Human Aspects of Nuclear Safety
NEA/CSNI/R(2016)17/ADD2
March 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

Appendix C

OECD
BETTER POLICIES FOR BETTER LIVES

NEA
NUCLEAR ENERGY AGENCY

Appendix C:

Speaker Presentations: Abstracts/Texts and Graphics

Appendic C - Table of Contents

Opening Remarks

Experts' Workshop on Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications

February 19 – 21, 2015
Charlotte, North Carolina

Participants

- Belgium
- Canada
- Finland
- France
- Japan
- Korea
- Norway
- Spain
- Sweden
- United States

- Regulatory Bodies
- Vendors
- Technical Support Organizations
- National Laboratories
- Universities
- Industry Consultants

Our Approach (1 of 3)

- Plenary Sessions
- Challenge Sessions
  - Challenge Presentations (45)
  - Issue Characterization (Q&A) (15)
  - Team Breakouts
    - Concept Generation and Organization (30)
    - Break (30)
    - Breakout discussion (90)
  - Summation of Recommendations (30)

7

Our Approach (2 of 3)

- Challenge Questions
  - Ordered chronologically
  - Focused on reasonable confidence
  - Designed to elicit recommendations

8

Our Approach (3 of 3)

- Panels
  - Both panels (Recommended Practices and Recommended Research) comprised of 4 panelists
  - Each panelists has been assigned to one of the 4 Challenge questions
  - Each panelist will recap the 4 team's recommendations for their assigned challenge question and provide preliminary commentary on their merits
  - Time will be allotted for participants questions and comments

9

Potential Challenges to Success

- Terminology
- Perspectives
- Focus
- "The" answer
- The question

10

Keys to Success

- Limit philosophical debate
- Make your assumptions explicit
- Refine the question if necessary
- Consensus need not be unanimous

11

Workshop Staff

- Facilitators
  - Amy D'Agostino
  - Steven Fleger
  - Brian Green
  - Dave Desaulniers
- Reporters
  - Niav Hughes
  - Stephanie Morrow
  - Jacqwan Walker
  - Andrew White

12

Task Working Group

- Per Oivind Braarud – IFE, Halden Reactor Project
- Amy D'Agostino – U.S. Nuclear Regulatory Commission
- Ceclia De-La-Garza – Électricité de France (EdF)
- David Desaulniers – U.S. Nuclear Regulatory Commission
- Stephen Fleger – U.S. Nuclear Regulatory Commission
- Bob Hall – REH Technology Solutions
- Jari Laarni – VTT Technical Research Centre of Finland
- Pierre Le Bot – Électricité de France (EdF)
- Leena Norros – VTT Technical Research Centre of Finland
- Dina Notte – AREVA / ERGODIN
- Cyril Rivere – AREVA
- Daniel Tasset – L'Institut de Radioprotection et de Sûreté Nucléaire (IRSN)
- Valeri Vassent – L'Institut de Radioprotection et de Sûreté Nucléaire (IRSN)

13

on to the program . . .

14

OECD Nuclear Energy Agency NEA

The NEA and Working Group on Human and Organizational Factors

Andrew White
NEA Secretariat for WGHOF

Experts' Workshop on Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications
February 2015

OECD Nuclear Energy Agency NEA

The NEA Mission

- 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.

OECD Nuclear Energy Agency NEA

OECD/NEA Membership

- Australia
- Austria
- Belgium
- Canada
- Chile
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Iceland
- Ireland
- Israel
- Italy
- Japan
- Korea
- Luxembourg
- Mexico
- Netherlands
- New Zealand
- Norway
- Poland
- Portugal
- Russia
- Slovak Republic
- Slovenia
- Spain
- Sweden
- Switzerland
- Turkey
- United Kingdom
- United States

OECD and NEA member
OECD member, not NEA
NEA member, not OECD

OECD Nuclear Energy Agency NEA

NEA Basic Facts and Figures

Governing body: the Steering Committee for Nuclear Energy

31 member countries (24 in the Data Bank)
~ 90% of global nuclear electricity generating capacity.

56 years of international service.

7 standing technical committees (including nuclear development, economics, safety, regulation…).

67 working parties and expert groups.

21 international joint projects funded by participants (17 in the safety area, and others in nuclear science, radiological protection and radioactive waste management).

Technical Secretariat of the Generation IV International Forum (GIF) and the Multinational Design Evaluation Programme (MDEP).

OECD Nuclear Energy Agency NEA

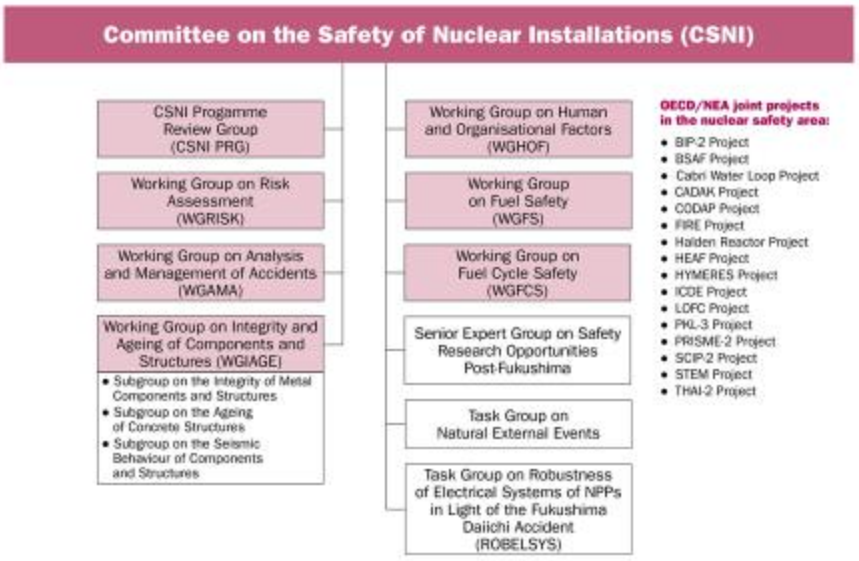
NEA Co-operation and Interactions

- International Energy Agency (IEA): OECD family,
- International Atomic Energy Agency (IAEA): agreement,
- European Commission (EC): full participant,
- China: Joint Declaration on Co-operation (CAEA) and Memorandum of Understanding (NNSA),
- India: expert invitations,
- Other participants and invited experts,
- Industry input to selected studies.

OECD Nuclear Energy Agency NEA

NEA Committees

OECD Nuclear Energy Agency NEA



OECD Nuclear Energy Agency NEA

WGHOF Mission

- Mission of WGHOF
  - To improve the understanding and treatment of human and organisational factors within the **nuclear industry** in order to support the continued **safety performance** of nuclear installations and improve the effectiveness of **regulatory practices** in member countries.
- Composition of the group
  - HOF experts (22 countries represented)
  - Regulators, TSO, Researchers, Operators
  - Representation of : Halden Project, IAEA, EU

WGHOF activities

- Scope of WGHOF activities
  - Act as a **forum** for exchange of information and experience
  - Identify and prioritise current and emergent **HOF issues**
  - Identify HOF **methodologies and practices**
  - Develop shared understanding and **common positions**
  - Facilitate international convergence on safety HOF issues
  - Sponsor specialists meetings, workshops and other means of fostering **international collaboration** with nuclear and other industries
  - advancing the current state of knowledge on HOF related issues important to nuclear safety though studies and technical reviews
  - …

WGHOF tasks and topics

- WGHOF tasks
  - A program and a task leader (CSNI approval)
  - Surveys, workshops, meetings, …
  - Deliverables: task report, proceedings, CSNI Topical Opinion paper
- Current tasks:
  - Establishing desirable attributes of current HRA techniques in nuclear risk assessment
  - Achieving reasonable confidence in validation test results of integrated system performance for nuclear power plant main control rooms
  - Human performance and intervention under extreme conditions

Recap

- The NEA is a valuable forum for cooperation and development of common positions on nuclear issues
- The Working Group on Human and Organisational Factors aims to improve the understanding and treatment of human and organisational factors within the nuclear industry
- The task on validation of integrated system performance is an important element of the overall WGHOF programme

Keynote Presentations

Validation Frameworks

Updating Guidance, Methods, and Techniques for Integrated System Validation

John O'Hara
Brookhaven National Laboratory
New York, 11779
ohara@bnl.gov

One key aspects of the U.S. Nuclear Regulatory Commission's human factors engineering (HFE) safety review is integrated system validation (ISV). ISV as an evaluation using performance-based tests to determine whether an integrated system design (i.e., hardware, software, and personnel elements) meets performance requirements and supports safe operation of the plant. NUREG-0711 provides review criteria for evaluating an applicant's ISV. The objective of the research reported here is to identify issues with ISV and suggest modifications and new guidance to address them. We evaluated nuclear industry ISV experience, technical literature on test and evaluation of complex system, and HFE standards and guidelines. This information was used to identify issues and promising approaches to address them. 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, we identified several high-level issues that have implications across many or all aspects of ISV methodology. They are alternative approaches to ISV, differences between new designs and plant modernizations, validating designs representing new concepts of operations, and grading the ISV effort.

1 Introduction

The United States (U.S.) Nuclear Regulatory Commission (NRC) performs human factors engineering (HFE) reviews of applicant submittals for new plants and for changes to existing plants. The reviews include the evaluation of the methods and tools (M&Ts) used by applicants as part of their HFE programs (NRC, 2007). HFE M&Ts are rapidly evolving. A study by the National Academy of Sciences concluded that advances in the M&Ts used by HFE professionals are revolutionizing the ways personnel are integrated into complex systems (Pew & Mavor, 2007). These advances in M&Ts are impacting HFE programs in the commercial nuclear industry as well (O'Hara, 2010; O'Hara et al., 2009).

As new M&Ts are adopted for use in the commercial nuclear industry, the NRC staff will need up-to-date review guidance to determine that the M&Ts are being used appropriately. In light of their importance to both design and safety evaluations, the NRC conducted evaluations of HFE M&T developments in the nuclear industry and identified 11 topics for potential new guidance development (O'Hara et al., 2009; O'Hara, 2010). These 11 topics were then prioritized, again using a "Phenomena Identification and Ranking Table" methodology to identify those of greater importance with respect to regulatory activities (Molino & O'Hara, 2010). The topics were evaluated by five subject matter experts (SMEs) along two dimensions: importance to regulatory review and the immediacy of the need for guidance. Using this methodology the highest priority topic was integrated system validation (ISV).

NUREG-0711 defines ISV as an evaluation using performance-based tests to determine whether an integrated system design (i.e., hardware, software, and personnel elements) meets performance requirements and supports the plant's safe operation (O'Hara, Higgins, Fleger, & Pieringer, 2012). The NRC uses the review criteria in NUREG-0711 to evaluate the applicant's detailed methodological for performing ISV, including: the validation team, test objectives, testbeds, plant personnel, performance measurement, test design, data analysis and human engineering discrepancy identification, and validation conclusions. An important input to ISV is another aspect of verification and validation which we call the "Sampling of Operational Conditions" (SOC). The review criteria for SOC verify that an applicant's scenarios (1) include conditions representative of the range of events that could be encountered during the

plant's operation, (2) reflect the characteristics expected to contribute to variations in performance, and (3) consider the safety significance of human-system interfaces (HSIs). The review criteria for ISV (and SOC) were originally developed in 1997 (O'Hara et al., 1997) and, although they have been periodically updated, no additional research has been conducted to address ISV since its original development.. Thus, it is not surprising that it was identified as a priority topic.

