaborative demands into scenarios. The challenge in doing so is to define the combination of scenarios and crews in order to demonstrate that the plant can be run safely.
- Verify scenarios to ensure they address all necessary areas (e.g. I&C, maintenance) and dimensions of operational conditions (e.g. plant conditions, personnel tasks, and situational factors).

Methodological considerations

The workshop participants identified several methodological approaches to developing scenarios:

- Capitalise on operating experience lessons learned in scenario selection.
- Use PSA/PRA/HRA in scenario design, even in preliminary validation stages.
- Use a requirements-based approach to scenario design for preliminary validation in which human-performance requirements guide the selection of operational conditions.

- One participant noted that a flowcharting methodology known as Business Process Model and Notation[5] has been successfully used in the steel industry for graphically identifying and portraying who is doing what, when, with what for emergency response and that this approach may be similarly useful in developing scenarios for edge case and beyond design basis events.

General framework for planning

- Use a multidisciplinary team to brainstorm the approach to be used for developing scenarios.
  - Structure is important. Must have some type of idea or general conception of how it should work; a person with the 'big picture' so things do not get out of hand.
  - Human factors specialists typically do not have the depth of knowledge to fully develop a scenario. Will also need trainers, operations staff, and procedure writers.
- Build a dialogue with the regulator to ensure there is general agreement on the scenarios and whether there is adequate coverage from the regulators perspective.
  - This approach depends on the regulatory framework. In some jurisdictions there is ongoing dialogue with the regulator. In other cases the regulatory review is post hoc.
- Build a framework or matrix to establish combinations of the critical elements of scenario content. The dimensions of the framework/matrix could include:
  - plant process and state;
  - I&C and HSI status and availability, including degraded I&C;
  - manpower availability, including minimum staff complement;
  - environmental factors;
  - complexity/novelty/safety-criticality of the operational conditions (e.g. long scenarios (4 hours or more) with multiple operating conditions.
- Develop a set of scenarios representative of operating conditions (see Sources of Scenario Content and Content Recommendations for details).
- Develop scenarios that cover as many accident sequences as possible.
  - develop scenarios which deal with as many components in the control room as possible;
  - building diversity into the evaluation is one way of developing a scenario that gives you the most information;
  - any one method used over and over is likely to leave you with blind spots.
- Use simplified operational test procedures for preliminary validation.
- The validation test for minimum staff compliment should take place during preliminary validation.
- Design tests to gain the most information by creating scenarios with maximum entropy (i.e. more uncertainty). For example, if you expect something to happen and it happens then the event yields little information. However, if something happens that you do not expect, then that event yields greater information. Tests involving maximum entropy are useful for preliminary

5. https://en.wikipedia.org/wiki/Business_Process_Model_and_Notation

validation where you want to learn from failure, opposed to final validation where you want success.

  - One approach to designing maximum entropy scenarios involves setting situation awareness (SA) to zero.
  - Such tests involve knowing where the critical line is that you do not want to cross and measuring the true margin you have based on task load.
  - If you want to know the margin, you have to push it to fail (e.g. stress test).
  - Caution is warranted with this approach. Using failure thinking rather than success thinking could be problematic for the regulator.

- Consideration should be given to running scenarios throughout the design process. To this end there is a continuum of evaluation tools that can be utilised, including:
  - Tabletop analysis with checklists for vendor equipment. Tabletop analysis should be performed with the "right people" (e.g. operators, PRA/HRA experts, personnel doing safety case) to make sure all-important scenario components are included.
  - Engineering simulation for subsystem validations.
  - Mockups for component and workstation usability tests.
  - Integrated testing at a fully functional training simulator.
- Many of the workshop participants opined that the availability of simulators is a limiting factor that has an impact on scenario selection.

3.3.3 Recommendations

Recommended practices

- Capitalise on operating experience lessons learned in scenario selection**.**
- Identify, select, and develop a general framework for scenario design. Examples include:
  - plant process and state;
  - I&C status/availability;
  - HSI status/availability;
  - manpower availability;
  - environmental factors;
  - complexity/novelty/safety criticality (e.g. long scenarios (4 hours or more) with multiple operating conditions)**.**
- Use a multidisciplinary team to:
  - brainstorm the evaluation approach; and
  - develop scenarios.
- Perform tabletop analyses with checklists prior to simulation.
- Validate the minimum staff compliment during preliminary validation.
- Use engineering simulation in conjunction with tabletop analysis.

- Develop a set of scenarios representative of operational conditions based on:
  - PSA/PRA/HRA;
  - normal, emergency, situation awareness (SA) situations;
  - hazards, e.g. fire;
  - links to outside control room;
  - the impact of degraded I&C on HSI.
- Define the edge case in terms of operator challenges (e.g. cognitive/workload/team/planning) for testing purposes since unexpected events may or may not be edge events.
- Pilot test the scenarios and build margin into the testing for the unexpected since things do not always run smoothly in ISV:
  - ensure collecting the performance data needed;
  - good time to do observer training.
- Ensure that evaluators, especially those with a training background, understand that they are evaluating the design, not the crew.
- For final validation, identify the interactions not tested during subsystem validations and develop scenarios specifically to test those interactions.
- Gain agreement between the regulator and vendor/licensee on the scenarios to be tested.
- Because there are many available methods for scenario generation there may be some value in compiling a comprehensive list of these into a best-practices guide or standard.

Recommended research

- Develop a standard set of acceptable scenarios for boiling water reactor and pressurised water reactors (PWRs) that include the events that are common (e.g. steam generator tube rupture events for all PWRs). A standard set could then be adapted for specific considerations for each design, thus saving licensees time and money in the scenario generation process. However, some participants expressed concern about protecting the security of these standard tests and the consequences for the integrity of the validation if operators were to know what was in them.
- Investigate how to use modelling to identify challenging scenarios and identify gaps in models?
- Develop a method for translating cognitive and collaborative demands into operational scenarios.
- Develop guidance concerning how to balance representative (realistic) and edge case (stress test) scenarios in the testing strategy.
- Investigate whether the ability to withstand edge events is a good indicator of the system's ability to perform well for the design basis.
- Revise NUREG-0711 to clarify the types of scenarios that need to be tested (e.g. whether edge event scenarios should be tested).

3.3.4 Related white papers

Two white papers were submitted that addressed the topic of validation test scenarios. These papers are summarised briefly below. The full texts of these papers are presented in Appendix B.

Test scenarios and test design by Kenji Mashio. Mr Mashio's white paper examines the five factors that he feels are crucial for a successful verification and validation with respect to test scenario design and test development. The first factor is the 'sampling dimension', which involves a representative sampling of the plants operating conditions (e.g. normal operations, accidents), operator actions (e.g. important human actions, credit manual actions), and situation factors (e.g. environment, workload). Next is the 'scenario build' factor, which Mr Mashio describes as the coordination scheme used to organise the operating conditions and operator actions into a range of sampling dimensions for subsequent integration into scenarios. The third factor is 'design verification'. In a manner similar to NUREG-0711, design verification involves a two-step process consisting of a verification of availability (i.e. HSI Task Support Verification) and a verification of suitability (i.e. HFE Design Verification). Because neither of the two types of verifications is intended to be an evaluation of system performance, both analyses can be conducted using desktop paper and pencil techniques. Of the five factors addressed in the paper, Mr Mashio reserves the most discussion to the forth factor, Integrated System Validation. He begins by providing the NUREG-0711 definition of ISV, followed by a discussion of human performance in the context of a system test environment and the important role performance measurement-type, – method, and – criteria play in achieving a successful ISV. The paper concludes with a brief discussion "test design", which Mr Mashio describes as the final logistics for integrating the testbed, scenarios, and task measurement accommodations into a test procedure for governing the test process in detail.

Discussion of ISV Scenario Scope with Operating Process Element by Yun Goo Kim. Dr Kim's white paper begins with an overview of the standing of their nuclear power plant (NPP) new build programme in Korea, including some insights into ISV planning, followed by a discussion about the scope of their ISV and limitations of the scenarios. Korea has 23 NPPs in operation and five under construction, two of which have already completed their ISVs. The ISV's occurred on a full scope simulator and lasted 3 weeks. The schedule called for 3 crews. Each crew participated in a one-week test involving 7 scenarios, with each scenario lasting between 1-2 hours. Dr Kim explained that the ISV scenarios included as many operational conditions as possible, but due to schedule limitations, it was not feasible to include all of the conditions (e.g. safety-related operator actions). To make up for this drawback they conducted a series of walkthroughs on the operational conditions that could not be included in the ISV scenarios. Some of the operation process elements could be validated utilising the representative ISV (i.e. operational walkthrough) such as monitoring, detection, and response implementation. Validation of the remaining elements (e.g. situation awareness, response planning, teamwork) utilising this approach would require a deeper dive, and Dr Kim posited a number of questions that would need to be addressed before the validity of this surrogate approach to ISV could be known.

3.4 Challenge 4: Analysing the results and drawing conclusions

There is one point on which there appears to be near universal agreement among individuals that have been involved in the validation of nuclear power plant control room designs or the validation of similar complex systems; these efforts can produce huge quantities and diverse types of data. There is much less consensus regarding how to analyse and draw conclusions from this data, particularly where the decision concerns the acceptability of the design for operation. In developing its plans for this workshop, the working group identified a number of questions pertaining to this phase of the validation effort. These questions included:

- Are inferential statistics meaningful in the context of MCR validation? If so, what practical methods might be used to gather sufficient data during the human factors validation of the MCR to support the legitimate use of inferential statistics?
- What substitutes might be proposed as alternatives to traditional statistical modelling approaches?

- How might qualitative measures be analysed or assessed to most effectively support achieving reasonable confidence?
- Is it possible (and desirable) to establish standardised acceptance criteria for the human factors validation of MCRs?
- Is there a rational basis for treating certain measures or data as only diagnostic while using others for pass/fail decisions during the human factors validation of the MCR?
- It is recognised that the validation process may result in changes to the integrated system (e.g. to the procedures, operator training, or a system interface). What should be the process or rationale for determining whether such changes should be evaluated through additional validation exercises?
- What should be the appropriate scope, format, and level of detail to be included in the final validation report?

Considering these and similar questions related to control room validation, the working group elected to pose the following challenge question to the workshop participants.

How should the validation results be aggregated and analysed to determine the final validation conclusions? Are inferential statistics meaningful in the context of MCR validation? If not, why not? What substitutes might be proposed as alternatives to traditional statistical modelling approaches?

Consistent with the format of Topical Sessions 1-3, workshop participant discussion of challenge question 4 was preceded by presentations from two invited speakers who provided additional background and perspectives for consideration during the session. These speakers were Gyrd Skraaning Jr., of the Organisation for Economic Co-operation and Development (OECD) Halden Reactor Project and Robert Fuld of Westinghouse Electric Company.[6]

3.4.1 Summaries of challenge presentations

Note to reader: Dr Skraaning's and Mr Fuld's presentation materials can be found in Appendix C.

Challenge presentation for Topical Session 4 (Gyrd Skraaning Jr., OECD Halden Reactor Project)

Dr Skraaning's presentation challenged participants to consider the presumptions of prescriptive (process based) validation methods which he identified as including the presumptions that: 1) human performance requirements can be fully understood, pre-defined and specified; 2) correct execution of prescribed process results in clear and convincing validation evidence; and 3) the procedure generalises across validation contexts. In contrast, he put forth the hypothesis that prescriptive validation methods fail to provide "…the statistical and logical bases for determining that performance of the integrated system is, and will be acceptable" (NUREG-0711 rev3, p. 93).

Dr Skraaning noted that to judge the outcome of a final control room validation, human performance scores from simulator trials have to be interpreted in light of established acceptance criteria. Thus, some observed performance scores are considered acceptable, while others may be unacceptable. However, he asserted that this is only a first necessary step to judge the acceptability of new control room designs. In the later stages of the data analysis and interpretation process, a myriad of detailed validation results have to be organised, weighted and judged together from multiple angles to reach a conclusion on whether the control room is, and will remain acceptable. He noted that this process is similar to a trial court or a safety case,

6. Biographical sketches of the speakers are available in Appendix A.

where structured arguments are used to evaluate a complex body of evidence in order to reach a clear and definitive conclusion; typically in the absence of formal and prescriptive methods. He in turn suggested that there is no universal formula or predefined psychometric procedure that can help us to reach overall conclusions on the acceptability of new control room designs.

Consistent with this line of reasoning, Dr Skraaning advocated that one should avoid simple approaches to acceptability judgment where the validation team checks if acceptance criteria are met for a selection of human performance measures. He asserted that such psychometrically oriented methodologies may enhance the interpretability of the observed human performance scores, but is useless during the comprehensive decision making process where detailed validation results are compiled, prioritised and compared to reach a trustworthy decision on acceptability.

Beyond skepticism: Induction, falsification, and reasonable confidence in MCR validation (Robert Fuld)

Mr Fuld began his presentation by highlighting difference between the processes of verification and validation. These differences, which are along multiple dimensions, include the processes' empirical goals (truth vs. strength), the types of reasoning they require (deductive vs. inductive), and whether one can hope to achieve a complete answer (yes and no, respectively). The comparison illustrated not only the differences but also highlighted the challenges of undertaking system validation. However, he noted that validation is necessary because, in part, adequate strength of the system must be demonstrated in order to put systems into service.

Mr Fuld suggests that we should not be skeptical of validation outcomes if: 1) we prepare sufficiently (e.g. per NUREG-0711); 2) we correct problems found during the process and progress incrementally to a refined state; 3) we accept the simulator modelling of safety parameters; 4) we identify a representative test set to rigorously challenge the system on expected (and other) operating conditions; and 5) we pass or fail on safety criteria, which are objective, conservative, and independently established and confirmed. Rather, he proposes that one approach validation from a "falsification" perspective in which a successful validation is the null result of passing all safety criteria and a failed validation is the failure of some safety criteria. Passing provides little information and only corroborates the adequacy of the current design (which may still fail someday), whereas failing 'proves' that the current design is not yet adequate. Under this line of reasoning, reasonable confidence in successful validation outcomes comes not from the accumulation of successful outcomes (e.g. through repetitions) but rather from performing appropriate preliminaries (i.e. design development).

Mr Fuld's presentation also touched upon the use of pass/fail and diagnostic criteria. He advocated pass/fail criteria to be based on safety limits and as few as reasonably possible. Diagnostic criteria should be as diverse as possible. He also proposed that signal detection theory might provide a useful model for the analysis of convergence between results for pass/fail and diagnostic measures. On the topic of risk-important human actions (RIHAs), he asserted that whereas RIHAs are a convenient sample of human actions (HAs) (e.g. for timing) and give insight on accuracy of PRA estimates; they are rarely related to safety, may not be significant to plant risk or to acceptable PRA results, may change entirely with the PRA, and may invite negative training. Accordingly, he proposed that RIHAs be used only as diagnostic criteria, and not as Pass/Fail criteria.

3.4.2 Discussion highlights

Challenge 4. How should the validation results be aggregated and analysed to determine the final validation conclusions? Are inferential statistics meaningful in the context of MCR validation? If not, why not? What substitutes might be proposed as alternatives to traditional statistical modelling approaches?

The discussion of Challenge 4 is summarised under the three major elements of the challenge question: aggregating the results, analysing the results, and drawing conclusions. Within each of these elements the workshop participants' comments are further organised into more specific themes that the working group identified as it reviewed the input it received from the workshop participants.

Aggregating the results

It was noted that there is variability in both crews and the scenarios and concern was expressed regarding whether it was possible to get to the heart of variability in ISV context. It was also observed that analysis of the large amounts of data that can result from an ISV is resource intensive and time-consuming.

Scope of the acceptance determination

- Aggregation should be done carefully by an inter-disciplinary group and with collection of all evidence (i.e. not just pass/fail).
- Pass/fail measures are the first step and only one piece of the picture. It is necessary to build a case by using all the evidence, including human performance measures, to ensure the integrated system is validated. The key is the integrated system.
- Qualitative data and measurements of multiple constructs are likely to be used in preliminary validation. We should develop better methods to support use of multiple and diverse sources and types of data in making acceptability determinations.

Quality and specification of the evidence – Individuals noted concerns and or the need for guidance in the following areas:

- What is reasonable evidence and sufficient quality of evidence? For example, what is sufficient variability of teams/operators and scenarios?
- Quality of evidence can be assessed by looking at consistency between various types of evidence and their inferences and relationships. Does it meet expectations for convergence or divergence with variation in conditions (e.g. workload)? The trends should be consistent with expectations (e.g. deterioration in performance with increasing workload). There can be evidence from various sources and they should be consistent – observers, documentation, measures, etc.
- How to achieve statistical relevance (inference) with a low number of crews? Can advanced statistics help? How to deal with variability in performance in scenarios?
- Can we get closer to statistical significance by doing more, smaller trials?
- Determining the number of observations that are sufficient and ensuring that the amount and type of evidence is commensurate to the question at hand and to the expectations of the reviewers.
- Adequately testing the variability in human performance through repetition – people and teams of people can be highly variable.

Crediting preliminary validations

- Crediting preliminary validation as part of the ISV should be considered, but the best way to go about doing this remains to be determined.
- Formalising preliminary validation to make it a more prescriptive process, such as part of the "evidentiary approach" in helping to make a "safety case" sounds promising.

Using data from the plant life cycle – Some workshop participants advocated the potential use of data from the full life cycle of the plant.

- Currently, data is collected during operator licensing exams, operating data, start-up testing emergency drills, etc. To the best of our knowledge this data is never associated to validation data to see if the conclusions of the ISV and PV work are actually supported. Mining this data could be very useful in predicting and preventing human error, but practically speaking, it would need to be done by researchers outside of the design teams. Realistically, the design team is going to do the tests they need to do to solve their design issues and move on; so motivating industry to produce/keep/share this data with research entities will be key.

Analysing the results

Top-down and bottom-up approaches

- Unexpected results cannot be covered by predefined criteria. Need to include bottom up analysis of collected datasets for the purpose of capturing unexpected issues.
  - This also relates to a staged evaluation model with a graded approach and periodic regulatory reviews. For example, start with regulatory review of the operational philosophy, then periodic review of design and preliminary validations, and finally the ISV.
  - Graded approach and accommodating a bottoms-up evaluation as per the safety case methodology in final validation, implies that you need a periodic review process (phased evaluation) not just acceptance of the ISV.

Use of statistics – Differing views were offered concerning the role of inferential statistics and the appropriateness of using descriptive statistics to form inferences about performance. The workshop participants' comments concerning statistics included the following:

- Most inferential statistics may not apply.
- Logic behind hypothesis testing does not apply.
- Inferential statistics and descriptive statistics have a role in validation, primarily in preliminary validation.
- Statistics can be used to analyse the aggregation of longitudinal data, which in turn can also serve as "evidence" to support arguments in making a "safety case".

Alternative approaches

- Equivalence testing, structural equation modelling, and response surface methodology are possible alternatives to traditional inferential statistics.
- Bayesian Inference may be a more appropriate method than what is currently used.

Drawing conclusions

Acceptance criteria vs. evidentiary approach – Views were expressed that both use of acceptance criteria and the evidentiary approach are useful. In discussing the evidentiary approach, the participants offered the following comments:

- The analogy of a court case for assembling the validation evidence to demonstrate safety beyond a reasonable doubt is interesting. The burden is on the regulator to be a qualified judge, but it can be supported by independent review by expert bodies. You could have a judicial panel with different expertise.
  - Technical support organisations can provide independent expert assessment and opinion.
  - Care would be necessary if independent bodies are providing certification and one must question if they improve the integrity of the programme. The approach could promote independence in validations but also add uncertainty in their integrity, performance, etc.
  - Independence is required between the designer and validation process, and therefore, there should be independence between initial and final review.
  - In this regard, does the regulator have to ensure independence between staff following the design and those evaluating the design (similar to independence between designers and those conducting the validation)?
  - In developing a case, the testing should move from high entropy to low entropy.
  - The falsification process is similar to presumption of innocence – you have to prove failure or guilt.
  - The metaphor breaks down when one considers the adversarial nature of a court case.
- Judging acceptability
  - If a crew does not meet a mission objective (see functional decomposition model) it is a validation failure.
  - At the second level of the functional decomposition model (i.e. human performance requirements – see Figure 4.1), if there is a "major" human engineering discrepancy, the team could be split on whether it is a validation failure.
  - If there is one failure and there is no pattern, it is an aberration, it does not fail the mission. If there a fundamental flaw in this design, some pattern of evidence that supports failing the design should exist. If there is no pattern, then look for things that can be fixed or improved upon in the current design.

Multi-path approach

- The multi-path approach to analysis (Skraaning and Stine, 2015) separates data analysis (e.g. looking for systematic patterns, aggregating data at multiple levels) from acceptability analysis (i.e. the approach discriminates between how we judge observed performance scores and the judgment of the acceptability of the overall system).

3.4.3 Recommendations

Recommended practices

- Analysis should include evaluation of convergence–divergence of results.
- Perform a bottom up analysis to address unexpected results.
- Include periodic/phased evaluations of results.
- Use ISV results to inform the plant PRA.

Recommended research

- Develop methods/guidance for application of statistics to ISV results and for reporting the results in a form that is understandable.
- Evaluate equivalence testing, structural equation modelling, response surface methodology, and Bayesian statistical approaches as potential alternatives to traditional inferential statistics.
- Develop guidance for using preliminary validation results to inform the conduct/conclusions of the final integrated system validation.
- Develop a means to verify validation results through plant life cycle data.

3.4.4 Related white papers

Among the white papers submitted by workshop participants were several papers concerning analysing and drawing conclusions from validation data. These papers are summarised briefly here. Except where noted otherwise, the full texts of these papers are available in Appendix B.

Integrated System Validation (ISV): The Acceptability Analysis Process, by Dr Gyrd Skraaning Jr.[7] Dr Skraaning's paper outlines an approach to support decision making related to the results of control room validation. The white paper follows from the reasoning presented in his challenge presentation and advocates the use of a separate acceptability analysis stage of the validation as a possible alternative to the use of pre-determined acceptance criteria.

Recommended Research on Data Analysis and Drawing Conclusions, by Dr Nathan Lau. Dr Lau's white paper provides an overview of the practical and technical constraints on conducting performance-based validations of MCR designs, including crew and scenario sampling. He notes that these constraints have implications for our ability to generalise from the performance tests but due to the lack of applicable formal methods, the content validity or comprehensiveness of performance testing is difficult to assess. He also touches upon other challenges, such as those affecting the quality of expert judgment, which are frequently used as performance measures in validation efforts. Dr Lau proposes that prior to formulating any research programmes to improve confidence; the nuclear community must answer "Drawing conclusions for what?" He introduces the notion of "consequential validity" and suggests that whereas all evidence that could support prediction of future performance should be admissible, not all individual pieces of evidence are equal in merit. Dr Lau proposes that research can provide an empirical foundation for estimating the "errors" or variability associated with various performance testing conditions and thereby a means to improve confidence in ISV conclusions and decisions. He concludes by describing the

7. For a full text of Dr Skraaning's paper see Proceedings of the 9th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technology.

merits of, and a strategy for, establishing a "databank" of simulator training data that could be used to study measurement variability and to provide reference points for data collected through ISV.

Is Structural Equation Modeling as a component of the Integrated System Validation Toolbox an alternative to traditional statistical modeling approaches?, by Yvonne Johansson and Martin Castor. Ms. Johansson's white paper briefly describes an effort undertaken by the Swedish nuclear regulatory authority to investigate statistical modelling methods, such as structural equation modelling (SEM), as potential components of an ISV analysis toolbox. She notes that SEM is a second generation statistical analysis method that combines the benefits of path analysis, factor analysis, and multiple regression analysis. Ms. Johansson concludes that SEM is not a "silver bullet", but represents a powerful statistical analysis method which is useful for analysis of large and heterogeneous datasets from operational settings and processes.

4. POST-WORKSHOP ANALYSIS

As a whole, the workshop resulted in a substantial compilation of information and professional viewpoints concerning the validation of main control room systems and modifications. These inputs came in the form of the plenary and topical session presentations, multiple white papers, and the three days of discussions amongst the nearly 30 professionals that gathered in Charlotte. In this post-workshop analysis section we attempt to look at these contributions collectively to identify the key theses and potential paths forward toward achieving reasonable confidence in control room validations.

4.1 Establishing reasonable confidence

The workshop discussions emphasised the importance of validation in determining the acceptability of the control room design and provided an opportunity for experts to share experiences and concepts for improving the validation process. The discussions centered on the workshop's main theme of "establishing reasonable confidence in the human factors validation of main control room systems of nuclear power plants". Since the purpose of validation is to determine that the design "meets performance requirements and supports safe operation of the plant", we need to know that the validation process can support making that decision and when sufficient information is available to do so. The question of reasonable confidence is a complex one; and to address it, we must parse it into two considerations: what contributes to confidence in validation conclusions and at what point is that confidence sufficiently reasonable. In other words, are we addressing the right topics and have the topics been sufficiently investigated.