2 Objectives and Methodology

The objective of this research is to develop HFE review guidance for ISV by updating existing guidance currently contained in NUREG-0711 and preparing new guidance as necessary. The research reported in this paper is the first step in the process; i.e., to identify the aspects of ISV that need to be updated and to identify the technical basis available to support guidance development. This paper will summarize the findings. Readers wishing additional detail are referred to O'Hara (2014). ISV guidance needs and the technical bases to address them were identified from evaluations of nuclear industry ISV experience, general technical literature, and standards and guidelines. The general findings from each are summarized in Section 3. Section 4 will present the technical issues and recommendations.

3 Results

3.1 Nuclear Industry ISV Experience

We evaluated industry ISV experience in three categories: plant modernizations, new plant designs, and NRC reviews of ISV implementation plans. While many, if not most, of the ISVs that have been performed used the NRC's guidance, there has been interest in examining alternatives approaches. One such alternative is a phased approach (e.g., Malcolm et al., 2000) where ISV is performed is stepwise fashion as the design is developed rather than a single evaluation at the end of design.

Much of the industry experience has focused on challenges and the needs for additional guidance in specific areas of ISV methodology. The topics most often identified are:

- ISV challenges that are unique to plant modifications rather than new designs.
- The need to expand the scope of ISV to include evaluations of specific design features, such as automation, alarms, group view displays, computer-based procedures, and controls.
- Determining how much testing to do.
- Designing scenarios; specifically how to identify scenarios and determining how many are necessary.
- Challenges posed by working with simulators, e.g., changes to simulators, competing demands for use of the simulator, simulator functionality and completeness, and crew unfamiliarity with the simulator.
- Challenges posed by performance measure selection; specifically what to measure, how measure aspects of performance of interest, and how much measurement should be used.
- Determining how and where to use expert observers and operator comments.
- Challenges posed by acceptance criteria definition; specifically how to establish criteria (especially for cognitive measures and for opinions and comments from observers and operators).
- Use of benchmark criteria; specifically how to pinpoint when comparisons between a benchmark and new designs are appropriate (in light of plant changes) and how to interpret the results, e.g., how much of a difference is a concern.
- Challenges posed by data analysis; specifically how to examine the data obtained across multiple measures and multiple scenarios, and draw conclusions from them. Sometimes, interpreting the results has been challenging for ISV teams.
- Importance of considering cultural and operational differences when attempting to generalize ISV results from one setting to another, e.g., when a standardized design that has been validated in one

setting (e.g., in the U.S.) is then built in another setting, such as Korea. How are the results generalized?

3.2 General Technical Literature

The general technical literature includes papers on ISV and complex system testing and evaluation that address work from the nuclear industry as well as other industries. We identified several ISV guidance needs in the technical literature and potential technical bases for addressing them. These needs include alternative models of ISV, scenario selection and design, performance measures and the development of criteria, and data analysis.

Alternative models of ISV have been proposed, i.e., the phased approach (discussed in Section 3.1), the usability approach, and the contextual approach. These approaches suggest a number of potential ways in which ISV guidance can be modified including the expansion of ISV activities beyond a single period of testing, expanding the focus to operator interactions with specific aspects of the HSI, and extending the assessment of the integrated system to usability and user experience.

Scenario selection and design is a frequently identified challenge and the guidance provided by edge-centered testing (Chua and Feigh, 2011; Elm et al., 2008) and by Patterson et al. (2010) may provide a methodology to comprehensively address these challenges. 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.

Perhaps the most often identified challenges to ISV teams pertain to the selection of performance measures and the development of criteria to assess the results. A considerable amount of work has been done to define overall performance frameworks and the individual metrics used to assess various aspects of performance (e.g., Braarud & Rø Eitrheim , 2013; Ha et al., 2007, 2009; Hsu, Wu, & Lee, 2006; Lee et al., 2009; Skranning et al., 2013). Despite this work, many questions still remain and echo many of the challenges already expressed:

- The use of multiple measures and their relationship to each other
- The role of cognitive measurement in the ISV process
- The selection of specific measures for use in NPP ISV, e.g., the specific instrument used to measure situation assessment
- The psychometric suitability of measures being used
- The specification of measures to serve as pass/fail criteria
- The development of criteria

3.3 Standards and Guidelines

HFE standards and guidelines (S&Gs) documents play an important role in the design and evaluation of complex systems (Karwowski, 2006). S&Gs provide users with principles to help ensure that the physiological, cognitive, and social characteristics of personnel are accommodated in system development. They also support standardization and consistency of HFE practices. Many HFE S&Gs are developed by professional organizations such as the International Standards Organization (ISO), the Institute of Electrical and Electronics Engineers (IEEE), and the International Electrotechnical Commission (IEC) using a consensus process. Consensus S&Gs are periodically updated to keep them current with new research and technological developments. We reviewed approaches to ISV and system testing and evaluation from ISO, IEEE, and IEC. Government organizations also develop HFE S&Gs. We reviewed the ISV related S&Gs from several organizations, including the Department of Defense (DoD), Federal Aviation Administration (FAA), National Aeronautics and Space Administration (NASA), and Food and Drug Administration (FDA).

Our assessment of the S&Gs varied based of the methodological detail they provided. Many S&Gs provide high-level guidance (or requirements) with little explicit detail. In such cases, we assessed the extent to which the S&Gs are similar to NUREG-0711 in term of general scope and approach. When the S&Gs provide more detail, we compared the guidance to the methodological topics presented in the subsection of NUREG-0711's ISV guidance and organized our evaluation accordingly. We were especially interested in areas where the standards differed from NUREG-0711, since such differences may highlight areas where changes or modifications may be needed.

The S&G documents were largely consistent with NUREG-0711 in their approach to validation. Many S&Gs present high-level requirements for validation and leave the implementation details up to the validation team. There were topics addressed in the S&G's that provide a potential technical basis that can be used to improve NUREG-0711's ISV guidance. Examples include: grading or tailoring ISV, validating multiple interacting plants and support systems, evaluating maintenance tasks, using diverse testbeds, determining sample size, and broadening the conceptualization of performance measures.

4 Technical Issues and Recommendations

This section summarizes the main technical issues identified and also some tentative recommendations for addressing them using the available technical basis. Section 4.1 addresses high-level issues that capture needs associated with topics not well addressed in the NRC's current ISV guidance. Section 4.2 discusses topics where industry experience has identified a need for additional guidance on the detailed aspects of ISV methodology.

4.1 High-Level Issues

Alternative Approaches to ISV

One of the topics identified is whether there are alternatives to the NRC's approach to ISV. We identified several that may provide a basis for improvements. For example, one alternative is the phased approach to validation. Using a phased approach, validations are conducted at various points during the design process rather than being a single evaluation when the design is nearly completed. The APWR (Hanada et al., 2010) and the APR 1400 (Shin et al., 2006) used such an approach. Designers like this approach because it reduces design risk and provides an opportunity to gain confidence in the design. This approach may make sense from a regulatory perspective as well. The ultimate goal of the HFE review is to support personnel and public safety. The earlier issues are defined, the more likely it is that effective solutions can be developed to address them. Malcolm et al. (2000) and, Laarni et al., (2013) provide other examples of using phased approaches to ISV, although the phases were defined differently for each. Some of the technical issues to be addressed for a phased approach to ISV are: (1) defining how are the phases characterized, (2) ensuring that each phase has validation objectives and methods rather than design-oriented tests, (3) integrating the validation results across the phases, and (4) using prior ISV results from a related, but different designs.

Usability (Dumas & Salzman, 2006) and contextual (Savioja & Norros, 2013) approaches are also alternatives approaches that warrant a closer look. Each has technical issues, yet may provide insights to support ISV improvements.

New Designs vs Modernizations

One of the issues identified in a Halden ISV workshop (Braarud et al., 2010) was the need for guidance for ISVs of plant modernizations. Most of the ISVs conducted to date have been for plant modernizations. The ISV guidance in NUREG-0711 is mainly directed at new designs with little methodological consideration for plant modernizations. One very important consideration is that modernizations are implemented in many different ways, e.g., many small modifications, a large modification during a single outage, and a large modification during multiple outages. Many of the

approaches will necessitate some type of phased approach to ISV, especially when plant modernizations are made of a series of outages. After each outage the plant will be operated in a temporary, interim configuration until the full modernization program is completed.

Another issue is that plant modernizations affect personnel in various ways that are unique. Changes to systems and components can impact their role and the way well learned tasks are performed. Modernizations often lead to changes in HSIs, procedures, and training, as well as in the physical equipment. Furthermore, modifications also may involve the HFE aspects of the plant (e.g., the main control room), even though the plant's systems and components are unchanged. ISV should address the challenges to human performance posed by how plant modifications are implemented. The specific challenges posed by each plant's modernization may be unique. Guidance for identifying these types of challenges is needed so they can be addressed in ISV.

Validating Designs Representing New Concepts of Operations (ConOps)

DoD identified a need to extend validation beyond a single system when the mission is accomplished by multiple interacting systems, referred to as "systems-of-systems" (Kometer, et al., 2011). While this need has not been identified in the nuclear industry literature we reviewed, it may be applicable to new designs based on ConOps that are different from current designs. For example, O'Hara, Higgins, and Pena (2010) discussed the ConOps differences between plants current plants and small modular reactors (SMRs), e.g., (1) monitoring and controlling multiple reactor units by a single crew, and (2) the use of the reactors to perform multiple missions, such as power generations, hydrogen production, and industrial steam. The ISV for SMRs may need guidance that reflects multi-unit simulation, especially for plant designs with shared systems. DoD's guidance on validating systems-of-systems may provide some useful input to the development of such guidance.

Grading the ISV Effort

There is a need for additional guidance to define the scope of ISV and to identify the minimum amount of testing needed. A technical basis is available to support such guidance development. For example, IEC 61771 (IEC, 1995) offers some guidance based on degree of innovation and qualification by similarity, especially for evolutionary designs and perhaps plant modernizations. Similarly, DoD has addressed this issue for determining the extent of validation needed for new systems (DoD, 2003b). Sources such as these can provide a basis upon which guidance for grading can be developed.

4.2 Needs Associated With ISV Methodological Details

Most of the ISV needs we identified focus on issues related to the need for additional guidance in specific areas of ISV methodology. These are addressed in this section.

Test Objectives

One of the lessons learned from industry ISV experience is the need for well-defined objectives, since they impact the ISV methodology. Our review identified the potential need for additional guidance with respect to objectives for maintenance tasks and individual HSIs.

Maintenance tasks are a key consideration is DoD's and FAA's testing requirements. While they are also within the scope of a NUREG-0711 review, they are only indirectly referred to in SOC and are not specifically addressed in ISV objectives. Given the importance of maintenance tasks to equipment availability and plant safety (Gertman et al., 2002; O'Hara et al., 1996), more specific guidance for including these tasks should be considered.