With respect to establishing confidence in validation, the workshop results suggest two general principles:

- Use a validation process that conforms to *accepted industry guidance*.
- Attend to the *specifics of the actual system* and its context.

One theme of the workshop discussions was the importance of using *accepted industry guidance* in conducting validation. There are consensus industry standards that provide guidance on validation methodology, such as IEC-61771 (IEC, 1995), as well as regulatory guidance, such as NUREG-0711 (O'Hara et al., 2012). Although the workshop touched upon several areas where additional guidance is needed, using widely accepted guidance to develop an approach to validation provides confidence that important aspects of the validation process will be addressed. The discussions during the workshop suggested that for some topics, industry guidance could be developed from recommended industry practice; while for other topics, industry guidance must be based on an improved technical basis. These bases are described under the Section 4.2 Recommended practices and Section 4.3 New directions for research below.

Another theme of the discussions was the need to address the specific characteristics of the plant and systems being evaluated (e.g. whether a new design is evolutionary or revolutionary, the plant's operating philosophy/concept of operations, the modifications to an existing plant, and its cultural and operational context). The scope and objectives of validation are affected by both the extent of changes being made and the specific nature of those changes. The workshop discussed the need to take into account the specific system's operational philosophy/concept of operations regarding what should be evaluated and the

development of performance criteria, thus avoiding the misconception that one validation plan will work for all plant designs. It is important to recognise that each validation is for a particular operating philosophy/concept of operations, which can be specific to a country, or reflect particular constraints that are required by the operating organisation (e.g. level of acceptable automation, crew member versatility and interaction, etc.). Thus the workshop has provided some general principles for establishing confidence in the validation process. While the general principles are a necessary first step, the workshop emphasised that additional guidance is needed before the guidance is sufficient to provide reasonable confidence. Determining when that confidence has achieved *reasonableness is a challenging issue.* This results in the situation where reasonable confidence means different things to different people (Salway, see Appendix B).

In a hypothesis testing environment, reasonableness is quantitatively defined by setting the statistical parameters of the test; e.g. the alpha (probability of a Type 1 error, concluding that the null hypothesis is rejected when in fact it is not) and beta levels (probability of a Type 2 error, concluding that the null hypothesis is not rejected, when in fact it should be) of the test. Knowing the alpha and beta levels of the test, one can decide whether there is reasonable confidence in the results. Further, there are accepted standards for setting these parameters (see Lau, Appendix B, and O'Hara and Higgins, 2015, for a discussion of these concepts as applied to validation; and O'Hara and Persensky, 2011, as applied to design modifications in general). However, this type of hypothesis testing model does not work in the context of validating the overall design or determining its reasonableness. Design validation is a complex decision making activity that is based on actual results obtained and additional considerations pertaining to issues such as how well the selection of scenarios reflects the real operational demands on the design.

Reasonableness addresses a critical aspect of validation in that it provides a basis for determining whether a test plan is adequate and when sufficient testing has been conducted. One common theme discussed at the workshop was the resource and schedule intensive nature of validation (e.g. see the Alonso paper and Hall presentation in Appendices B and C, respectively). Simulators must be developed and be available for validation tests. Operators need to be selected and trained. Scenarios and test protocols have to be developed. Crews perform scenarios with repeated trials and data is collected. The data must be analysed and conclusions regarding the acceptability of performance reached. Any issues encountered must be resolved. Validation is a costly and time consuming process that often occurs near the end of the design phase where pressures to finalise and implement the design are strong (Notte, Appendix B). While we need to know how validation should be performed to gain confidence in the results, we equally need to know when sufficient testing has been performed to know that reasonableness has been achieved. Without such an understanding, the validation team cannot answer such basic questions as:

- Should additional crews be used?
- Should additional scenarios be included?
- Should additional performance measures be taken?
- Should additional analyses be performed?

One can always do more testing, but given validation's cost and schedule considerations, it is important to know when the process has achieved a point where reasonable confidence exists in the conclusions.

The reason validation methodology has yet to reach a point where reasonableness can be defined is that there are too many gaps in the technical bases, available guidance and recommended practices. For example, whereas the guidance may state that a range of operational conditions are needed to validate the system, limited technical basis and little guidance is available to specifically identify how many scenarios

should be used. Similarly, although the available guidance states that multiple crews should be used to account for human variability, little guidance is available to identify how many should be used.

In addition to detailed methodological considerations, the available guidance is limited in other ways reflecting bigger picture issues. To illustrate, consider the guidance provided in NUREG-0711. NUREG-0711 is widely used to guide both regulatory reviews and designer validation plan development. Yet, there are gaps and limitations in that guidance that make it difficult to use for certain aspects of validation. For example, NUREG-0711:

- Focuses on validation at the end of the design process and does not address preliminary validations that are performed during the process and how they contribute to the overall validation results.
- Focuses on the validation of new plants and provides limited guidance on validation of plant modifications.
- Does not adequately address establishing validation conclusions.

These issues were discussed in detail at the workshop and furthered the objective of identifying the research needed to improve the validation process and help better define when confidence in validation results is reasonable.

4.2 Recommended practices

One of the objectives of the workshop was to identify *recommended practices* derived from the experience of the workshop participants in conducting validation. Several participants identified recommended practices in their papers and presentations, (e.g. see Alonso in Appendix B and Fuld in Appendix C), such as:

- Validate modifications even if not related to safety systems because they can change operator tasks, human system interfaces (HSIs), and degrade performance.
- Identify a representative test set to rigorously challenge the system.
- Pass or fail the design on safety criteria, which are objective, conservative, and independently established and confirmed.

Recommended practices were also a main focus of the discussion groups and the recommendations included considerations such as:

- Establishing independence of the validation team from the design team.
- Using pilot tests to ensure the simulator and the methodology perform acceptably before validation trials begin.
- Using a multidisciplinary approach to establishing performance criteria and the interpretation of validation data.

During the workshop, we prioritised the elicitation of recommendations and, considering time constraints, did not endeavor to vet them in accordance with guidance development procedures, such as subjecting them to a consensus process or peer review. Nevertheless, the recommended practices proposed by the workshop participants can be valuable inputs to future guidance development efforts (see Section 4.4, Broader validation challenges, Bullet 1). Accordingly, we have deferred analysis of the recommended practices as we believe this is best accomplished through guidance development procedures.

4.3 New directions for research

While detailed research needs were identified in Section 3 of this report, we would like to highlight at a general level those areas of research likely to be most critical:

- multi-stage validation;
- unique consideration for validating plant modifications;
- testbeds/simulators;
- scenarios;
- performance measures and acceptance criteria;
- validation conclusions.

4.3.1 Multi-stage validation

There was considerable interest, both in the workshop papers/presentations and during the workshop breakout groups, about the need to begin the validation process before integrated system validation (ISV) as a means to achieving reasonable confidence and the need to develop further guidance to support this approach.

Several different ways of describing this concept were used, such as "phased validation", "stepwise validation" and "multi-stage validation". Regardless of the term used, common to the discussions was the notion that validation could be conducted in phases starting with preliminary validations and culminating with integrated system validation (ISV), or perhaps validations conducted after the plant has started operations. Here we have elected to use "multi-stage" as a generic, descriptive term to summarise discussions that focused on these terms and concepts.

Whereas NUREG-0711 and some of the other guidance documents focus mainly on ISV, which is a test conducted towards the end of the design process, many workshop participants argued that the validation process begins early in design process and continues throughout, culminating in ISV (e.g. see the Alonso; and Laarni, Norros and Salo; papers in Appendix B; the De la Garza; Hall; and Norros presentations in Appendix C, and Rivere, 2015). One of the workshop's working groups went even further and suggested a life-cycle approach where validation is extended through start-up testing and operations. The multi-stage approach to validation has already been employed in the validation of new plants as well as modifications (see papers listed above). Designers like this approach because it reduces risk and provides an opportunity to gain confidence in the design. It makes sense from a regulatory perspective as well. The earlier that issues are defined, the more likely it is that effective solutions can be developed to address them.

O'Hara and Higgins (2015) identified some of the technical issues to be addressed for a multi-stage approach to validation:

- How are the stages characterised?
- How to ensure that each stage has validation objectives and methods rather than design oriented tests?
- How do earlier stages of ISV impact the design of the final ISV?
- How are the validation results integrated together across the stages?
- How can prior ISV results from a related, but different design, be used?

Information is available to begin addressing these questions, e.g. Rivere (2015) proposed ways of bridging the gap between validations conducted early in the design process and ISV.

4.3.2 Unique consideration for validating of plant modifications

The human performance challenges associated with plant modifications have been generally discussed by the industry (International Atomic Energy Agency (IAEA), 2010), in case studies (Roth and O'Hara, 2002), and by regulatory bodies (Nuclear Energy Agency (NEA), 2009). A topic of considerable discussion at the workshop was the unique consideration for validating plant modifications. It was noted that there is little guidance for validating modifications and that improvements in this area are needed (see the Alonso and O'Hara papers in Appendices B and C, respectively). Most of the validation guidance available addresses new plant designs. The importance of this topic stems from the fact that most of the control room validations conducted to date have been for plants with modifications. Therefore, it is not surprising that the need for additional guidance for validating modifications has been identified in other validation workshops as well (see Braarud et al., 2010). Modifications may address only part of the control room design and can be implemented in many different ways. For example, modifications can be implemented slowly while the plant is still operational, during a single outage, during multiple outages (thus the plant is operated in a series of interim configurations). They can also be introduced into the control room in parallel with the human-system interface (HSI) being replaced; providing operators with options for using old or new HSI and to become confident in the new design. Thus modifications pose unique challenge to human performance and to the validation process (O'Hara and Higgins, 2015).

When modifications are implemented over multiple outage periods, their validation may necessitate a multi-step approach, where interim configurations are validated and used by operators, as the full modification evolves.

Guidance on validating plant modifications is a pressing need.

4.3.3 Testbeds/simulators

Considerable discussion centered on the importance of testbeds. Much of the guidance in the available standards is based on the assumption of high simulator fidelity. Issues related to simulator readiness, e.g. functionality limitations, completeness issues, and newness of the simulator giving rise to less than optimal performance. Issues also include competing demands for simulator time and timing of availability for ISV (Gibson, 2015). In addition, there are also situations where validation needs to be conducted and no simulator is available, e.g. in the design and review of fuel-cycle facilities.

A possible solution is the use of a greater diversity of testbeds, an approach that may be necessary for "preliminary validations" conducted as part of a multi-stage approach to validation. One example is the use of non-full-scope simulation capabilities for selected validation objectives. Berntson's et al. (2004, 2006) proposed a three-stage validation that may be useful in considering these situations. The stages include table-top validation early in the design process, a table-top walkthrough using the procedures, and a full operational trial of the final design. Elsner, Freitag, and Rivere (2015) discuss the use of mockups and full-scope simulators in the validation process.

Another possible alternative to full-scope simulators discussed during the workshop is computer simulations of the control room and crew performance. The technology for such simulator models to play a role in design and evaluation is rapidly advancing and they are finding acceptance in the nuclear industry as well as the broader engineering design community. The simulation models can support early design validations as well as helping to better allocate resources for ISV by simulating validation scenarios to determine which are more important to consider in human trials.

A technical basis is available to consider alternatives to full-mission simulation for some aspects of ISV. These alternatives may help address testbed needs and may help reduce pressures experienced by ISV teams for simulator time.

4.3.4 Scenarios

What scenarios should be used for validation? This is an important consideration in determining the cost of validation and confidence placed in the conclusions.

NUREG-0711 provides detailed guidance on the sampling considerations that go into identifying a broad field of conditions from which scenarios can be developed. It was generally concluded that this provides a good starting place, but more guidance is needed to identify specifically the scenarios that should be used for validation trials, e.g. ensuring the right scenarios are used to support validation conclusions concerning the achievement of operational and safety goals. With the need for additional guidance clearly established, the workshop focused on promising approaches to address this issue.

Some workshop participants focused on the "criteria" that can be used to help select from the field of possible scenarios to include in validation. For example, Labarthe and Graglia (see Appendix B) identified the following scenario characteristics as important considerations in the selection process:

- frequency of use (routine operations);
- operating issues (safety, production, etc.);
- level of innovation relative to the existing situation;
- level of operating complexity (accumulation of failures creating a significant workload and requiring problem-solving, etc.);
- variety of resources used (technical systems implemented, types of procedure adopted, etc.).

Similarly, Gibson (2015) provided principles to help selection including requirements, available scenarios, and significant safety events; and Jordi and Elsner (2015) proposed a method to select the most significant operator tasks from the enormous number of possible tasks.

Other workshop participants focused on the need to identify situations that challenge the integrated system. For example, both O'Hara (see Appendix C) and Roth (see Appendix C) proposed the use of "edge" cases. Edge-centered testing is an approach that focuses on testing the demanding decision-making situations that constitute the "edge" of a human-machine system that may represent potential weaknesses. Scenarios are developed to test the edge so that weakness in the design can be addressed.

O'Hara and Roth also suggested the use of generic lists of "complicating situation factors" that are available in the literature (e.g. Patterson, Roth and Woods, 2010).

Roth (see Appendix C) noted the importance of considering the implication of lessons learned from actual nuclear plant events in the consideration of potential challenges a future plant may face. Roth provided the following example:

> "Another important lesson from examination of actual accidents is the need to expand the validation scenario coverage beyond activities in the control room and beyond abnormal and emergency events. The Fukushima event made clear that in serious accidents, the locus of diagnosis, planning, prioritising, and decision making, largely shifts from the control room to emergency response centers. The Fukushima accident, illustrated saliently the need to consider beyond design

basis and severe accidents in defining the set of validation scenarios to achieve reasonable confidence that the support systems in place will enable the broader emergency response organisation to respond resiliently."

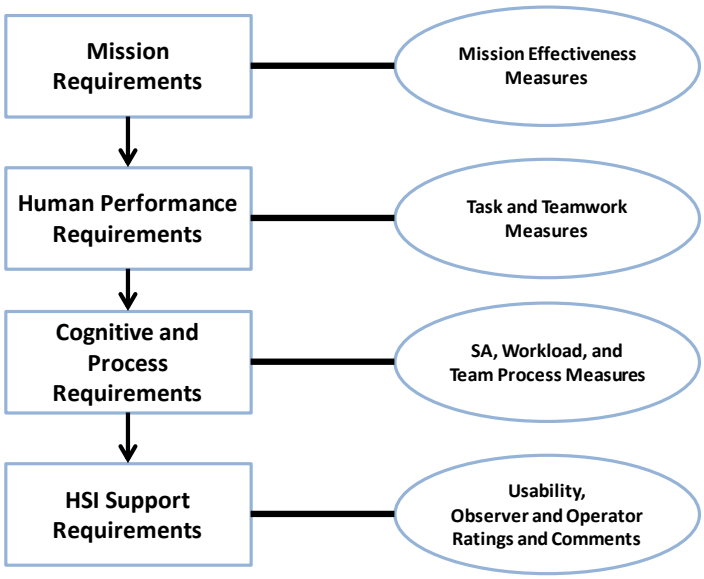
While the workshop provided some insights regarding areas to improve the guidance on scenario development, there remains a need to improve the technical basis for resolving important questions such as how many observations of crew and operator performance are necessary to achieve reasonable confidence. Although there is not likely to be one answer to this question that is suitable to all validation efforts, guidance is needed to help validation teams answer it for their validation test.

4.3.5 Performance measures and acceptance criteria

The performance measures to be used for validation and the identification of acceptance criteria were the subject of considerable debate. The available guidance typically suggests that a multi-variate approach be used encompassing categories of measures for plant, operator, team, cognitive, and physiological performance. Some of the measures are scenario dependent, such as the specific plant measures used to assess performances during specific scenarios, while others may be scenario independent, such as measures of cognitive workload that can be used for any scenario. This framework provides significant challenges to validation teams. There are nearly infinite choices of measures that can be used to assess these categories of performance and additional technical bases and guidance is needed to make decisions on the best measures for use in the nuclear environment and for a specific validation effort.

Several promising approaches to address these questions were discussed. For example, both the Malcolm and Braarud papers (see Appendix C) discussed the importance of using requirements to drive performance measurement selection. This approach was elaborated during the workshop and one working group proposed the use of a hierarchal functional decomposition approach where measures are tied to increasing levels of design specificity (see Figure 4.1).

Figure 4.1: Hierarchical functional decomposition model relating levels of requirements to types of measures



Another approach is to use measurement frameworks that take an integrative approach (Braarud and Rø Eitrheim, 2013). Braarud and Rø Eitrheim's framework supports the identification of performance measures and criteria. The main purpose of the framework is "to support the development of criteria, and to support the development of observations to evaluate to what extent criteria are met" (p. 17).

Yet another approach was the use of usability measures to augment the available measures. The use of usability theory and measures to frame validation evaluation has been developed by VTT Technical Research Centre of Finland Ltd (VTT) (Savioja and Norros, 2013; Savioja, 2014).

One subject that came up frequently during the workshop was operational resilience; the ability of the integrated system to manage and recover from a wide range of events, including novel events a crew has not encountered before. This gave rise to a recommendation to identifying resilience indicators. These are indicators that could predict good practice in future scenarios and provide a basis for predicting performance beyond the ISV scenarios tested. Additional research in this area could provide a basis for enhancing confidence in integrated system validation.

These are some of the recommendations stemming from the workshop discussions for developing performance measures and criteria. Additional development of these concepts will provide a basis to improve industry practices and the available guidance.

4.3.6 Validation conclusions

There was much discussion at the workshop on the challenges associated with analysing the large amount of data collected during validation and processing it to arrive at conclusions regarding the validation of the design. We have divided this discussion into these two parts: analysing the data and arriving at validation conclusions.

Analysing data

When validation trials are completed, the validation team is left with a very large amount of data. The data often reflect performance of individual crews across a range of scenarios, repeated trials of the same scenario by several crews, and non-overlapping data sets (not all crews perform the same scenarios creating a crew by scenario confound). Some measures are qualitative (e.g. observational assessments of subject matter experts or operator opinion) and others are quantitative (e.g. simulator parameters). Some of the data are expected to correlate well, providing an opportunity to establish convergent validity. The number of crews is usually fairly small, and relatively speaking, the number of scenarios fairly large. How to analyse this data to arrive at a simple "validated or not" is a daunting task.

The available validation guidance provides little to no support for addressing this important aspect of validation. The discussions casted doubt on the use of traditional parametric statistical approaches to data analysis, since they are rooted in hypothesis testing and assume much "cleaner" experimental paradigms and much more focused analyses. In addition, they were essentially designed to answer a different question. Indeed, validation is not an experiment.

This led the discussion into the possibilities of using more novel approaches. Alternatives were discussed in the plenary/challenge/white papers as well. We will illustrate some examples of the alternative approaches discussed.

Johansson (see Appendix B) proposed the use of structural equation modelling (SEM). SEM is a "quantitative second generation statistical analysis method that combines the benefits of path analysis, factor analysis, and multiple regression analysis". Johansson described some promising results from the aviation industry in the analysis of large sets of data. Johansson identified the benefits of the approach as:

- SEM can combine different data types; resulting models simultaneously analyse many statistical relations and present them in a comprehensive format.

- SEM provides a goodness of fit value of the model vs the variance in the dataset, which can be used when comparing alternative models.

Whereas these attributes of SEM suggest that it may have utility in the analysis of validation data, we note that it is likely that substantial conceptual development would be needed to incorporate SEM analysis into ISV acceptability analyses.

Another alternative approach is based on the work of Snow, Reising, Barry, and Harstock (see O'Hara, Appendix C). The focus is on demonstrating "practical equivalent", rather than "statistical significance", when comparing a new design to a baseline. This method is used in bioequivalence testing from drug research.

Yet another approach is to analyse data in an effort to arrive at a single "figure of merit". An example is the use of a mission capability level (MCL) scale (see O'Hara, Appendix C). This is used by the United States Department of Defense to assess how well operators using the system under testing can be expected to fulfill their intended mission in a realistic environment. The MCL assessment:

- provides a systematic methodology for arriving at conclusion regarding measure of effectiveness, measures of performance, and measures of suitability
- provides a framework for aggregation when multiple critical operational issues exist
- normalises evaluation results to a common scale, allowing comparisons across systems

See also Braarud, Eithreim, and Fernandes (2015) for another approach to new, higher-level measures.

These are examples of the types of data analysis methods that can be explored to support validation teams in analysing and understanding large sets of data. The goal of such methods is to help analysts better see relationships in the data and the big picture emerging when all the results are considered.

Validation conclusions

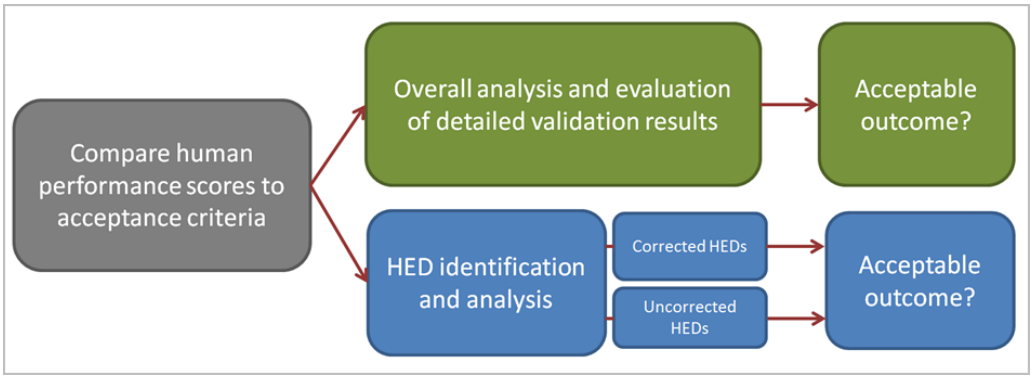
While advances in the tools available to analyse data are greatly needed, arriving at an overall conclusion regarding the acceptability of the design is not simply based on data analysis. As we noted in the discussion of what constitutes reasonable confidence in validation above, conclusions involve making judgements about the overall results and the process upon which those results were obtained. Confidence in the ability to generalise the findings beyond the simulator and into actual plant operations is essential for being reasonably confident in the results and the conclusions derived from them.

Norros (see Appendix C) and Laarni (see Appendix B) discussed the utilisation of the safety case methodology for ISV. They suggested that the analysis and the assessment of acceptability could be structured according to the concepts of requirements, arguments, and evidence. This point was also effectively made at the workshop by Skraaning (see Appendix C) and in the detailed discussion in Skraaning and Strand (2015). They argue that the arrival at a validation conclusion does not end with data analysis; there is an extra step involved which they call the "acceptability analysis process". The process is illustrated in Figure 4.2 below (reproduced from their paper). Skraaning and Strand summarise this process as follows:

> "In the second stage of the acceptability analysis, the validation team compiles all the detailed ISV results and judge the overall acceptability of the new control room. This analysis includes findings for each crew and individual operator; each scenario and scenario type; and for every performance measure included in the validation. The validation team has to weigh the importance of

the detailed performance observations and search for meaningful results that can substantiate whether or not the new control room is acceptable. To achieve this goal, it is necessary to reveal systematic effects, converging results, and other consistent patterns that are hidden in the huge and complex ISV data set. In addition, the relative importance of plant performance, task performance, team performance and cognitive performance for the operational safety, should be clarified. We also have to interpret conflicting evidence and contra-intuitive results. Furthermore, the validation team has to take into consideration that some performance indicators may have more predictive value than others, i.e. some measures may express the acceptability of human performance only in the sampled scenarios, while other measures anticipate operator performance in future scenarios. The validation team needs to evaluate minimum performance as well as typical performance, i.e. focus both on the acceptability of the lowest observed performance scores and average performance scores. In addition, we have to consider the principal and practical weaknesses of the validation methodologies, such as benchmarking, operationalisation of requirements or expert rating in the specific ISV. Generic methodological issues like simulator fidelity, scenario representativeness, participant training, statistical conclusion validity, etc. should also be evaluated to estimate the trustworthiness of the findings in the particular validation study (see NUREG-0711 rev 3, pp. 85-93)[1]. Finally, possible unanticipated adverse effects of the new control room design have to be taken into account."

Figure 4.2: Acceptability analysis process



Source: Skraaning and Strand, 2015.