Another issue raised by industry experience is the need for more specific objectives related to individual design features and individual HSIs, such as automation, group view displays, and computer-based procedures. This is a fundamental aspect of the contextual approach to ISV (Savioja & Norros, 2013). NUREG-0711's approach focuses on the integrated system, thus any data on the issues related to individual aspects of the design comes indirectly from observers and operators and when doing 'root cause" analysis of performance issues. This approach should be reconsidered. Inclusion of specific objectives for detailed aspects of the design can help support the root cause analysis as well.

Testbeds

ISV experience illustrates the importance of testbeds and we identified several needs related to them. The industry has identified issues related to simulator readiness, e.g., functionality limitations, completeness issues, and newness of the simulator giving rise to less than optimal performance, particularly when using a phased approach to ISV. Additionally, competing demands for use of simulators is an issue. These issues may be partly addressed by adhering to the industry lesson learned to resolve any HED identified through earlier tests and verifications prior to performing ISV. NUREG-0711 reinforces this be recommending that issues identified in verification be addressed prior to ISV.

Another possible solution may be the use of a greater diversity of testbeds. NUREG-0711's ISV guidance is based on evaluations largely being conducted using a full-scope simulator, e.g., ANSI 3.5 compliant (ANS, 2009). However, other platforms may be available and their role in ISV should be explored. One is the use of "other-than-full-scope" simulation capabilities for selected ISV objectives. The FDA validation guidance (FDA, 2011) recommends the use of "real-world" trials to increase realism. Use of actual plant trials may be applicable to modernization programs where the plant is operational. They may also play a role in calibrating results.

There are also times when ISV needs to be conducted and no simulator is available. The NRC encounters this situation when reviewing fuel cycle facilities. Berntson's et al. (2004) three-stage validation (table-top validation early in the design process, a table-top walkthrough using the procedures, and a full operational trial of the final design) may be useful in considering these situations.

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.

Plant Personnel

There is a need for guidance on the number of participants and crews to include in ISV. Currently, no specific guidance is provided for determining sample size, yet this is a significant consideration for several reasons. First, typically a limited number of crews are available for ISV, especially for an ISV at a specific plant (in contrast to a standardized design where crews from multiple utilities can participate). The number is further limited when one considers that operators involved in the design cannot participate in ISV. Second, as the number of crews increases, there is an escalation of the overall cost of conducting ISV. Third, the number of crews impacts the types of data analysis that can be performed.

A technical basis is available that can potentially support improved guidance in this area. For example, DoD (2003a) discusses the use of statistical and operational information to help determine sample size. Appendix B of the FDA HFE guideline (FDA, 2011) provides some guidance on sample size determination that can support the development of improved guidance in this area for ISV.

Scenarios

NUREG-0711, Section 11.4.1, Sampling of Operational Conditions, provides fairly detailed guidance on the operational conditions to be developed into scenarios for ISV. However, there are frequently identified needs related to scenarios that include: determining the number of scenarios to use for ISV, selecting which specific scenarios to use, and designing the detailed scenarios. We found a limited technical basis for addressing these issues. The guidance provided by edge-centered testing discussed in Section 3.2 may help the scenario identification challenge.

However, determining the number of scenarios is complex and needs further research. Some of the technical considerations are similar to determining the number of participants to include. Both are random variables that are appropriately sampled to provide a basis for generalization. Novel approaches to this issue should be considered.

Performance Measures

Issues related to performance measurement are frequently identified, for example:

- What aspects of human-system performance should be measured
- What specific metrics should be used
- What is the role or contribution of different aspects of performance (such as cognitive measures)
- How can better use be made of expert observer and operator comments
- What are the implications of taking data, such as situation awareness ratings, during vs. after scenarios in terms of intrusiveness and what is actually being measured
- What is the psychometric suitability of measures being used, such as construct validity and inter-rater reliability
- How many performance measures should be used
- When using multiple measures, how should convergence between them be established
- How should acceptance criteria be determined (especially for cognitive measures and for opinions and comments from observers and operators)
- When are comparisons between benchmark and new designs appropriate (in light of plant changes) and how should the results be interpreted, e.g., how much of a difference is a concern
- How should pass/fail measures be identified

While a considerable amount of work has been done to define overall performance frameworks and the individual metrics of performance, additional research will be needed to resolve many of the issues identified above. We will summarize some of the technical basis and more detailed issues below.

Conceptualization of Performance Measures

While the validation approaches taken by DoD (DoD, 2009, 2012) and NRC are quite similar, some interesting differences exist that should be explored for application to ISV, specifically:

- the explicit linkage of performance measures to higher-level mission characteristics
- the inclusion of measure of effectiveness (MOE), measures of performance (MOPs) and measures of suitability (MOEs) in the evaluation process
- the use of the mission capability level (MCL) scale to assess how well operators using the system under testing can be expected to fulfill their intended mission in a realistic environment (discussed further in “Data Analysis and Conclusions” below)

The work of Braarud and Rø Eitrheim (2013), Savioja and Norros (2013), and others in the nuclear industry can support the development of improved guidance for this aspect of ISV as well.

More Complete Treatment of Teamwork

The NRC's approach to performance measurement lacks sufficient guidance on teamwork and team processes. More complete treatment of teamwork is needed, such as (1) adding teamwork-specific objectives; (2) ensuring scenarios address work between operators within the control room and between the control room and outside-the-control-room staff and (3) identifying more complete performance measurement for teamwork. IEC 61771 (IEC, 1995) provides one source of information for teamwork.

Usability and User Experience

Many researchers identified the need to include usability and user experience measures (e.g., effectiveness, efficiency and satisfaction) in ISV programs. ISO 11064 (2006) and other publications on usability testing can support the development of such measures for NPP ISV applications. The work on usability and contextual approaches to ISV also address performance measures in this category.

Use of Observer and Operator Subjective Report

Many ISV researchers have called for an increased use of expert observer and operator subjective reports (e.g., Braarud et al., 2010). These reports are based on observations, ratings, and opinions. Subjective reports are not without technical issues, such as determining how to collect this information in a structured reliable manor. However, information of this type can be invaluable for diagnosing performance difficulties encountered by crews.

Consistent with the need identified above concerning the role or contribution of different aspects of performance, the role of usability and user experience metrics needs to be identified. For example, research suggests that user preference and their performance are not highly correlated (e.g., Andre & Wickens, 1995; Bailey, 1993; Barnum et al.; 2004; Nielsen & Levy, 1994). Thus use of such data alone for making design decisions may be questionable. However, the use of such measures in IS might provide information to support HED assessment and validation criteria.

Measures Related to New Technology

Considerations should be given to the use of new measures for new technological developments in NPP design and operations that pose new human performance challenges. An example is the use of "trust" measures. As NPPs become more highly automated, operator trust in automation becomes a very important consideration in operator-system interaction (O'Hara & Higgins, 2010). Another example related to monitoring of multiple SMR units is "neglect time" (Crandall & Cummings (2007) and "change detection/blindness" (Parasuraman et al., 2009; Simons & Ambinder, 2005) that can result when the operator's monitoring burden gets large.

Standardization of Measures

Achieving a consensus between stakeholders, such as applicants and regulators, on what should be measured during ISV tests and the specific metrics that are acceptable would be a worthwhile effort. While not all measures and metrics can be standardized, having some generally accepted approaches would reduce uncertainty considerably, making the development and review of ISV plans simpler.

Data Analysis and Conclusions

Industry ISV experience has identified challenges posed by data analysis, specifically how to examine the data obtained across multiple measures and multiple scenarios, and to draw conclusions from them. ISV teams usually pursue traditional statistical modeling approaches which guards against a Type 1 error at the expense of making Type 2 errors. However, Wickens (1998) has questioned such an approach. He stated that "In the case of a Type II error, it is the user who suffers by not gaining access to a system that was superior and may even be safety enhancing" (p. 19). The importance of Type 2 errors has been addressed in other domains, such as pharmacology. For example, Snow, Reising, Barry, and Harstock (1999) discussed the importance of "practical equivalent," rather than "statistical significance," when comparing a new design to a baseline. They proposed an approach based on bioequivalence testing from drug research. Alternative statistical models should be examined to address the analysis of ISV results.

With respect to formulating conclusions, DoD's (2009) use of the MCL score is an example. The MCL scale is used to assess how well operators using the system under test can be expected to fulfill their intended mission in a realistic environment. The MCL assessment:

- provides a systematic methodology for arriving at MOE and MOS conclusions
- provides a framework for aggregation when multiple critical operational issues exist
- normalizes evaluation results to a common scale, allowing comparisons across systems

An MCL-type approach may be a promising means to compare new designs to baseline designs and combining different measures to support decision making.

There are additional technical issues to be addressed:

- Combining qualitative and quantitative analysis
- Using results from both within and across trial analyses
- Integrating large data sets from many different types of measures and drawing conclusions
- Identifying the factors that need to be considered when generalizing results from one context to others, such as when there are cultural and operational differences

Research should address data analysis issues and provide a basis for guidance improvements.

5 Discussion

The objective of this research is to update the NRC's ISV guidance and prepare new guidance as necessary. In this step of the project, we identified aspects of ISV that need to be updated and the technical basis available to support guidance development. As this work continues, the NRC will seek input from other stakeholders concerned with ISV and will use that information to help prioritize the ISV needs identified. Once this is accomplished, research will be devoted to addressing the selected issues and developing the needed review guidance.

6 References

Andre, A. & Wickens, C. (1995) When Users Want What's not Best for Them. *Ergonomics in Design*, 3 (4), 10-13.

American Nuclear Society (2009). *Nuclear Power Plant Simulators for Use in Operator Training* (ANSI/ANS-3.5-2009). Washington, DC: American National Standards Institute.

Bailey, R. (1993). Performance vs. Preference. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (37), 282-286.

Barnum, C., Henderson, E., Hood, A. & Jordan, R. (2004). A Usability Study of User Preference and Performance. *Technical Communication, 51* (2), 185-206.

Berntson, K., Kozak, A., Malcolm J., McGregor, D. (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).* La Grange, IL: American Nuclear Society.

Braarud, P. Ø. (2010). HWR Lessons learned and insights from using a benchmark approach to ISV. In Braarud, P., Nystad, E., Strand, S., Skråning, G., Bye, A., Hildebrandt, M. & Massaiu, S. (Eds). *The Halden Reactor Project Meeting on Integrated System Validation (ISV) - Status of Current Approaches to ISV and R&D Needs* (HWR-939). Halden, Norway: OECD Halden Reactor Project.

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

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

Chua, Z & Feigh, K. (2011) Integrating Human Factors Principles Into Systems Engineering. In the *Proceeding of the 2011 IEEE/AIAA 30th Digital Avionics Systems Conference (DASC 2011)*. New York, NY: The Institute of Electrical and Electronics Engineers.

Crandall, J. & Cummings, M. (2007). Developing Performance Metrics for the Supervisory Control of Multiple Robots. In *Proceedings of the HRI '07 Proceedings of the ACM/IEEE International Conference on Human-robot Interaction.*

DoD (2012). *Defense Acquisition Guidebook*. Washington, D.C: U.S. Department of Defense.

DoD (2009). *Operational Test and Evaluation Manual* (U.S. Marines Instruction). Washington, D.C: U.S. Department of Defense.

DoD (2003b). *Guidelines for Conducting Operational Test and Evaluation for Software-Intensive System Increments*. Washington, D.C: U.S. Department of Defense.

DoD (2003a). *Test and Evaluation in Support of Systems Acquisition* (Department of the Army Pamphlet 73–1). Washington, D.C: U.S. Department of Defense.

Dumas, J. & Salzman, M. (2006). Usability Assessment Method. In R. Williges (Ed.), *Reviews of Human Factors and Ergonomics* (Volume 2). Santa Monica, CA: Human Factors and Ergonomics Society.

Elm, W., Gualtieri, J., McKenna, B., Tittle, J., Peffer, J., Szymczak, S. & Grossman, J. (2008). Integrating Cognitive Systems Engineering Throughout the Systems Engineering Process. *Journal of Cognitive Engineering and Decision Making, 2* (3), 249–273.

FDA (2011). *Applying Human Factors and Usability Engineering to Optimize Medical Device Design* (Draft Guidance). Washington, DC: U.S. Department of Health and Human Services - Food and Drug Administration.

Gertman, D., Halbert, B., Parrish, M., Sattison, M., Brownson, D., & P. Tortorelli. P. (2002). Review of Findings for Human Performance Contribution to Risk in Operating Events. (NUREG/CR-6753). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Ha, J. & Seong, P. (2009) HUPESS: Human Performance Evaluation Support System. In P. Seong (Ed.) *Reliability and Risk Issues in Large Scale Safety-critical Digital Control Systems.* New York, NY: Springer Link.

Ha, J., Seong, P., Lee, M., & Hong, J. (2007). Development of Human Performance Measures for Human Factors Validation in the Advanced MCR of APR-1400. *IEEE Transactions on Nuclear Science, Vol. 54, No. 6.*

Hanada, S., Mashio, K., Hirahatake, M. & Takahashi, K. (2010). US-APWR Human System Interface System Verification & Validation Program for Digital I&C Design. In *American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-machine Interface Technologies (NPIC & HMIT).* La Grange, IL: American Nuclear Society.

Hsu, S., Wu, T., & Lee, J. (2006). Development of Human Performance Evaluation Battery for Integrated System Validation of the HSI for an Advanced Control Room. In *Proceedings of NPIC&HMIT 2006.* Albuquerque, NM: American Nuclear Society.

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

ISO (2006). *Software Product Quality Requirements and Evaluation - Common Industry Format for Usability Test Reports* (ISO/IEC 25062). Geneva, Switzerland: International Standards Organization.

Karwowski , W. (2006). *Handbook of Standards and Guidelines in Ergonomics and Human Factors*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kometer, M., Burkhart, K., McKee, S. & Polk, D. (2011). Operational Testing: From Basics to System-of-Systems Capabilities. *ITEA Journal, 32,* 39–51.

Laarni, J., Karvonen, H., Koskinen, H.., Liinasuo, M., Norros L, & Savioja P. (2013). *A Stepwise Validation Process for the Main Control Room of Fortum Loviisa Nuclear Power Plant*. Paper Presented at the Enlarged Halden Program Group Meeting. Halden, Norway: OECD Halden Reactor Project.

Lee, M., Hong, J., Suh, J., Lee, S., & Hwang, D. (2009). Development of Human Factors Validation System for the Advanced Control Room of APR1400. *Journal of Nuclear Science and Technology, Vol. 46, No. 1, p. 90-101.*

Malcolm, S., Holford, K. & Gillard, D. (2000). HF Verification and Validation Activities: Simulator Based Operational Trials. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

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

Nielsen, J. & Levy, J. (1994) Measuring Usability: Preference vs. Performance. *Communications of the ACM*, 37 (4), 66-75.

NRC (2007). *Standard Review Plan*, Chapter 18 - Human Factors Engineering (NUREG-0800). Washington, DC: U.S. Nuclear Regulatory Commission.

O'Hara, J. (2014). *Integrated System Validation: Models, Methods, and Issues*. (BNL Technical Report No. 6859-1-2010, Rev 2). Upton, NY: Brookhaven National Laboratory.

O'Hara, J. (2010). Current Trends in HFE Methods and Tools Update (BNL Technical Report No. 6859-1-2010, Rev 2). Upton, NY: Brookhaven National Laboratory.

O'Hara, J, & Higgins, J. (2010). *Human-System Interfaces to Automatic Systems: Review Guidance and Technical Basis* (BNL Technical Report 91017-2010). Upton, NY: Brookhaven National Laboratory.

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

O'Hara, J., Higgins, J. & Pena, M. (2012). *Human-Performance Issues Related to the Design and Operation of Small Modular Reactors* (NUREG/CR-7126). Washington, D.C.: U. S. Nuclear Regulatory Commission

O'Hara, J., Plott, C., Milanski, J., Ronan, A., Scheff, S., Laux, L. & Bzostek, J. (2009). *Trends in HFE Methods and Tools and Their Applicability to Safety Reviews* (BNL Technical Report BNL-90424-2009). Upton, NY: Brookhaven National Laboratory.

O'Hara, J., Stubler, W., & Higgins, J. (1996). *Hybrid Human-System Interfaces: Human Factors Considerations* (BNL Report J6012-T1-4/96). Upton, New York: Brookhaven National Laboratory.

O'Hara, J., Stubler, W., Higgins, J. & Brown, W. (1997). *Integrated System Validation: Methodology and Review Criteria* (NUREG/CR-6393). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Parasuraman, R., Cosenzo, K. & De Visser, E. (2009). Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change Detection, Situation Awareness, and Mental Workload. *Military Psychology, 21*(2), 270-297.

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

Pew, R. & Mavor, A. (2007). *Human-system Integration in the System Development Process: A New Look*. Washington, D.C.: The National Academies Press.

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

Shin, Y.-C., Moon, H-K. & Jong-Hyun Kim, J.H. (2006). Human Factors Engineering Verification and Validation for APR1400 Computerized Control Room. In proceedings of *Fifth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-machine Interface Technologies (NPIC & HMIT).* La Grange, IL: American Nuclear Society.

Simons, D. & Ambinder, M. (2005). Change Blindness: Theory and Consequences. *Current Directions in Psychological Science*, *14*(1), 44-48.

Skråning, G., Bye, A. Nystad, E., & Eitrheim, M. (2013). Workshop Meeting on Human Performance Measurement for Simulator Experiments in Nuclear Process Control (HWR-1034). Halden, Norway: OECD Halden Reactor Project.

Snow. M., Reising, J., Barry, T., and Harstock, D. (1999). Comparing new designs with baselines. *Ergonomics in Design* (October), 28-33.

Wickens, C. (1998). Commonsense Statistics. *Ergonomics in Design*, October, 18-22.

Updating Guidance, Methods, and Techniques for Integrated System Validation (ISV)

John O'Hara
Brookhaven National Laboratory

WGHOF Experts Workshop on Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications

February 2015

BROOKHAVEN NATIONAL LABORATORY
a passion for discovery

Office of Science

Topics

- Background
- Overview of ISV Review Guidance
- Project Objectives
- Technical Approach to Guidance Update
- Preliminary Results
- Discussion

Background

- The concept of control room validation has been around since the importance of human performance to plant safety became evident following the Three-Mile Island (TMI) accident
- In 1981, NUREG-0700, Rev 0 was published as part of the U.S. Nuclear Regulatory Commission's (NRC) TMI actions to improve safety
  - validation was part of the overall process

Section 3.2.3 Process 6 – Validation of Control Room Functions and Integrated Performance Capabilities

Control room adequacy must also be considered from the perspective of integrated functional requirements. Process 6 involves examination of the interactions and dependencies of the operating crew and equipment. (p. 3-4)

Background

- The concept of control room validation included:
  - a focus on functional aspects of control room design
  - integration of personnel and equipment
  - consideration of performance
- In the early 1980's, technology imposed some limitations on what could be accomplished
  - simulation was not integral to the engineering design and evaluation process
  - simulation technology was not as developed as it is today
- Advances in the use of simulation and in human factors engineering (HFE) methods and tools has improved our ability to perform control room validation

Background

- Today, the NRC defines integrated system validation (ISV) is an evaluation to determine whether an integrated system's design (hardware, software, and personnel elements) meets performance requirements and supports safe operation
- The NRC uses NUREG-0711 to review an applicant's ISV methodology
- The review guidance was originally developed in 1997
  - NUREG/CR-6393
- While the guidance has been periodically updated, no additional research has been conducted to address ISV since its original development

Background

- The need for additional ISV guidance was identified in NRC research
  - One study conducted in 2008 examined a broad range of issues associated with emerging technology
    - 14 subject matter experts (SMEs) evaluated the issues
    - ISV was identified as one of the top-priority issues
  - Another study conducted in 2010 focused specifically on topics related to HFE methods and tools
    - five SMEs evaluated the topics
    - ISV was the highest rated topic
- Based on these findings, the NRC initiated an effort to develop additional guidance

Overview of ISV Review Guidance

- ISV is an evaluation of the integrated design (hardware, software, procedures, and personnel)
- The evaluation is challenging
  - it's not possible to test all operational conditions, so decisions have to be based on a sample of conditions
  - it's not possible to test the actual system in vivo for many scenarios, so decisions have to be based on simulations of scenarios
  - it's not possible to test all operators/crews, so decisions have to be based on a sample of operators

Overview of ISV Review Guidance

- The methodology to meet these challenges involves
  - performance-based testing
  - good sampling processes
  - representing the integrated system and the scenarios to be performed as realistically as possible
  - exercising good test practices to minimize error and bias
  - collecting appropriate measures of performance
  - comparing observed performance to acceptance criteria

Overview of ISV Review Guidance

- ISV review topics
  - ISV team
  - test objectives
  - testbeds
  - test participants
  - scenarios
  - performance measurement
  - test design
  - data analysis and human engineering discrepancy (HED) identification
  - validation conclusions

Overview of ISV Review Guidance
Methodology Elements

- Team
  - unbiased
- Test objectives
  - staffing levels
  - human-system interfaces (HSIs)
  - error management
  - important human actions
- Testbeds
  - high-fidelity simulator
  - near final HSI and procedure designs

Overview of ISV Review Guidance
Methodology Elements

- Test participants
  - actual operators
  - trained on the new design
- Scenarios (sampling of operational conditions)
  - sample of plant conditions, personnel tasks, and situational factors
  - identify and define scenarios to embody sampled conditions
- Performance measurement
  - types of measures
    - comprehensive measurement approach, including plant, task, and cognitive measures
  - characteristics
  - criteria
  - selection of pass/fail measures

Overview of ISV Review Guidance
Methodology Elements

- Test design
  - repeated trials to account for performance variability
  - rigorous methods to avoid bias and noise
- Data analysis and human engineering discrepancy (HED) identification
  - qualitative and quantitative factors
  - convergence of measures
  - HED identified when performance doesn't meet criteria
  - HED resolution
- Validation conclusions
  - bases for determining that performance of the integrated system is/will be acceptable

Project Objectives

- Our objective is to revise the ISV review guidance
  - updating existing guidance currently contained in NUREG-0711
  - preparing new guidance as necessary
- The current research is the first step in the process
  - to identify the aspects of ISV that need to be updated
  - to identify the technical basis available to support guidance development