The authors consider this process similar to building a safety case and cite a United Kingdom Office of Nuclear Regulation (ONR) document (ONR, 2013). That document describes what a safety case should involve:

> "A safety case should communicate a clear and comprehensive argument that a facility can be operated or that an activity can be undertaken safely. The safety case for a facility or activity should demonstrate that the associated risk and hazards have been assessed, appropriate limits and conditions have been defined, and adequate safety measures have been identified and put in place." (p. 6)

Of course, validation conclusions do not only pertain to safety. A parallel argument needs to be made for the production side of validation objectives.

This aspect of arriving at a conclusion beyond the data analysis is a key element to understanding what constitutes reasonable confidence. It is also an aspect of validation decision-making that is not addressed in available validation standards. The work of Laarni and Skraaning and Strand provides a basis to begin considering this question in more detail.

4.4 Broader validation challenges

In this section, we have chosen to focus on a subset of the key issues and areas of needed research identified from the workshop discussions (see Sections 3.1.3, 3.2.3, 3.3.3, and 3.4.3 for discussions of additional topics). There were also discussion of other big picture issues that are important to address, but may fall outside the use of research as a means to resolve them. That's not to say research cannot contribute to their resolution, only that their resolution may be more dependent on developing consensus than new technical insights or capabilities. These issues included:

- The consideration by validation guidance developers of recommended practices based on industry experience, for example as discussed in this workshop, for inclusion in industry guidance and standards documents.
- The need to consider validation's scope and objectives more broadly to define:
  - validation objectives when a similar design has been validated previously or validated in another context;
  - which control room activities need to be evaluated for specific types of validation efforts;
  - which activities outside the control room should be included in validation efforts, e.g. activities that are carried out in secondary control areas such as the remote shutdown station and in a technical support and outage control centers and maintenance activities;
  - validation objectives reflecting operation philosophy and concept of operation, including those reflecting the safety concept; and
  - validation objectives reflecting the safety concept within the context of the plant's safety case.
- The general need for more standardisation of approaches to validation, which is perhaps achieved through efforts to harmonise the standards and guidance used by designers and regulators.
- The challenges posed by different regulatory approaches to the review and evaluation of validation approaches.
- The need to address cultural and operational differences when conducting validation for designs that will be used in different parts of the world.
- The need for a "repository of lessons learned" (Gibson, 2015), i.e. "The creation of a global ISV lessons learned database, across companies and regulators, would provide information to the regulators, and the human factors community about the issues surrounding ISVs."
- The challenge of conducting a validation of the final design while supporting timely licensing of operators.

Addressing needs such as these will enhance the efforts to better define the degree of confidence that can be placed in validation.

4.5 Conclusions

The workshop set out to address a challenging topic – "establishing reasonable confidence in the human factors validation of main control room systems of nuclear power plants"; and in so doing, to

- Critically examine preliminary and final (integrated) validation activities to better understand their strengths, limitations, and potential inter-relationships with respect to the technical and

practical considerations for achieving reasonable confidence in nuclear power plant control room designs and modifications.

- Identify recommended practices, potential solutions, and available technical bases for addressing current limitations.
- Identify priority areas for future research.

The papers and discussions provided important insights into these topics. With respect to the issue of reasonable confidence, the workshop identified two general principles for establishing confidence in validation: use of industry guidance and attending to the specifics of the actual system being validated and its context. The types of recommended practices identified at the workshop can provide input to guidance development efforts and eventually become integrated into them. For example, adoption of a multi-stage approach to validation may support the establishment of reasonable confidence. The workshop emphasised that more commonly accepted approaches are needed. Industry guidance provides a focal point for gaining acceptance of the validation process by both designers and regulators.

It was generally concluded that, while industry validation guidelines are valuable in achieving confidence in validation, there is considerable work to be done. The available guidance documents have limitations in terms of the level of detail needed (e.g. to support the identification of specific scenarios to be used to support generalisation of results) and gaps, i.e. areas not addressed adequately at all (e.g. achieving validation conclusions). These needs speak directly to the issue of "reasonableness" of the validation process. Addressing these needs will provide a clearer picture of what is needed for results to be reasonable.

Providing promising pathways to address these issues was the workshop's most significant contribution. Research needs were identified and the technical bases to develop more complete guidance were discussed. We have identified some of the important issues and pathways to resolution in Section 4.3, and Section 3 has more complete coverage of these discussions. Another important conclusion from the workshop was that the issue of reasonableness is not strictly a validation methodology consideration. There are many broader validation challenges that have to be addressed as well.

Defining and achieving reasonable confidence is a complex challenge but one that can be substantially addressed by establishing guidance that is sufficiently detailed to support the development and execution of validation processes that is agreed upon by the community of stakeholders as being technically sound and meeting standards of reasonableness. The workshop resulted in a better understanding as to what it will take to define and establish reasonable confidence.

5. REFERENCES

Berntson, K., A. Kozak and J. Malcolm (2006), "Shift Manager Workload Assessment – A Case Study", In *Proceedings of the Fifth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2006), American Nuclear Society, La Grange, IL, United States.

Berntson, K., A. Kozak, J. Malcolm and D. McGregor (2004), "Validation of the Bruce A Unit 3 and 4 SCA – A Case Study", In *Proceedings of the Fourth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2004), American Nuclear Society, La Grange, IL, United States.

Braarud, P. Ø. (2015), "An Overall Framework for the Definition of Requirements, Criteria and Human Engineering Discrepancies for Control Room Validation", In *Proceedings of the Ninth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2015), American Nuclear Society, La Grange, IL, United States.

Braarud, P. Ø., M. Eithreim and A. Fernandes (2015), "SCORE" – "An Integrated Performance Measure for Control Room Validation", In *Proceedings of the Ninth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2015), American Nuclear Society, La Grange, IL, United States.

Braarud, P. Ø., E. Nystad, S. Strand, G. Skråning, A. Bye, M. Hildebrandt and S. Massaiu (2010), *The Halden Reactor Project Meeting on Integrated System Validation (ISV) – Status of Current Approaches to ISV and R&D Needs*, (HWR-939), OECD Halden Reactor Project, Halden, Norway.

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

De la Garza, C., H. Pesme, P. Le Bot and J.-P. Labarthe (2013), "Interest of combining two human factor approaches for the evaluation of a future teamwork", In *Safety, Reliability and Risk Analysis: Beyond the Horizon*, Steenbergen et al. (Eds.), ESREL Conference, 29 September-2 October 2013, Amsterdam, The Netherlands, pp. 523-529.

De la Garza, C., J.-P. Labarthe and L. Graglia (2015), "Feedback of an iterative evaluation process as part of a design project to support a human factors engineering program in the nuclear field", in *Proceedings 19th Triennial Congress if the IEA, Reaching out*, 9-14, August, Melbourne, Australia.

Elsner, T., T. Freitag and C. Rivere (2015), "Lessons Learned during the Human Factors Guided Validation of Human-System Interfaces in New Build NPPs and Modernization Projects". In *Proceedings of the Ninth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2015), American Nuclear Society, La Grange, IL, United States.

IAEA (2010), *Integration of Analog and Digital Instrumentation and Control Systems in Hybrid Control Rooms*, International Atomic Energy Agency, Vienna, Austria.

IEC (1995), *Nuclear Power Plants Main Control Rooms – Verification and Validation of Design*, IEC Standard 61771, International Electrotechnical Commission, Geneva, Switzerland.

Jordi, C. and T. Elsner (2015), "An Approach to Identify and Analyze the Most Significant Operator Tasks in a New Built Nuclear Power Plant", In *Proceedings of the Ninth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2015), American Nuclear Society, La Grange, IL, United States.

Labarthe, J-P. and C. De la Garza (2011), "The human factors evaluation program of a control room: the French EPR approach", In *Human Factors and Ergonomics in Manufacturing and Service Industries*, 21(4), Wiley Periodicals, Inc., pp. 331-349.

Molino, J. and J. O'Hara (2010), *HFE Methods and Tools Evaluation Report: Topic Priorities*, BNL Technical Report, No. 6859-3-2010, Brookhaven National Laboratory, Uptown New York, NY, United States.

NEA (2011), *Summary of Survey and Workshop Results on Areas of Research in Human Factors for the Design and Operation of New Nuclear Plant Technology: Final Report*, OECD, Paris.

NEA (2009), *The Role of Human and Organisational Factors in Nuclear Power Plant Modifications*, CSNI Technical Opinion Papers, No. 10, NEA/CSNI/R(2008)11, OECD, Paris.

O'Hara, J. and J. Higgins (2015), *Integrated System Validation: Models, Methods, and Issues*, BNL Technical Report No. 6859-1-2015, Brookhaven National Laboratory, Uptown New York, NY, United States.

O'Hara, J., J. Higgins, S. Fleger and P. Pieringer (2012), *Human Factors Engineering Program Review Model*, NUREG-0711, Rev. 3, US Nuclear Regulatory Commission, Washington, D.C., United States.

O'Hara, J. and J. Persensky (2011), "Human Performance and Plant Safety Performance: Establishing a Technical Basis and Framework for Evaluating New Human-System Interfaces", In *Simulator-Based Human Factors Studies*, A. Skjerve and A. Bye (Eds.), Chapter 6, Springer, New York, NY, United States.

O'Hara, J., J. Higgins, W. Brown, R. Fink, J. Persensky, P. Lewis, J. Kramer, A. Szabo, A. and M. Boggi (2008), *Human Factors Considerations with Respect to Emerging Technology in Nuclear Power Plants*, NUREG/CR-6947, US Nuclear Regulatory Commission, Washington, D.C., United States.

ONR (2013), *The Purpose, Scope, and Content of Safety Cases*, ONR NS-TAST-GD-051, Revision 3, Office of Nuclear Regulation, London, United Kingdom.

Patterson, E., Roth, E., and D. Woods (2010), "Facets of complexity in situated work", In *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*, Patterson E. and J. Miller (Eds.), Ashgate Publishing, New York, NY, United States, pp. 221-251.

Rivere, C. (2015), "Human Factors Engineering Verification and Validation Process of New Nuclear Power Plant Control Room: How to Bridge the Gap from Stepwise V&V to final ISV", In *Proceedings of the Ninth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2015), American Nuclear Society, La Grange, IL, United States.

Roth, E. and J. O'Hara (2002), *Integrating digital and conventional human system interface technology: Lessons learned from a control room modernization program*, (NUREG/CR-6749), US Nuclear Regulatory Commission, Washington, D.C., United States.

Savioja, P. (2014), *Evaluating systems usability in complex work – Development of a systemic usability concept to benefit control room design*, VTT Science 57, Espoo, VTT, Finland.

Savioja, P. and L. Norros (2013), "Systems Usability Framework for Evaluating Tools in Safety-Critical Work", In *Cognition, Technology, and Work*, 15(3), pp. 255-275.

Skraaning, G. and S. Strand (2015), "Integrated System Validation: The Acceptability Analysis Process", In *Proceedings of the Ninth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, (NPIC & HMIT 2015), American Nuclear Society, La Grange, IL, United States.

6. LIST OF PARTICIPANTS

Pedro Trueba Alonso	Human Factors Engineering Senior Specialist/Project Manager Tecnatom, S.A.
Per Oivind Braarud	OECD Halden Reactor Project/Institute for Energy Technology
Ronald Laurids Boring	Idaho National Laboratory
Cecilia De la Garza	Electricité de France
Thomas Elsner	AREVA
Valerie Gawron	MITRE Corporation
Michael P. Gibson	AREVA
Bob Hall	REH Technology Solutions
Greg Jamieson	University of Toronto
Shawn Jerrow	NuScale Power
Yvonne Johansson	Swedish Radiation Safety Authority
Jeffrey Jones	Nuclear Safety Engineering NuScale Power
Yun Goo Kim	Korea Hydro & Nuclear Power Co., Ltd.
Jari Laarni	VTT Technical Research Centre of Finland
Nathan Lau	Virginia Tech
Scott Malcom	Candu Energy Inc.
Kenji Mashio	Mitsubishi Heavy Industries, Ltd.
Leena Norros	VTT
Dina Notte	AREVA
John O'Hara	Brookhaven National Laboratory
Paul Pieringer	US Nuclear Regulatory Commission
Julie Reed	Westinghouse
Cyril Rivere	AREVA
Emilie Roth	Roth Cognitive Engineering
Alice Salway	Specialist Canadian Nuclear Safety Commission
Poon Hyun Seong	Korean Nuclear Society (KNS)
Gyrd Skraaning Jr	OECD Halden Reactor Project
Daniel Tasset	IRSN
Facilitators and Reporters	
Amy D'Agostino	US Nuclear Regulatory Commission
David Desaulniers	US Nuclear Regulatory Commission
Stephen Fleger	US Nuclear Regulatory Commission
Brian Green	US Nuclear Regulatory Commission
Niav Hughes	US Nuclear Regulatory Commission
Stephanie Morrow	US Nuclear Regulatory Commission
Jacqwan Walker	US Nuclear Regulatory Commission

APPENDIX A
WORKSHOP PARTICIPANTS

Pedro Trueba Alonso
Human Factors Engineering Senior Specialist / Project Manager Tecnatom, S.A,
Avda. Montes de Oca, 1, 28703 San Sebastián de los Reyes-Madrid, Spain
Phone: +34 916598600/+34 916598847
Email: ptrueba@tecnatom.es

Mr Pedro Trueba Alonso is a Mining Engineer specialised in Energy and Fuels and has been working for almost 20 years in Tecnatom, S.A. in many projects related in Human Factors Engineering and Modernization in nuclear power plants, including: Licensing activities for the AP600(Westinghouse), New plants: Lungmen (ABWR for TPC), Fuqing, Fanjiashan, Chanjiang (M310+ for CNNC), Existing plants: Beznau, Angra I, Atucha I, Kalinin, South Ukraine and in Spain like Almaraz I & II, Trillo, Cofrentes, Vandellós II.