Technical Approach to Guidance Update

- Evaluation of technical information from several sources
  - nuclear industry ISV experience
  - standards and guidelines
  - general technical literature related to validation topics
- Nuclear industry ISV experience
  - plant modernizations
  - new plant designs
  - NRC reviews of ISV plans

Technical Approach to Guidance Update

- Standards and guidelines
  - consensus standards
    - International Electrotechnical Commission (IEC)
    - International Organization for Standardization (ISO)
    - Institute of Electrical and Electronics Engineers (IEEE)
  - government requirements and guidance documents
    - Department of Defense (DoD)
    - National Aeronautics and Space Administration (NASA)
    - Federal Aviation Administration (FAA)
    - Food and Drug Administration (FDA)
    - Canadian Nuclear Safety Commission
    - Swedish Radiation Safety Authority

Technical Approach to Guidance Update

- General technical literature related to validation topics
  - alternative approaches to ISV
  - scenario selection
  - scenario design
  - performance measures
    - selection of performance measures
    - selection of pass/fail measures
    - identification of criteria
  - data analysis

Preliminary Results

- We identified technical issues from each of the sources we examined and some tentative recommendations for addressing them
- Issues were organized into
  - *high-level issues* that capture bigger picture topics
  - *detailed issues* that capture the need for additional guidance on specific aspects of ISV methodology
- A summary of these issues is presented in next six slides

Preliminary Results - High-Level Issues

- Alternative approaches to ISV
  - approaches to validation that are different from the NRC's approach
  - examples
    - phased (stepwise) approach
    - usability approach
    - contextual approach
- New designs vs modernizations
  - NUREG-0711 is mainly directed at new designs
  - most of the ISVs conducted to date have been for plant modernizations
  - modernizations have unique issues
    - differences in how modernizations are implemented
    - may necessitate some type of phased approach to ISV, especially when plant modernizations are made across a series of outages

Preliminary Results
- High-Level Issues (continued)

- Validating designs representing new concepts of operations (ConOps)
  - new missions, such as hydrogen production and industrial steam
  - multi-unit monitoring and control
- Grading the ISV effort
  - scoping ISV to identify the minimum amount of testing needed

Preliminary Results
- Detailed Issues

- Test objectives
  - consider objectives more broadly, e.g.:
    - maintenance tasks
    - individual design features of HSIs
- Testbeds
  - consider alternative testbeds
    - simulator readiness for ISV
    - competing demands for simulator
    - ISV when no simulator is available
- Plant personnel
  - additional guidance on sample size
- Scenarios
  - number of scenarios
  - selecting which scenarios to use

Preliminary Results
- Detailed Issues (continued)

- Performance measures
  - aspects of performance not addressed well in current guidance, e.g.:
    - higher-level measures, such as measures of effectiveness
    - teamwork
    - usability and user experience
    - expert observer and operator comments
    - measures related to new concepts of operations and new technology (e.g., trust in automation)
  - contribution of different aspects of performance
    - such as cognitive measures
  - number of performance measures to be used
  - specific metrics to be used
  - implications of taking data during vs. after scenarios
    - intrusiveness of situation awareness ratings

Preliminary Results
- Detailed Issues (continued)

- Performance measures (continued)
  - establishing the psychometric suitability of measures
    - construct validity
    - inter-rater reliability
  - choosing pass/fail measures
  - determining acceptance criteria
    - especially for cognitive measures and for opinions and comments from observers and operators
  - interpreting results when comparing new designs to benchmarks, e.g., how much of a difference is a concern
  - establishing convergence when using multiple measures
  - establishing an accepted standardization of measurement categories and metrics

Preliminary Results
- Detailed Issues (continued)

- Data analysis and conclusions
  - combining qualitative and quantitative analysis
  - alternatives to traditional statistical tests for data analysis and interpretation, e.g., equivalence testing (Snow, Reising, Barry & Harstock, 1999)
  - analyzing data across multiple scenarios and measures
  - using results from both within and across trial analyses
  - integrating large data sets from many different types of measures and drawing conclusions
  - identifying the factors that need to be considered when generalizing results from one context to others
    - such as when there are cultural and operational differences

Preliminary Results
- Resolving Issues

- Resolution of these issue will require coordinated research and development efforts on the part of the commercial nuclear power community
- The issues are often more complex then they appear to be at face value
  - one example is the use of operator comments and evaluations
    - operators of often asked their preferences
    - yet there is ample research to show operators often prefer designs that are not associated with better performance
    - we need to determine the best way to obtain and use operator comments and evaluations to best meet evaluation goals
- Issues are often inter-related
  - example – determining how many teams to participate impacts decisions concerning how to analyze the data

Discussion

- In this step of the project, we are identifying aspects of ISV that need to be updated and the technical basis available to support guidance development
- As this work continues, the NRC will seek input from stakeholders concerned with ISV and will use that information to help prioritize the ISV needs identified
- Once this is accomplished, research will be devoted to addressing the selected issues and developing the needed review guidance

Leena Norros

Preliminary validation – or a life-cycle perspective to validation of complex systems

Preliminary validation (PV) is a concept and process that has been considered necessary for ensuring the quality of complex tools, e.g. the nuclear power plant control rooms, for their intended use. One possibility to define what is meant by PV is to contrast it to the integrated system validation (ISV) that has been defined in the existing standards and guidelines. In an ISV the evaluation activity focuses on the safe functioning of a whole complex tool, and the evaluation serves decision for acceptance of this system for use. In a preliminary validation, the evaluation focuses on parts of the whole and guides design so that a "good" tool will be achieved. Preliminary validation is independent from design, and it tests the achievements of the design against a conceptual reference of a good outcome. Clearly, a concept of the good outcome needs to be defined. Moreover, as the tool under evaluation is not yet ready, 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. The professional users, due to their insight of the requirements of the work, play a key role in foreseeing the appropriateness of the future tool. Therefore, the contribution of the professional users must be facilitated by exploiting appropriate test methodology.

The presentation will give arguments for extending the validation approach by a preliminary validation. Some methodological and methodical key characteristics of the preliminary validation will be considered. These include:

- *Defining what is a good control room.* Introduction of the concept of Systems Usability, and connecting it to an extended concept of safety. The latter emphasizes the contribution of human operators to safety via creating resilience to the sociotechnical system.
- *Providing reasons for extending ISV with preliminary validation.* Preliminary sub-system validations are needed for providing a sufficiently detailed focus and sufficient coverage in the evaluation of very complex tools, the intended concept of operations and the operating procedures.
- *Integrating the PV in the design process and its quality control.* The detailed regime of sub-system validations must be designed case-by-case and depending of the type of the overall design process. Also the connection between PV and ISV must be defined.
- *Developing of a formative methodology including qualitative methods in the acquisition of the evaluation results.* This is needed in order to improve the predictive value of the evaluation. Traditional quality criteria for evaluations need to be completed by new criteria.
- *Creating a systems usability case to establish connection to the design requirements, to structure the reasoning about the evaluation results, and to support the accumulating of results.* The systems usability case can be considered as a living document.
- *Considering the role and timing of operator training in the design process.* The basic idea needs to be acknowledged that a tool is not a tool before people have appropriated it for a meaningful use.

The presentation will exploit the experience that the VTT Technical Research Centre human factors research group has gained in a joint project with Fortum Nuclear Services on the renewal of the I&C and

control room systems of the Loviisa NPP. The experience demonstrates an intimate connection between the constructive and evaluative functions in design. It also shows that a necessary independence between these two functions requires new conceptual means for defining the targets of design, for creating test results about the use the designed solutions, and for reasoning about the acceptance of the solutions. Finally the experience about PV supports adoption of a life-cycle perspective to validation, and sees it as a realistic path for the future development of validation methodologies.

Preliminary validation – or a life-cycle perspective to validation of complex systems

Leena Norros
Research professor (emerita)
VTT Technical Research Centre of Finland

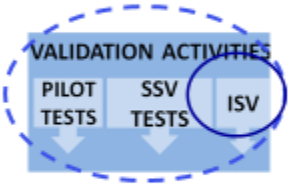
Experts' Workshop on Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications, February 19-21, 2015

CONTENTS OF THE PRESENTATION

I. Introduction
II. Case Loviisa I&C renewal
III. Procedure in a multistage V&V process
IV. Underlying methodological choices
V. Conclusions

I. INTRODUCTION

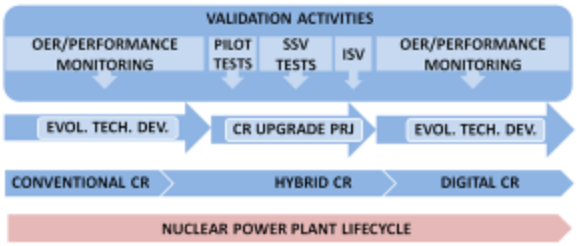
Currently expressed aim is to extend the V&V ....



Optional terms for the extension of the V&V concept:

- Preliminary validation
- Phased approach
- Stepwise validation
- Multistage validation
- Sub-system validation SSV

… but maybe it would necessary to put V&V in the plant life-cycle perspective

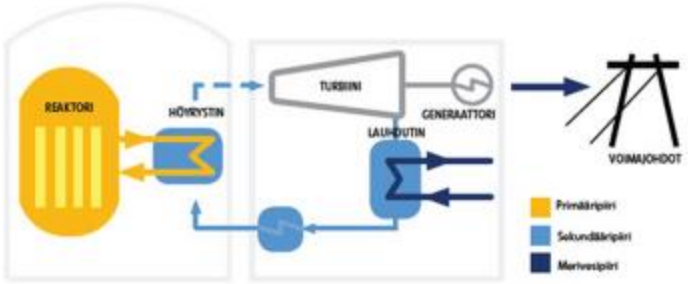


- The life cycle of control rooms stretches over decades
- The technical bases of automation and control systems change radically
- Several development phases during their operative use
- Human factors data is gathered during the life cycle with multiple methods

Extending the V&V concept

- There is a need to
  - Systematise HF data
  - Build a holistic picture of the fulfilment of human factors and safety requirements
  - Integrate HF knowledge better into the design process

II. CASE LOVIISA I&C RENEWAL



Specifics of Fortum Loviisa automation and control room upgrade

- VVER type nuclear power plant in Loviisa in Southern Finland, two units commissioned in 1977 and 1980
- I&C designed by Siemens
- I&C and control room modernization project started in 2005, and was planned to take place in several phases

- Fortum has an in-house design department with
  - A competent CR interface design team, including a HF specialist
  - A full-scope APROS plant simulator for design and testing
- According to the DiD-based safety architecture the functions of the main automation system and the safety automation were separated, as well as corresponding HS-interfaces
- Changes in the concept of operations were anticipated

Acknowledged general requirements for HFE V&V process

- National regulatory standards and guidelines (YVL-guides)
- ISO 11064-7: Ergonomic design of control centres. Part 7: Principles for the evaluation of control centres
- IEC 1771: Nuclear power plants – Main control-room – Verification and validation of design
- NUREG-0711: Human Factors Engineering Program Review Model
- NUREG/CR-6393: Integrated system validation: Methodology and review criteria