He is specialised in Human Factors Verification and Validation and has participated in several Control Room validations for new designs (Lungmen) and for Design Modifications (Beznau, Vandellos II, Almaraz and Trillo). His present position is V&V Project Coordinator for the above mentioned CNNC Chinese plants.

Per Oivind Braarud
Senior Researcher
OECD Halden Reactor Project/Institute for Energy Technology
Phone: +47 40721464
Email: per.oivind.braarud@hrp.no

Per Oivind Braarud is a Senior Researcher at the OECD Halden Reactor Project / Institute for Energy Technology. He has a Msc in psychology from the Norwegian university of science and technology (NTNU). He has worked on several Human Factors Integrated System Validation (ISV) projects for Swedish modernised reactors. This work has included scenario specifications, specification of human performance measures, and establishing applied approaches for the analysis of data and development of conclusions about control room status. He has also worked on several Human Factors simulator Experiments for the Halden Reactor Project (HRP) on topics such as ISV, Human reliability Analysis, and Teamwork. He is currently project leader for the HRP research on ISV focusing on the development of valid human performance measures for ISV applications.

Ronald Boring
Principal Human Factors Scientist, Idaho National Laboratory
Phone: +1 (208) 526-0966
Email: ronald.boring@inl.gov

Ronald Laurids Boring, Ph.D., is a Principal Human Factors Scientist at Idaho National Laboratory, where he has worked as a researcher and principal investigator on projects for the US Nuclear Regulatory Commission, NASA, the US Department of Energy, the Canadian Nuclear Safety Commission, and the Norwegian Research Council. He previously worked as a human reliability researcher at Sandia National Laboratories, a usability engineer for Microsoft Corporation and Expedia Corporation, a guest researcher in human-computer interaction at the National Research Council of Canada, and a visiting human factors scientist at OECD Halden Reactor Project. He is currently principle investigator for the US Department of

Energy's Light Water Reactor Sustainability Program's project on control room modernisation. He is an adjunct faculty member at the University of Idaho.

Dr Boring has a B.A. in psychology and German from the University of Montana, an M.A. in human factors and experimental psychology from New Mexico State University, and a Ph.D. in cognitive science from Carleton University. He was a Fulbright Academic Scholar to the University of Heidelberg, Germany. He has published over 130 research articles in a wide variety of human reliability, human factors, and human-computer interaction forums. He has served on the organising committees for international conferences held by the Human Factors and Ergonomics Society, IEEE, the Association for Computing Machinery, and the American Nuclear Society.

Cecilia De la Garza
Senior Researcher on Ergonomics, EDF
Phone: + 33 (0)1 47 65 29 20
Email: cecilia.de-la-garza@edf.fr

Large experience on ergonomics studies: analysis and solutions oriented safe design, safety and accident prevention in different industrial fields: nuclear power plants, railways, printing, etc. A psychologist profile applied in the cognitive analysis and the understanding of the human behaviour in risky and dynamic situations, and the analysis of cooperation and coordination of the teams. These last 7 years I have been working in different design projects (new build and modification) in the phases of specifications and evaluation/validation particularly in control rooms. I developed with other colleagues a multidisciplinary approach having for aim to structure a framework for an Engineering Human Factors Program.

Core Competencies: Ergonomics and Human factors integration into design; Safety design: Contribution to the specifications and the evaluation in human factors engineering programmes; Cognitive analysis in dynamic situations; Teamwork skills analysis; Human factors evaluation of technical devices, control means; Risk analysis and prevention; Accident and incident analysis.

Education/Certifications: Thesis submitted in support of the accreditation to supervise research, René Descartes University, Paris, 2004. PhD on Ergonomics, *Ecole Pratique des Hautes Etudes*, Paris, 1995.

Thomas Elsner
AREVA
Phone: +49 (0) 9131 900-92878
Email: thomas.elsner@areva.com

Thomas Elsner's experience includes:

- Studies of Psychology at the University of Regensburg (1999-2004): Experimental Psychology; Cognitive Psychology; Multifactorial and Multidimensional Statistics.
- Freelancing (2004-2007): Usability Testing; Usability Inspections; Job-Analysis.
- Automotive Industries (2007-2010): HSI Specification; HSI Design Verification; Requirements Management.
- Nuclear Industries/AREVA (since 2010): HSI Reviews of MCR and local HIS; Preparing and Conduction of (Sub-) System Validation Studies in Full-Scope Simulator Environment: Scenario Selection; Definition of Cognitive Constructs and its measurement (observation, ratings, qualitative and quantitative questionnaires); Test Planning; HFE Observation; Data Analysis.

Valerie Gawron
Human Systems Integrator MITRE Corporation
Phone: 703-983-7001
Email: vgawron@mitre.org

Valerie Gawron has a BA in psychology from the State University College at Buffalo, a MA also in psychology from the State University College at Geneseo, a PhD in Engineering Psychology from the University of Illinois, and a MS in Industrial Engineering and MBA both from the State University of New York at Buffalo. She completed postdoctoral work in environmental effects on performance at the New Mexico State University in Las Cruces and began work for Calspan directly following. She remained at Calspan for 26 years until it was eventually acquired by General Dynamics and she was made a technology fellow. She is presently a human systems integrator at the MITRE Corporation. Dr Gawron has provided technical leadership in Research, Development, Test, and Evaluation of small prototype systems through large mass produced systems, managed million dollar system development programmes, led the design of information systems to support war fighters and intelligence personnel, fielded computer aided engineering tools to government agencies and industry, tested state-of-the-art displays including Helmet Mounted Displays, Night Vision Goggles, and Synthetic Vision Displays in military and commercial aircraft, evaluated security systems for airports and United States Embassies, conducted research in both system and human performance optimisation, applied the full range of evaluation tools from digital models through human-in-the-loop simulation to field operational tests for military, intelligence, and commercial systems, directed accident re-enactments, consulted on driver distraction, accident investigation, and drug effects on operator performance, and written over 380 publications including the Human Performance, Workload, and Situation Awareness Measures Handbook (second edition) and 2001 Hearts: The Jane Gawron Story. Both of these are being used internationally in graduate classes, the former in human factors and the latter in patient safety.

Dr Gawron has served on Air Force Scientific Advisory Board, the Army Science Board, Naval Research Advisory Committee, and National Research Council. She gives workshops on a wide range of topics to very diverse audiences from parachute testing given as part of the Sally Ride Science Festival for girls ages 8 to 14 to training applications of simulation to managers and engineers. She has worked programmes for the United States Air Force, Army, Navy, Marines, NASA, the Departments of State and Justice, the Federal Aviation Administration, the Transportation Security Administration, the National Transportation Safety Board, the National Traffic Safety Administration, as well as for commercial customers. Some of this work has been international and Dr Gawron has been to 172 countries. Dr Gawron is an associate fellow of the American Institute of Aeronautics and Astronautics, a fellow of the Human Factors and Ergonomics Society, and a fellow of the International Ergonomics Association.

Mike Gibson
Principal Engineer
Human Factors AREVA INC. Charlotte, NC, United States
Phone: 704-805-2245
Email: Mike.P.Gibson@areva.com

Mr Michael P. Gibson has a Master of Science in Industrial and Systems Engineering and Human Factors from Virginia Tech University. He has served as a Principal Engineer at AREVA since March 2009. His experience during this time includes being the Lead Human Factors Engineer on the AECL Solid Liquid Waste Cementation Conceptual Design project, the ITER Tokamak Cooling Water System project, and analysing and designing human machine interfaces and maintenance activities on multiple nuclear power plant projects. Michael is also leading the development of AREVA's global human factors Integrated System Validation methodology. Prior to AREVA, Michael worked as a human factors engineer at General Motors (GM) in the design, development, and validation of user interfaces, occupant accommodation, and in the GM manufacturing ergonomics laboratory. Michael also worked as a human factors engineer at the

Pantex Plant in Amarillo, Texas evaluating nuclear weapon dismantlement programme procedures and processes.

Bob Hall
CEO of REH Technology Solutions REH Technology Solutions, LLC
Phone: (631)928-5456
Email: rehtec@optonline.net

Mr Hall is the founder and current CEO of REH Technology Solutions, LLC (RTS), a small high technology consulting and development company specialising in human factors engineering located in New York. He is a member of the IEEE's Nuclear Power Engineering Committee and EPRI's Technical Advisory Group. Mr Hall has an international client base in advanced nuclear power plant human system interface (HSI) design and US NRC chapter 18 design certification application (DCA), including:

- USAPWR-Mitsubishi Heavy Industries/Mitsubishi Electric/Mitsubishi Electric Power Products.
- APR1400-Korea Electric Power-E&C/Korea Hydro & Nuclear Power.
- Human Factors Guidance For Control Room & Digital Human System Interface Design and Modification (EPRI 1010042) revision project- EPRI through CDF Services.
- mPower Reactor Program-B&W through WTS.

Mr Hall's current activities include; validation testing, human factors programme development, digital control room HSI design, review of and principal author of NUREG-0711, rev 2 and 3, development of DCA programmes and documentation, and client representation to the US NRC.

With over 40 years of engineering and management experience, 30 of which was with Brookhaven National Laboratory (BNL), a Department of Energy National Laboratory, Mr Hall has authored over 190 technical papers and coauthored a college text book on risk assessment techniques in the nuclear power industry. While at BNL, Mr Hall headed the Engineering Technology Division with a staff of 60 scientists and engineers and an annual research and development budget in human factors engineering, risk assessment regulatory applications, of over USD 20 million.

Prior to joining the scientific staff of BNL Mr Hall worked for a NY power utility on the construction of its first boiling water reactor.

Mr Hall holds advanced degrees in mechanical and nuclear engineering.

Greg Jamieson
Associate Professor
Department of Mechanical and Industrial Engineering, University of Toronto
Phone: +1 416 946-8504
Email: jamieson@mie.utoronto.ca

Greg A. Jamieson is Associate Professor and Clarice Chalmers Chair in Engineering Design in the Department of Mechanical and Industrial Engineering at the University of Toronto. Professor Jamieson's research and consulting work address the design of artifacts for decision support. He specialises in human-automation interaction and human-machine interface design with applications in the process and energy industries.

Shawn Jerrow
Supervisor Operations Engineering NuScale Power, LLC
Phone: 541-360-0727
Email: sjerrow@nuscalepower.com

Mr Jerrow is currently the Supervisor of Operations Engineering for NuScale Power. Mr Jerrow spent 8 years in the nuclear navy where he qualified and operated two different types of pressurised water reactors. He then spent 20 years at the Columbia Generating Station commercial nuclear power plant (BWR) mainly in Operations where he held a Senior Reactor Operator (SRO) license, worked as an operator in all the control room positions, and held several management positions such as Senior License, Operations Manager, and Operations Training Manager. He also spent 4 years working various operationally related jobs on the DOE Hanford Site including the Waste Treatment Plant supporting task analysis, HSI design, procedures, and training.

Mr Jerrow has extensive experience in the operations, maintenance, plant design modification, training, and procedure development of nuclear facilities. Additional relevant experience includes integration of digital technology and control systems into an analog control room, control room re-design, scenario development and validation, and human performance improvement of the operating crews. In his current role, Mr Jerrow is leading an HFE design team focused on developing and implementing the NuScale HFE programme which includes OER, analysis, HSI development, simulator development, control room design and layout, procedures, and training.