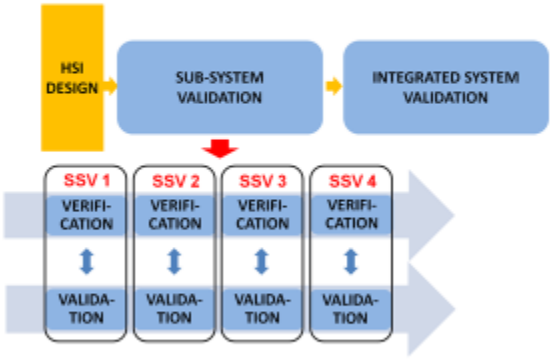
Requirements for multistage V&V at the Loviisa plant

- Phased approach
- Graded approach
- Requirement-based
- Structured and well-documented
- Continuous

Phased approach

- Multiple stages can be recognized on different levels
  - **Technological** upgrade phases e.g. from conventional to hybrid and to digital
  - **Project** may be divided in phases, for example first renewal of reactor I&C and then turbine I&C
  - **Within a project phase** verification and validation can be conducted in several steps with focus on different sub-systems
  - **Within one test** the focus can be put on different aspects of operator activity
  - Distinction between **verification** that correctly designed, **validation** that design is correct

Phased approach in the Loviisa case – Sub-system validation (SSV) (within each project phase)



Requirement-based

- Main reference of evaluation is requirements as opposed to normative or benchmark-referenced approaches
- In a requirement-based evaluation the general claim of system safety is divided further into design requirements

…. and into theoretically derived requirements or claims **TBD**

- The theoretical reference would qualify the evaluation as a validation (in contrast to verification)

Graded approach

- Utilized in selecting the sub-systems to be evaluated in the multistage V&V process
- Depth of treatment is based on the risk and importance that is associated with each modification
  - Main criteria for the grading are *safety criticality* and *degree of novelty*
- In grading the largest effort is placed on the evaluation of the most critical systems

Structured and well documented. Continuous

- Proposed was to exploit the idea of Safety Case (Bishop & Bloomfield 1998) to develop a "Systems Usability Case"
- Producing Systems Usability Case is:

  Creating an accumulated documented body of evidence throughout the design process that provides a convincing and valid argument of the degree of systems usability of a system for a given application in a given environment.

- VTT had tested the idea earlier with regard to evaluating an innovative design concept (EU-MMOTION Norros et al. 2011)

III. THE EVALUATION PROCEDURE

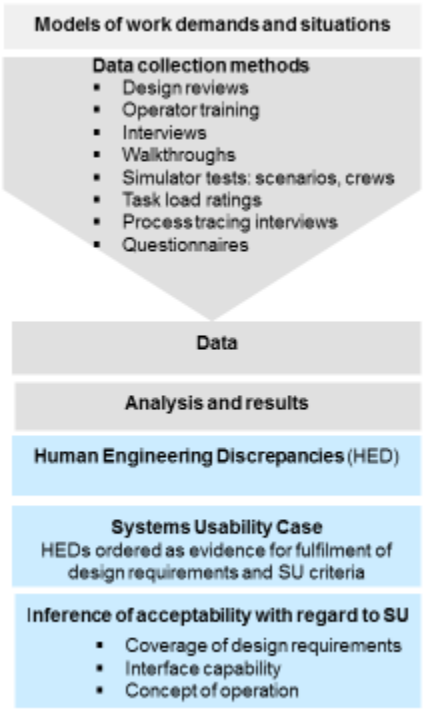
The rationale of sub-system validations (SSV)

- SSVs are needed for providing a sufficiently detailed focus and sufficient coverage in the evaluation of very complex tools

NOTE

- In SSVs the focus is on functional parts of a system which are tested with regard to their capability to support safe work – i.e., the parts are embedded in an emerging whole
- A conceptual apparatus is needed to represent the whole during design

Phases of the evaluation process in each SSV



IV. METHODOLOGICAL CHOICES

1. Resilience perspective to safety

- Aim is at ensuring that the designed system adequately supports plant personnel in the safe operation of the plant processes
- The need to elaborate the concept of safety has been raised
  - "Safety 1" considers safety as non-existence of failure
  - "Safety 2" extends the concept by considering safety as capability to manage the unexpected, i.e. resilience

- Safety 1 is the starting point of current evaluation methods – adequate and failure free performance outcome (in tested situations) needs to be shown
- Safety 2 is applied in SSV - both adeqaue performance outcome, and way of operating demonstrating generic capability for approporiate acting need to be shown
- The concept of Systems Usability focuses on the capability for adequate acting

2. Evaluation tests portray formative characteristics

- A new attribute to the quality of testing is applied : the test should be formative, i.e., developmental
  - Aim is at creation of new knowledge – not confirmation of a hypothesis
  - Encourages the professional users in constructing new knowledge during tests
- Main instruments
  - Valuing the professional users' perspective – good ethnography including user experience
  - Use of the "double stimulation" approach (Vygotsky), i.e., provide external tools for problem solving (operators' attention on the tools not process problem; the specific functionality issues of interfaces registered and reflected dialogically)
  - Generalized solution – CR evaluation session among operators and establishing the Systems Usability Case

3. Contextual approach to human activity

- In the proposed evaluation methodology a contextual approach to human activity is applied
  - Activity is object-oriented – definition of the core task
  - Activity is mediated by tools – definition of necessary tool functions
- The approach is applied in defining the Systems Usability quality attribute:

Systems usability (SU) denotes the capability of the technology to fulfil the instrumental, psychological, and communicative functions of a tool in the activity and to support the fulfilment of the core-task functions in the work.
Systems usability is evidenced in technology's usage by appropriate performance outcome, way of acting, and user experience.

(Savioja 2014, Savioja & Norros 2008, Norros et al. submitted)

The Systems Usability concept

Core Task

	Performance: Outcome	Way of Acting: Orienting to core-task	User Experience: The development potential of use
Instrumental Function	Effective	Appropriate	Professional quality
Psychological Function	Effortless	Meaningful	Embodied. Flow
Communicative Function	Interactive	Shared	Value for actors

4. Focus on the joint human-technology system

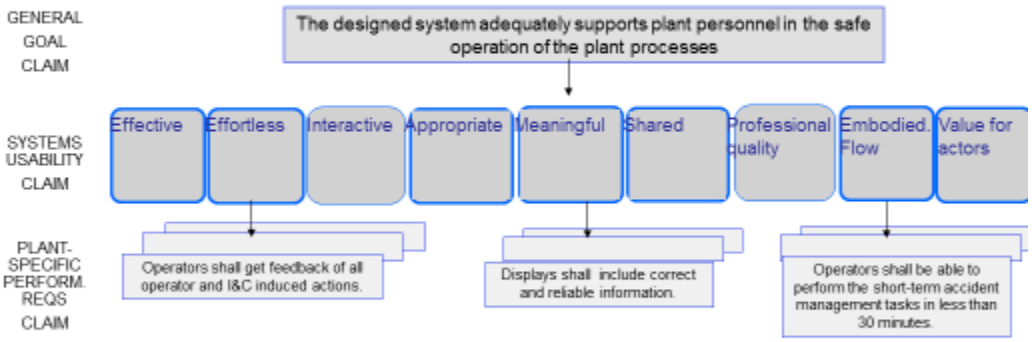
- Correlative results between type of control room and operator performance are not informative about the H-T teamwork
- It is necessary to understand the mutual mechanism by which certain features of the tool and human users form a new teamwork
- Professional user experience is valuable in revealing
  - The mechanism of forming H-T teamwork
  - The promise of the tools in their future use
  - What training is relevant
- Expert observation is effective in judging the functionality of the joint system

5. Systematic way of making assessments and accumulating of results – Systems Usability Case

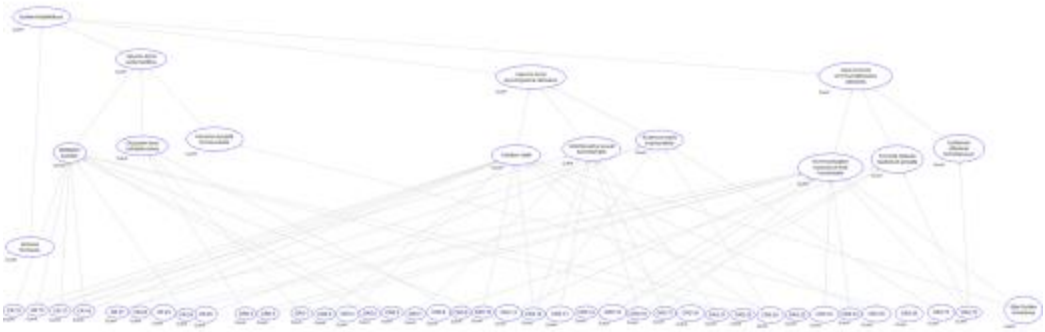
- Assessment deals with interpreting the significance of results for safety
- A reference is needed - the reference is SU requirements

- Using the Systems Usability Case the SU and design requirements are defined as CLAIMS
- Included are also EVIDENCE that are interpreted to either confirm or disconfirm a certain claim
- ARGUMENT explains the mechanism for the identified connection between the evidence and claim.

Requirements as claims

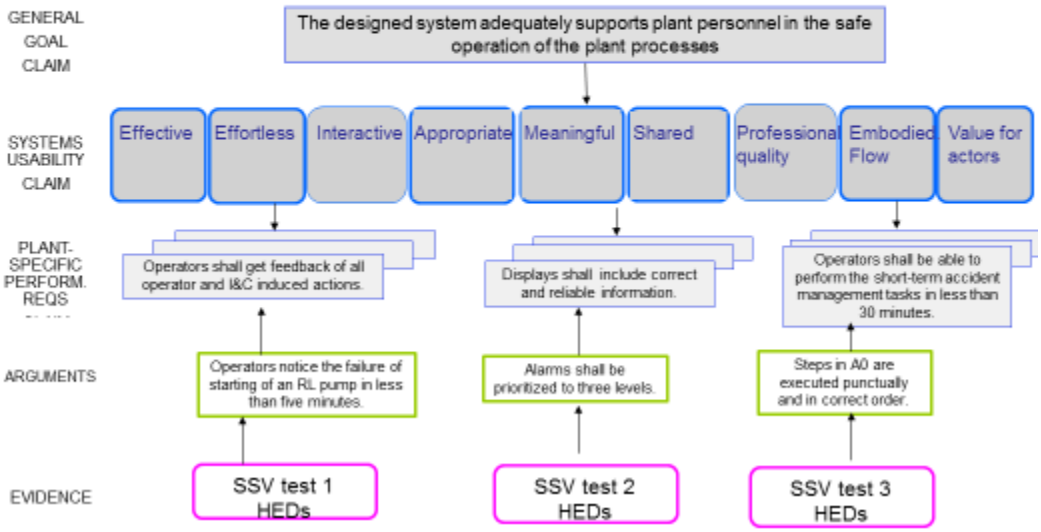


An example of the Loviisa SSV1



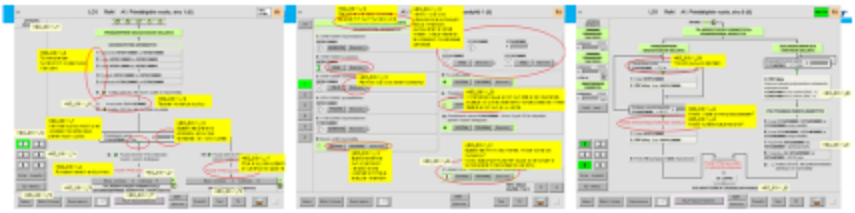
- Selected relevant design requirements were mapped to systems usability claims (SSV1)
- The ASCE tool is used

Evidence and arguments to test the fulfilment of requirements



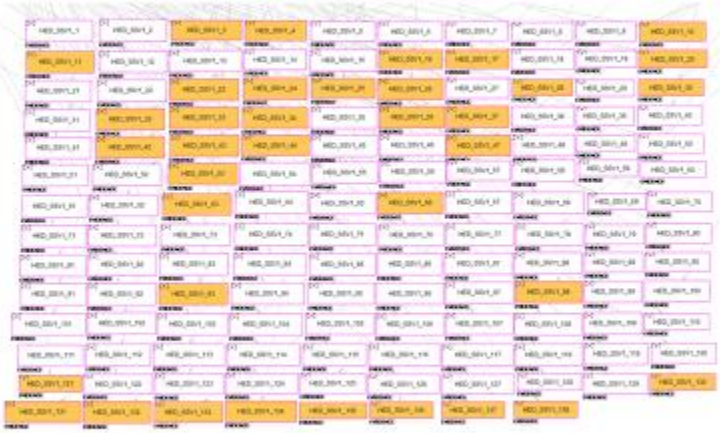
Example of Loviisa SSV1: Human Enegineering Discrepeancies (HED) represented on to the user interface

- Altogether 138 Human Engineering Discrepancies (HED) were identified in the analyses of the test data



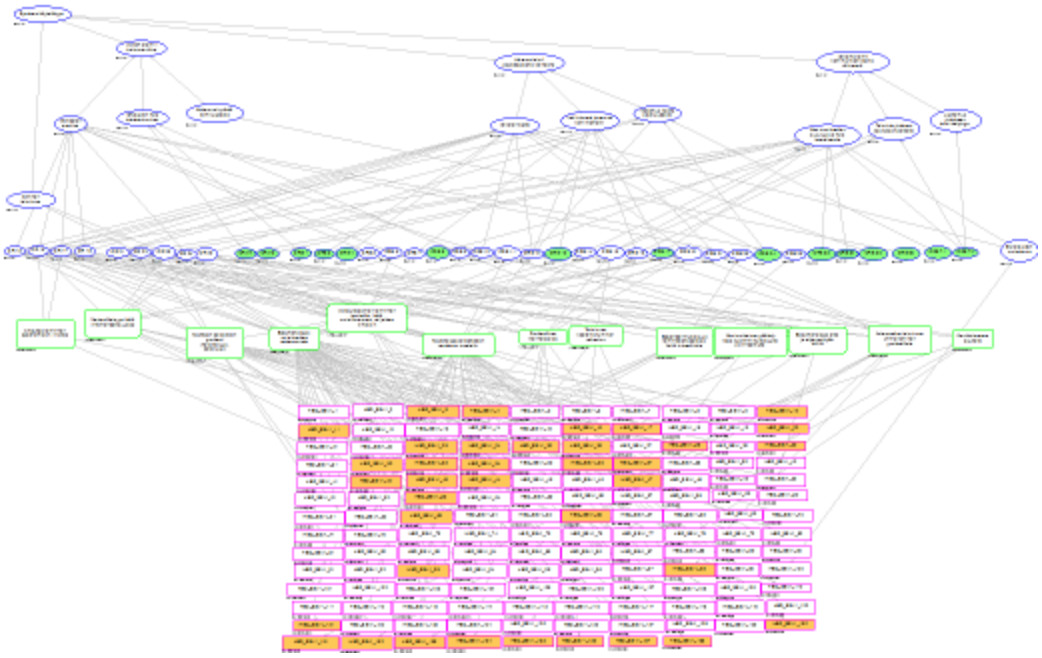
- 96 concerned user interface features
- 42 concerned the operational concept

Example of Loviisa SSV1: Identified Human Engineering Discrepancies



- Each HED was given a unique identifier and a description in a separate file (orange refers to concept level discrepancy)

Example of Loviisa SSV1: Totality of the Systems Usability Case



Example of Loviisa SSV3: Arguments as source of new knowledge of the interface

Type of argument	Argument
Insufficent HSI quality	Against usability principles **Against Fortum styleguide** Discrepancy between procedure and display Error in interface design
Impedes activity or increases workload	Does not serve activity **Hinders fluent activity** Insufficient SA or formation of SA **Impedes crew coordination** Supervisor role hindered Insufficcient automation awaraness Impedes control of own activity Workoad decreases reliability of actions Wording of procecdure impedes action or collaboration Increases secondary task workload
Error or identified error possibility	Possibility for false interpretation **Possibility for failure in operation or procedure use, or due to overconfidence with automation** Reduction of reliability of action
Number of arguments	278

Aspects in assessment of acceptance

- **Coverage of design requirements with regard to systems usability**
  - Results from mapping design requirements to systems usability claims
- **The capability of the human-system interfaces of the CR**
  - Evidence concerning fulfilment of claims (design requirement claims and SU claims)
  - Arguments inform of the mechanism of not fulfilling the claim
- **Maturity of the concept of operations**
  - Emerging insights of the concept of operation

Form of the assessment

- Assessment is a narrative
- It is based on the evaluation team's (including human factors and domain experts) shared expert opinion
- It focuses on each of the aspects of assessment
- It makes explicit
  - Need for design improvements regarding the tested systems
  - Development from the previous evaluations
  - Need for re-testing or testing in further evaluations

Benefits of the Systems Usability Case method in HFE validation

- Systematic conductance of validation tests
- The chain of reasoning is explicit and transparent
- Enables evaluation of the design solution
- Enables a review of fulfilment of requirements in a longitudinal manner
- Accumulating results support focusing of the integrated system validation ISV

=> An "Integrated Phased Validation" IPV approach emerges

V. CONCLUSIONS

The "IPV" approach

- Bridges gaps between
  - design and evaluation
  - sub-system and integrated system validations
- Facilitates construction of new knowledge of the human-technology teamwork
- Enables independent evaluation and realistic objectivity via theoretical generalization, formative approach and systematic treatment of data
- These features need to be developed further

THANK YOU!

Invited Presentation – Final Integrated Control Room Validation

Scott Malcolm
Candu Energy Inc.

This presentation contains the views of the author developed over 20 years of experience in planning and completing validation work on both new control room designs and modifications to existing facilities. It will cover experiences and lessons associated with establishing design requirements, completing design work and consolidating all of these inputs into Final Integrated Control Room Validation (FICRV). The presentation will also provide suggestions for further enhancements that will bolster confidence in the results – the theme of the present workshop.

FICRV has various definitions, although most are similar and codified in industry consensus design standards and regulatory requirements or guidance. To this end there has been good convergence of definitions in recent years with the creation or revision of standards and guides in the nuclear industry. Broadly, the purpose of FICRV is to validate that the integrated Control Room (CR) system supports safe operation of the plant and thus drives everything the design team will do up to and including the final tests.

This presentation will discuss approaches and methods adopted for validating integrated CR systems, where integrated means containing the design (hardware, software), either actual operating procedures or example operating procedures (new plant design), and typical or actual licensed operators and shift supervisors. It is important to distinguish FICRV as it relates to design from validation of site operating procedures, training programs and staff licensing programs. The latter have a separate edifice of requirements and processes to achieve their ends. For the design team these aspects are approximated and provided as inputs to site specific operating organizations for their use in meeting licensing requirements.

The 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, in the end the conclusion is acceptable or not. If not, identifying the remedies for making it acceptable fall to the design team and this may include issues for the operating organization to resolve, although these must be limited or the plant owner will return the issues to the design team.

The confidence in the validation must be formulated in a manner that allows this acceptable/not acceptable decision to be made clearly. Validation of design work is a common and expected aspect of design from the hardware and software to plant layout and civil design. In simple terms, FICRV is a test or series of tests that are completed to determine if requirements have been met. This is most clearly seen when tests fail and the subsequent remedies. Remedies that may be seen as new requirements for design are unwelcome additions to any project.

The genesis of the validation plan and hence the validation it self, 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 ground and must be navigated with careful consideration. In fact, I will argue here that there is much to be gained in advancing our confidence in the outputs by improving the inputs - the requirements. Thus, it is imperative to have the challenging discussions early.

Let me be clear - setting requirements for CR design with respect to the integrated system have come a long way in the last 25 years and are producing designs that are demonstrably better than what went before. To-date, the design of new control rooms and modifications has produced acceptable results and hence are authorized for construction. We can and must do better because the bar for acceptability of all aspects of plant design and operations is continually moving higher.

To this end my presentation will look at existing methods, challenges and special considerations in conducting FICRV concluding with next steps .

Invited Presentation
Final Integrated Control Validation Thoughts on Past and Future

Scott Malcolm, SNC-Lavalin

US NRC Special Workshop on Control Room Validation

2015 February, USA

A world leader

SNC-Lavalin's Nuclear team provides leading nuclear technology products and full-service solutions to nuclear utilities around the globe. Our team of 1,300 engineering, procurement, construction and project management experts offer customized operations, maintenance and plant life management services, including waste management and decommissioning for light water and CANDU-type reactors. Our experts in nuclear steam plant and balance of plant engineering carry out life extension projects, and design and deliver state-of-the-art CANDU® reactors, which are capable of operating on many types of fuel including natural uranium, mixed oxide (MOX) fuel, recycled uranium (RU) and thorium.

Outline – Final Integrated CR Validation

Introduction

Methods

Challenges

Special Considerations

Requirements

Next Steps

Introduction

In the role of invited paper for a workshop with a distinguished audience

- Get right to the point
- Little time on definitions and background
- Share topical examples and thoughts from 25 years of CR design work
- Provocative to stimulate workshop discussion

Acknowledgenment

- To the late, great Dr. J. Persensky – colleague and friend

Introduction

In the role of invited paper for a workshop with a distinguished audience

- Get right to the point
- Little time on definitions and background
- Share topical examples and thoughts from 25 years of CR design work
- Provocative to stimulate workshop discussion

Acknowledgenment

- To the late, great Dr. J. Persensky – colleague and friend

Introduction…cont'd

Context for design validation

- Standards and Guides
  - IEC, IEEE, national standards in nuclear power for both programmatic and detailed design activities (both consensus and public comment processes)
  - Generally good convergence
  - Most robust basis for establishing and integrating human based design activities and requirements

Today we have…

HF design plans, evolved over many years of lessons

High fidelity test beds

Integration with allied design disciplines such as I&C, process, mechanical and civil engineering

Broad application across the NSP and BOP

More functional, maintainable and human usable designs in both plant retrofits and new plants

Shift from establishing a position to making the outcomes stronger

Methods

Fundamentally Design Validation is a test not an exploration, although there are aspects of exploration in understanding and interpreting test results

Every CR design change no matter how small is subject to engineering V&V – processes scaled to the nature of the change.