Yvonne Johansson
Human Factors Specialist, Swedish Radiation Safety Authority
Phone: + 46 8 799 43 62
Email: Yvonne.johansson@ssm.se

Dr Johansson has been Human Factors Specialist at the Unit for Human-Technology-Organization (MTO), at Swedish Radiation Safety Authority (SSM) since 2001. Parallel with this she obtained her PhD in Psychosocial Medicine at Karolinska Institute, Stockholm, Public Health, in 2010. Dr Johansson's primary work is supervision of nuclear power plant safety through inspections and reviews, especially plant modifications and modernisation that may have impacts on humans. This work includes a variety of Man-Technology-Organization-related areas. She has additional interest in the methods of evaluation and takes care of SSM's research interests in this field.

Jeffrey Jones
Supervisor, Nuclear Safety Engineering NuScale Power, LLC
Phone: 704-526-3412
Email: jjones@nuscalepower.com

Jeffrey Jones has been active in the nuclear industry for 27 years. His background includes 6 years as a US Navy submarine reactor operator and bachelors and masters degrees in nuclear engineering. Jeffrey has been a reactor engineer, a licensing engineer, a mechanical engineering manager, and a project engineering manager for an EPC firm. Working on the design certification project for AREVA's US EPR, Jeffrey was asked to initiate and manage the human factors engineering programme and write Chapter 18 of the application. That experience grew into a global role coordinating HFE expertise across AREVA's international offices, writing design standards for company use, and building the international HFE business line. Jeffrey has also worked as HFE programme consultant for the US APWR and the NuScale plant.

Yun Goo Kim
Senior Researcher, Korea Hydro & Nuclear Power Co., Ltd. Central Research Institute
Phone: +82-(0)42-870-5737
Email: ygkim.stpn@khnp.co.kr

Dr Yun Goo Kim is senior researcher of Korea Hydro & Nuclear Power Co., Ltd. He received a Ph.D degree in nuclear and quantum engineering from the Korea Advanced Institute of Science and Technology (KAIST). He has been active in the area of digital I&C, advanced control room, and human factors engineering for over 15 years including computerised procedure system design for many years. He also has been involved APR1400 digital control room verification and validation for Shin-Kori 3, 4 and Shin-Hanul 1,2. He currently leads I&C and HFE field licensing for APR1400 and APR+.

Jari Laarni
Principal Scientist, VTT Technical Research Centre of Finland
Phone: +358 40 595 2313
Email: jari.laarni@vtt.fi

Jari Laarni is Principal Scientist at the Systems Research Centre of VTT Research Centre of Finland. He has a Ph.D. and Master's degree in psychology from the University of Helsinki, Finland. He is specialised in the areas of cognitive psychology, cognitive science, human factors, ergonomics, user-centred design, systems usability, human well-being and stress, and he has participated in several national and international research projects on these topics. He has also been involved in several projects in Finland concerning V&V of NPP control room systems.

Nathan Lau
Assistant Professor, The Grado Department of Industrial and Systems Engineering, Virginia Tech
Phone: +1 540 231-5936
Email: nathan.lau@vt.edu

Nathan Lau is Assistant Professor in the Grado Department of Industrial and Systems Engineering at Virginia Polytechnic Institute and State University (Virginia Tech). He received his Bachelor and PhD degrees from the University of Toronto on 2004 and 2012, respectively. Prior to joining Virginia Tech, he has conducted full-scope simulator experiments at the OECD Halden Reactor Project, Halden, Norway, and the Center for Engineering and Research, VA, USA. Professor Lau specialises in human-machine interface design and human performance assessment with applications in the process and energy industries.

Scott Malcom
Vice President, Candu Services Candu Energy Inc.
Phone: 905-823-9040 x33157
Email: scott.malcolm@candu.com

Scott Malcolm is currently the Vice President, Candu Services, responsible for business development and sales for the CANDU Reactor Services business line. He has 30 years of experience in design and commercial operations of high technology companies, including 25 years at Atomic Energy of Canada Ltd. (AECL)/Candu Energy Inc.

Scott spent 18 years in nuclear design at AECL and 5 years outside the nuclear industry in naval surface ship design. More recently he spent the last 5 years in senior commercial roles at AECL/Candu Energy in the after sales services organisation where as General Manager, he was responsible for managing USD 130 million/yr business, including product development, proposals and customer relations.

At AECL/Candu Energy, Scott held positions of Manager, Instrumentation, Control and Electrical Engineering, Section Head, Control Centre and Human Factors Design, and Principal Design specialist in Human Factors. He has been responsible for a wide range of engineering services, including control centre modifications to existing plants and support to advanced control centre designs such as the Advanced CANDU Reactor. Scott was a key developer of the process for incorporating human factors engineering into the Canadian nuclear industry, including a strong focus on the development of control centre functionality and supporting technology.

Scott has over 15 years' experience developing nuclear design standards and is a former Chair of the Institute of Electrical and Electronic Engineers (IEEE) Nuclear Power Engineering Committee and the Chief Canadian Delegate to the International Electrotechnical Commission (IEC) in the area of nuclear instrumentation. He has a Masters degree in Human Factors/Ergonomics from the University of Guelph in Ontario, Canada (1985) and is a Certified Human Factors Specialist (CHFP) from the Board of Certification in Professional Ergonomics (US).

Kenji Mashio
Engineering Manager, Mitsubishi Heavy Industries, Ltd. (MHI)
Phone: +81-78-672-3327
Email: kenji_mashio@mhi.co.jp

Kenji Mashio has 18 years' experience for NPP HFE and I&C technologies. He has been involved in various NPP HFE & I&C licensing, engineering, and construction projects, including Genkai-1,2 MCR control board replacement (1996-2003, the first control board replacement in Japan) with engineering in charge, Tomari-Unit 3 (a new construction with digital I&C & HSI platform in 2003-2005 as HFE/HSI licensing & engineering in charge), Ikata-1,2 control board replacement (the first screen based control board replacement project in 2005 as HFE/HSI licensing in charge, AP1000 DCD Chapter 18 support in WEC (2006), US-APWR DCD Chapter 7, 13, and 18 licensing and engineering in charge (2008-2013), and ATMEA 1 standard design (2014-present). During various projects, he has also been involved in the HFE V&V.

Leena Norros
Research professor (emerita) VTT
Phone: + 358 40 5616774
Email: Leena.Norros@elisanet.fi

Research professor emerita Leena Norros is an industrial psychologist working on the problems of cognitive ergonomics and design of complex industrial systems. She has spent most of her career by leading a human factors research group at VTT Technical Research Centre of Finland. She studied at University of Helsinki and the Technical University of Dresden, in which institutions she took her doctors degrees. She acts as docent (adjunct professor) at Helsinki University and teaches human factors engineering both at Helsinki University and Helsinki Technical University. In her work she has focused on the analysis of cognitive systems in context. Research has been conducted in several safety-critical domains like nuclear power plants, aviation (ATC), anaesthesia, maritime. In these studies and development projects she has worked together with experts of automation and reliability engineering, mathematics and information technology, and various domain experts. She has participated as human factors expert in many accident investigations in collaboration with the Finnish Accident Investigation Board. She has been actively associated with the European Association of Cognitive Ergonomics, the North-American naturalistic decision-making community and the International Ergonomics Association, OECD Nuclear Energy Agency. She has about 170 scientific publications in international journals, books and conference proceedings.

Dina Notte
Human Factors Expert, AREVA
Phone: 0033 134961787
Email: dina.notte.ext@areva.com

Dina Notte has an industrial psychologist and ergonomist education. She has founded ERGODIN consulting company and has worked for 35 years in the design of computerised control rooms and HSI in high-risk industrial processes like nuclear power plant, petrochemical sites, railways traffic controls, steel industry and robotics in surgery and telemedicine. Dina is used to be in charge of HF integration in long term and complex projects encompassing technical challenges (i.e. artificial intelligence, robotics, high automation and computerisation) and human reliability issues. During these projects, she had to build-up sophisticated HSI validation and verification methods and experimental protocols (i.e. Wizard of Oz techniques, Satellite Communication Simulation) based upon human performance metrics (i.e. cognitive workload assessment, cognitive walkthrough, team work evaluation). Dina is member of HFES Society since 1985 and member of the SELF since 1984. She is an HF expert for European Commission DG III; DG XII and DG XIII. She has been general secretary of the French Ergonomic Society during six years, and is Certified European Ergonomist since 1995. Today, she has a position of HF Expert in AREVA NP since 2009, in charge of developing an integrated HF team in close relationship with engineering disciplines and of building HFE training programme. She supervises HEF programme/process for new builds including next generation of NPP and modernisation projects with a focus on methodological issues notably concerning V&V and ISV steps. She issues recently a detailed and generic ISV Work Instruction as a practical guide to prepare, develop and implement ISV.

John O'Hara
Senior Scientist, Brookhaven National Laboratory
Phone: 631-344-3638
Email: ohara@bnl.gov

Dr John O'Hara is a Senior Scientist at Brookhaven National Laboratory and its Human Factors Research Manager. His research programmes address the effects of advanced technology on individual and crew performance in complex systems. Specific programmes address: (1) human factors engineering methods and tools; (2) the development of human factors design guidance for advanced systems including alarms, information systems, computer-based procedures, and controls; (3) the role of cognitive factors, such as attention, situation assessment, and workload in complex system operation and human error; and (4) evaluation methods of individual and integrated human-system performance. John's research has focused on many types of industrial systems, including: nuclear power, space, aviation, robotics, maritime, and homeland security. He has also performed numerous safety evaluations and design reviews of various types of complex systems, including nuclear power plants and NASA control centers. John is a Certified Human Factors Professional; Fellow of the Human Factors and Ergonomic Society; and Past Chair of American Nuclear Society's Division of Human Factors, Instrumentation, and Control.

Paul Pieringer
Sr. Reactor Operations Engineer,
US Nuclear Regulatory Commission
Phone: 301-415-4664
Email: paul.pieringer@nrc.gov

Mr Pieringer's work experience includes 27 years at Calvert Cliffs Nuclear Power Plant split between Operations, Quality Assurance, and Engineering positions. In 2008, Paul started work at the Nuclear Regulatory Commission where he currently fills the position of Sr. Reactor Operations Engineer in the Operating Licensing and Human Performance Branch of New Reactor Operations. His primary

responsibility in this capacity has been the review of Design Certification and Combined License applications. Paul has been the primary reviewer for the Verification and Validation element of these applications and has been actively involved in incorporating lessons learned from these reviews back into the review guidance (Standard Review Plan, RG 1.206, NUREG-0711, and NUREG-0700).

Julie Reed
Human Factors Technical Lead, Westinghouse
Phone: 412-374-5069
Email: reedji@westinghouse.com

Julie has a BSc in Psychology (City of London) and a MSc in Human Factors (University of Birmingham, UK). She has worked in the nuclear industry for twenty-seven years, providing effective and practical integration of safety and human factors into the design, licensing, construction, commissioning, operational and decommissioning phases of nuclear facilities. She was employed by British Nuclear Fuels (BNFL) in the UK, and transferred to BNFL Inc. in the USA in 1998. Since 2007, she has been employed by Westinghouse as the Human Factors Technical Lead. She has experience in dealing with human factors licensing issues for a range of facilities across a number of different countries. Julie's current main projects are associated with the build and licensing of the AP1000 'First of a Kind' PWRs in China and the USA.

Cyril Rivere
Human Factors Specialist AREVA
Phone: +33 1 34 96 25 19
Email: cyril.rivere@areva.com

Cyril Rivere is a Human Factors specialist, with a Master and PhD in Ergonomics completing a Biology and Physiology educational background. Prior to working for AREVA, he worked in a large company specialised in high performance materials and packaging manufacturing, where he developed methods for ergonomics of product applied to packaging design, favoring ergonomics and end-users integration in the design process. Since 2009, Cyril is specialised in Industrial Complex systems addressing Human and Organisational Factors issues as Human Factors Engineering Specialist for AREVA NP Company, where he is in charge of integrating the Human Factors discipline within the Engineering and Projects organisation with a focus on large Nuclear Power Plant New Builds projects. In this position, Cyril pilots (with an engineering and project management standpoint) and implements Human Factors principles, methods and requirements for several NPP project, with a large part on Human Factors preliminary analyses (FRA/FA, TA and OER) and Control Center, Control Room and HSI design. In addition, he developed for his projects the whole Verification and Validation approach, and is currently in charge of leading the Integrated System Validation process for one of them. As AREVA Human Factors specialist, Cyril is also in charge of training activities for engineers but also customers and safety authorities, and support the development of R&D projects.

Emilie Roth
Owner and Principal Scientist, Roth Cognitive Engineering, Roth Cognitive Engineering
Phone: 650-800-7188
Email: emroth@mindspring.com

Emilie Roth is the owner and principal scientist of Roth Cognitive Engineering. A cognitive psychologist by training, Dr Roth has been involved in analysis of cognitive challenges and design of decision support in a variety of domains ranging from nuclear power plant control rooms to railroad operations, military command and control, medical operating rooms and intelligence analysis. Dr Roth supported design and evaluation of the Westinghouse AP600 and the Mitsubishi APWR next-generation nuclear power plants. She is a fellow of the Human Factors and Ergonomics Society, is an Associate Editor of the Journal of

Cognitive Engineering and Decision Making, and serves on the editorial board of the journal Human Factors. She recently participated in the National Research Council Committee on Human-System Design Support for Changing Technology and is currently a member of the National Research Council Committee examining lessons learned from the Fukushima nuclear accident for improving safety and security of US nuclear plants.

Alice Salway
Human and Organisational Factors Specialist Canadian Nuclear Safety Commission
Phone: 613-991-3271
Email: Alice.Salway@cnsc-ccsn.gc.ca

Alice Salway has worked in Canada, the USA, the European Union, Australia, and Norway. She has a broad, practical, knowledge of ergonomics/human factors, psychology (organisational and cognitive), accident investigation and safety analysis. Alice is familiar with a range of user-centred methods, structured methods, and analysis techniques used for systems design projects. She has worked in nuclear, aviation, transportation, mining, oil and gas, chemical, and defence domains, although her non-nuclear experience is from 9 years ago or more.

While working in design teams, she managed, planned and conducted test, evaluation, and validation activities. Alice was also involved in developing military human factors integration / human systems integration approaches and standards in the UK and Canada, which position validation activities in the systems design process as well as specifying detailed guidance and requirements for validation activities.

In her current role, she contributes to Canadian standards and regulatory documents, develops approaches and criteria for regulatory inspections and technical assessments, and carries out inspections and assessments of Canadian nuclear licensee's facilities and activities.

Her qualifications and affiliations are:

- B.Sc. (Hons.) Ergonomics, Loughborough University of Technology, England, 1987
- Ph.D. Cognitive Psychology, University of Aberdeen, Scotland, 1991
- Registered European Ergonomist (CREE – Eur. Erg.), since 1998
- The Human Factors and Ergonomics Society (USA), Full Member since 1994

Poong Hyun Seong
Professor, Dept. of Nuclear and Quantum Engineering/Vice President/President – Elect Korean Nuclear Society (KNS); KAIST/Korean Nuclear Society (KNS)
Phone: +82-42-350-3820
Email: phseong@kaist.ac.kr

Dr Poong Hyun SEONG is currently a Professor in Nuclear Engineering at KAIST, Korea. He is also the vice president/president-elect of Korean Nuclear Society now. Prof. Poong Hyun SEONG had his BS degree from Seoul National University in 1977, and MS and PhD degrees in nuclear engineering from Massachusetts Institute of Technology (MIT) in 1984 and 1987, respectively. He worked as the chief editor of "Nuclear Engineering and Technology" from 2003 to 2008. He was a commissioner of the Korea Nuclear Safety Commission from 2006 to 2009. He was the chair of the HFICD (Human Factors and Instrumentation and Control Division) of the ANS (American Nuclear Society) from 2006 to 2007. He is now an editorial board member of "Reliability Engineering and System Safety". His research interest includes Digital Instrumentation and Control systems developments for Nuclear Power Plants, Software V/V, Human Reliability Analysis, and Cognitive Systems Engineering. He published numerous technical

papers (http://niclab.kaist.ac.kr) and he published a book "Reliability and Risk Issues in Large Scale Safety-critical Digital Control Systems", Springer in 2009.

Gyrd Skraaning Jr
Principal Scientist, OECD Halden Reactor Project
Phone: +47 905 43 402
Email: gyrds@hrp.no

Gyrd Skraaning Jr. received his Ph.D. in psychology from the Norwegian University of Science and Technology in 2003. He has worked as a Research Scientist at the OECD Halden Reactor Project since 1996, focusing on the evaluation of control environments in the nuclear domain, air traffic control and petroleum industry. His research interests include human performance measurement, experimental design and applied statistical analysis, and more recently; control room design for future nuclear plants.

Daniel Tasset
Deputy Head of the Department of Organisational and Human Factors in Risk Management IRSN
Phone: +33 (0)1 58 35 90 09
Email: daniel.tasset@irsn.fr

Daniel Tasset has over 30 years' experience in Ergonomics and Human Factors engineering in different areas such as military systems, aeronautics, and since 1995 in nuclear safety. He holds a Master of Ergonomics from Paris 5 University and a Master of Management from Paris 1 university.

Daniel worked for IRSN from 1995 to 2004 doing safety assessment of new reactors in France such as N4 plants and EPR. In 2004, he joined the French Safety Authority, ASN, and worked on developing human factors in the regulation and control activities of ASN. Daniel returned to IRSN in 2011 and is now deputy head of the department of organisational and human factors in risk management.

Daniel is a member of the CSNI Working Group on Human and Organisational Factors in the Nuclear Energy Agency of the OECD. He has been chairman of this group from 2009 to 2014. He is a member of the Halden project committee programme. Daniel has been involved in actions with IAEA on safety culture and human factors.

WORKSHOP FACILITATORS and REPORTERS

Amy D'Agostino
Human Factors Analyst, US Nuclear Regulatory Commission
Phone: 1 301-251-7915
Email: amy.dagostino@nrc.gov

Dr Amy D'Agostino is a Human Factors Analyst in the Office of Nuclear Regulatory Research at the US Nuclear Regulatory Commission. Dr D'Agostino has been with the NRC since 2009 and serves as the technical lead for various human performance projects. She established the NRC's human performance test facility and develops and oversees human performance studies using a nuclear power plant main control room simulator. In addition, Dr D'Agostino is the technical lead for a project that examines measurement of situation awareness, workload and teamwork, identifies the various measurement tools available and characterises the appropriate usage of these tools. Dr D'Agostino has also been involved with projects addressing automation and complexity in nuclear power plants, the concept of operations of small modular reactors and human factors concerns in a fire scenario. Dr D'Agostino has a Ph.D. and Master's degree in industrial and organisational psychology from the University of Connecticut.

David Desaulniers
Senior Technical Advisor for Human Factors and Human Performance Evaluation, US Nuclear Regulatory Commission
Phone: 1 301-415-5918
Email: david.desaulniers@nrc.gov

Dr Desaulniers is a senior level scientist at the United States Nuclear Regulatory Commission (NRC). He currently serves as NRC's Senior Technical Advisor for Human Factors and Human Performance Evaluation, providing expert technical advice on emerging technical and policy issues concerning human performance in nuclear safety. He obtained his doctorate in Psychology from Rice University in Houston, Texas where he specialised in Engineering Psychology. During the past 25+ years his work has addressed a wide range of technical and policy issues where there is a nexus between the design and operation of nuclear power plants, human and organisational performance, and the protection of public health and safety. Specific activities include serving as NRC's technical lead for the development of a federal regulation to require fatigue management programmes at all US commercial nuclear power plants and providing human factors technical expertise in agency initiatives concerning severe accident management, crediting manual actions, evaluating the cumulative impact of operator workarounds, and assessing control room conduct of operations. Dr Desaulniers' current work is focused on supporting NRC's response to the lessons learned from the nuclear accident at Fukushima Dai-ichi and the integrated system validation of main control room designs. Dr Desaulniers is also active in industry and international organisations, serving as Chair of IEEE's Nuclear Power and Energy Committee, Subcommittee 5 (Human Factors, Control Facilities, and Reliability) as well as serving as a Vice-Chair of NEA's Working Group on Human and Organisational Factors.

Stephen Fleger
Senior Human Factors Analyst, US Nuclear Regulatory Commission
Phone: 1 301-251-7905
Email: Stephen.fleger@nrc.gov

Steve Fleger is a Senior Human Factors Analyst in the Office of Nuclear Regulatory Research at the US Nuclear Regulatory Commission. He is a Certified Human Factors Professional (CHFP) from the

Board of Certification in a Professional Ergonomics with 36 years of human factors engineering design, analysis, and evaluation experience. Before joining the NRC in 2008 Mr Fleger worked in the private sector where he had technical and managerial responsibilities on over 70 consulting projects for clients across the commercial and private sectors, including business, industry, the US government and military, and foreign clients in Japan, Bulgaria, Spain, Lithuania and Russia. In 2006, Mr Fleger was nominated and elected as the United States expert to the International Organization for Standardization (ISO) WG8, Control Centers, by the US TAG to TC159/SC4, Ergonomics for Human-System Interaction. Mr Fleger served as the Chair of the Institute of Electrical and Electronics Engineers (IEEE) Nuclear Power Engineering Committee (NPEC) Subcommittee 5 (SC5) for six years. SC5 is responsible for developing human factors engineering consensus standards for nuclear facility control centers. In this capacity he led the effort to develop and publish IEEE Standard 1786TM-2011, IEEE Guide for Human Factors Applications of Computerized Operating Procedure Systems (COPS) at Nuclear Power Generating Stations and Other Nuclear Facilities. Mr Fleger is currently serving as the Chair of NPEC, as well as the Chair of SC5's Working Group (WG) 5.4 on human reliability.

Brian Green
US Nuclear Regulatory Commission
Phone: 1 301-415-6728
Email: brian.green@nrc.gov

Dr Brian Green is a Human Factors Engineer in the Office of Nuclear Reactor Regulation at the US Nuclear Regulatory Commission. He is responsible for the assessment of proposed licensing actions related to control room modifications, ex-control room manual actions, and other human factors issues at nuclear power plants. Previously, he worked in the Office of New Reactors where he wrote several human factors verification and validation inspection procedures for use in the construction of new plants. Dr Green has a Ph.D. and Master's degree in industrial and systems engineering/human factors from the University of Buffalo. Prior to joining the NRC, he worked as a human factors research assistant at the Research Institute for Safety and Security in Transportation. Dr Green also works as an adjunct professor of psychology at the George Washington University.

Niav Hughes
Human Factors Analyst, US Nuclear Regulatory Commission
Phone: 1 301-251-7964
Email: niav.hughes@nrc.gov

Niav Hughes is a human factors analyst in the Office of Nuclear Regulatory Research at the NRC. Her work at the NRC primarily supports human factors research in the area of human performance and control room simulation studies.

Stephanie Morrow
Human Factors Analyst, US Nuclear Regulatory Commission
Phone: 1 301-251-7607
Email: stephanie.morrow@nrc.gov

Dr Stephanie Morrow is a Human Factors Analyst in the Office of Nuclear Regulatory Research at the US Nuclear Regulatory Commission. Dr Morrow has been with the NRC since February 2011 and serves as a project manager for various safety culture and human performance research activities. She has conducted studies of safety culture and its relationship to organisational performance and served as a safety culture assessor during onsite inspections of nuclear power plants. She is currently a member of the NRC project team to test a new human reliability analysis (HRA) method called the Integrated Decision-Tree Human Event Analysis System (IDHEAS). Dr Morrow has a PhD and Master's degree in industrial

and organisational psychology from the University of Connecticut. Prior to joining the NRC, she worked as an engineering psychologist at the Department of Transportation's Volpe National Transportation Systems Research Center.

Jacqwan Walker
US Nuclear Regulatory Commission
Phone: 1 301-415-4035
Email: Jacqwan.walker@nrc.gov

Jacqwan Walker is a Human Factors Analyst in the Office of New Reactors at the US Nuclear Regulatory Commission. He serves as a technical reviewer for aspects related to human performance activities in the licensing of new reactor control room designs. Some of his technical activities range from evaluations of design certification applications, combined license applications, license amendment requests, or (more recently) inspections of the integrated system validation for main control room designs. Mr Walker has been with the NRC since February 2003.

APPENDIX B WHITE PAPERS

and

APPENDIX C. SPEAKER PRESENTATIONS: ABSTRACTS/TEXTS AND GRAPHICS

are available in Pdf format only on the NEA website