All design disciplines are subject to the above

Both HF design and safety critical SW design share a need to make processes structured and visible

Methods…cont'd

Create an operational environment to place SMEs into that will facilitate their response to operational conditions

- Large and small changes to CR designs
- Range from table tops with defined rating scales to dynamic conditions where responses occur in real time or close to real time

Use of SMEs from a range of areas

- Designers, Operators and trainers for test scenario development
- Operators, Shift supervisors, maintainers, field staff for testing

Forms of Fidelity

Interface physical

Interface functional

Environment

Data completeness

Data content

Data dynamics

Remote actions

Representation of support tasks such as shift logs

Scenario Selection

Factors to consider

- Critical tasks related to safety
- Frequently performed tasks
- Tasks with known or suspected performance issues
- Range of task behaviours common to many operations
- Bounding tasks
- Ability to reproduce task conditions in test-bed
- Amount of preparation for test required
- Duration of test and availability of SMEs

Assignment of SMEs

Common considerations

- Not random with small samples
- Counterbalance scenarios
- Balance experience levels
- Balance outliers (or eliminate)
- Try to use SMEs that were not involved or close to the design

Methods…Measures

Rating scales covering aspects of human behaviour based on a standard human model covering aspects from perception to response verification

- Acceptance criteria based industry "acceptable" performance

Objectives measures of various types with set acceptance criteria

- Time to complete task or task segment
- Completion of task (no time)
- Human errors
- Knowledge (e.g. Situational Awareness)
- OWL
- Detection and response to specified events (e.g alarm response)
- Time to achieve specified plant state or condition

Debrief sessions with SMEs augment measured data (no acceptance criteria)

Data Analysis

Statistical where possible

- Descriptive
  - Do this as much as possible to understand data
  - Use judgment to make inferences and draw conclusions
  - Look for convergence of measures (convergent validity)
- Inferential
  - Parametric if possible
  - Look for convergence of measures (convergent validity)
- With variances, pay attention to pattern over intervals or conditions (peculiar pattern can indicate methodological problem)

Preparation of SMEs

Some considerations

- SMEs require some level of training on design
  - Can range from a briefing to several days of high fidelity training
- Purpose is to reduce bias in data due to learning effects
- SMEs do not like to be surprised or embarrassed
- Make sure they realize we are using them as measurement instruments and they are not being evaluated. (Location of simulator based trials a special consideration)
- Ensure confidentiality of data (practical limits on this)

Methods---Relationship to Related Activities

Alignment with I&C design validation activities/plans (includes SW control/display)
Alignment with operator training programs
Alignment with plant change control that includes training of staff on changes
Application of selected aspects of training practices for development of measures and acceptance criteria

- Consider the paradigm
  - Training and licensing uses the design as the measurement tool
  - Design uses the operator as the measurement tool
  - On the surface, both paradigms look the same – operators interacting with high fidelity representations of the plant
- To this end
  - Time to complete
  - Human errors
  - Task completion
  - Knowledge
  - Ability to apply knowledge
  - Being monitored closely – over the shoulder, operator verbalization, halting scenarios without notice
  - All of the above are used continuously over the life of the plant by authorized staff training programs

Multi Unit MCR

Single Unit MCR

SNC·LAVALIN Candu

18

Candu MCR Mockup Facility – Past Generation

SNC·LAVALIN Candu

19

Reconfiguration of MCR Engineering Mockup for Validation of Changes

SNC·LAVALIN Candu

Candu MCR Mockup Facility – Current Generation

SNC·LAVALIN Candu

Challenges

Establishing scope, mechanism and schedule for FICRV

› Project discussion on how much validation is needed, particularly for retrofits

› For new plant designs, how much fidelity in procedures, training of SMEs and dynamic modelling of plant response

› Validation must be integrated with the engineering schedules so as to allow for some change to design

› Balancing generic/standard engineering from site specific which can be many years away

Challenges – Acceptance Criteria

One of the more challenging aspects

Technical basis required

Broad technical acceptance required (Face validity)

Examples are

› Reasonable time

› Error type committed

› Correct answer

› Task completed

› Specific time to target state

› Judgment on given dimension above acceptable level (e.g. usability dimensions)

Challenge…cont'd

Addressing Secondary Control Room and remote shutdown points FICRV

Tackling maintenance related issues

› Alignment with maintainability program/experts

Access to operations personnel for design input

› Developing/finding operations personnel that can work within a design environment (in this context for resolving performance issues identified during validation)

› Alignment with project groups responsible for operability of designs – for changes these are utility groups (e.g. COMs process)

Special Considerations

Style of licensing in different countries

Ensure visibility of design processes IAW licensing expectations

Independence from design

› Not more than practice for special safety systems

CR mockups are the most salient and visible aspect of plant design – include considerations for

› Tours

› Space for observers

› Canned demonstrations

Requirements…

Considerable evolution over past 20+ years

- Requirements are hard to write
  - Designers to think in terms of the solution too much
- Requirements coded for tracking through detailed design
- Cross references to related sections of DR
- Rationale
- Dedicated HFE section
- Reference to project HFE Program Plan
- Ownership by Control Centre design group and/or Allied discipline such as I&C.
- Clear path and relationship to other key DRs such as
  - Annunciation
  - Plant Display system
  - Special safety systems
  - SCA
- Changes tightly controlled by project Change Control Function

Requirements…examples

Functional Safety Requirements

- Design shall be such…which are initiated by automatic control logic in response to an accident can also be initiated manually.
- CRs shall be designed…such that presentation of information shall provide personnel with an adequate picture of the status and performance of the plant and support necessary operator actions…
- MCR, SCA & TSC shall contain SPDS that presents sufficient information on SCPs for the diagnosis and mitigation of DBAs and severe accidents…

Requirements…

Functional Safety Requirements

› Displays shall be provided for indicating bypassed or deliberately inoperable conditions of the plant auxiliaries…

› …MCR design shall provide the operator with accurate, complete, operationally relevant and timely information regarding the functional status of plant equipment & systems…

› …the MCR shall be designed to bring the plant back to a safe state after the onset of accident conditions

Requirements…

Performance Requirements

Following clear and unambiguous indication that operator action is necessary on a safety or safety support system, there shall be 30 min available before that action is required (inside MCR) and 60 min outside …

› ..the design of the MCR, SCA and TSC shall provide adequate situational awareness required to monitor and control plant safety (also a similar requirement to not have excessive OWL)

› …consideration shall be given in the MCR design to minimizing the occurrence of any undesired power reduction or trip caused by operators' erroneous decision making and actions…

Requirements

Performance Requirements

› …upon EOF activation…it shall be staffed and fully operational within 1 hr…TSC shall perform EOF functions in first hr

› Operation and Maintenance

› …maintenance program shall be developed and integrated into CR design (new plant designs warrant this level of effort)

› …the design shall permit regular training and practice in the use of the SCA without affecting plant availability

Requirements…

Layout

› CR shall have sufficient space to allow staff to perform necessary actions while minimizing the need for operator movement in abnormal conditions….the plant Operating Basis identifies normal and minimum staff complement including space requirements)

  › Operating Basis defines range of staff configurations because there is some variance between utilities in operational style, beyond this range the utility would need to define the requirements as part of site specific engineering

› …walking time to SCA < 2min

Requirements…

Human Factors

› Project HFEPP invoked to give it proper influence over CR design – on the scale of effort across the project, CRs are highest level

› …HF assessment shall be completed during detailed design…non conformances shall be corrected to the greatest extent practical/possible

› …an information system shall be provided to inform operators of the plant status on variable important to safety and availability, which allows operators to obtain a complete understanding of the plant state at all times [IEC 964 reference]

› …the MCR design shall provide an optimal assignment of functions which achieves maximum utilization of operator and system capabilities… [IEC 964 reference]

Next Steps

If I had resources to spend on advancing confidence in FCRISV

› Review requirements for opportunities to make them

- › More specific
- › More quantifiable
- › Push for stronger alignment with related design validation activities in I&C and human performance aspects of training and staff licensing groups
  - › Ultimately these needs cannot be opposing
  - › Greater credibility and acceptance by both design and operating organizations

Thank You

Topical Session 1

Defining the Scope and Objectives of a Control Room Validation

Human Factors Validation Of Nuclear Power Plant Control Rooms & Modifications

February 19 2015

Challenge Session 1

What are the critical considerations in defining validation scope & objectives; how do these impact achieving reasonable confidence

R. E. Hall and Julie Reed

Challenge Session Objectives

{Scope & Objectives}

- Critically examine preliminary and final integrated validation activities to understand strengths, limitations and relationships between technical and practical considerations
- Discuss, through case histories, variation in international practices and lessons learned
- Identify recommended practices, potential solutions and currently available technical bases for addressing existing limitations
- Identify priority areas for future research

Setting The Goals For Validation Testing

- Must focus and limit the scope
- Must be practical
- Must account for the availability of needed tools
- Must have a realistic estimate of resource commitments
- Must identify client & regulatory expectations
- Must understand the difference and relationship between preliminary and final (ISV) validation
- Must support achieving reasonable confidence
- Should include future marketplace demands

3

Too many times the scope

- is limited to either testing a design element or assessing final acceptance,
- is limited to the control room and operations,
- does not consider input from external analysis; such as risk (PRA) and reliability (HRA) studies,
- does not include past testing results; such as vendor acceptance tests and predecessor & reference plant testing programs,
- does not consider required tool availability,

resulting in limited usefulness & added impact on cost and schedule

4

Too many times objectives:

- do not consider the overall goals of achieving reasonable confidence,
- do not integrate preliminary and final validation,
- do not account for limitations of technology,
- do not consider required tool availability,
- do not consider costs and schedule restraints
- do not consider required resources (human factors, operations)
- do not realistically identify the application to the world market place,

resulting in limited usefulness & added impact on cost and schedule!

5

One Challenge

Julie and I have selected one challenge that must be addressed in defining validation scope and objectives:

Considerations for successfully introducing a design across country boarders

{application to the world market}

6

Application to the world market

{moving into the USA}

- Many new plant designs and control room platforms have their country of design origin outside of the US boarders
- This introduces a set of similar issues across designs when setting validation scope and objectives

•7

Application to the world market

{moving outside the USA}

- Taking credit for USA verifications in another country
  - Identifying the differences
  - Determining applicability
    - Design
    - Procedures
    - Training
    - Culture and personnel
- This introduces a set of similar issues across designs when setting validation scope and objectives

•8

Application to the world market

{First-of-a-kind – USA, another country, or both}

- No predecessor or reference plant
- Different regulatory expectations of scope, degree of compliance
- The 'standard plant' misconception
- Development of alternative approaches
- Unique Challenges:
  - Costs
  - Schedule (from Engineering Completion to Commissioning)
  - Simulator fidelity and availability
  - PRA updates and scope
  - Ongoing design changes
  - Trained and experienced operators
  - Validated operating procedures
  - An integrated system – design, procedures, training, simulator

9

Challenge!

How can validation manage

- differences in country conduct of operations and crew models
- regulatory requirements and expectations, interpretation & application of NUREG-0711 & NUREG-700
- differences in stakeholder requirements and expectations
- accounting for all past validation efforts

?

10

Primary Topics For Consideration when defining scope and objectives (1 of 6)

- How can validation be optimized to account for cross country migration, designed in one country and built in another? Is there a way to have a world accepted validation process? What should be considered when setting the validation objectives for a design that may enter the world market place?
- Should the objectives of validation be limited to evaluation of performance of the integrated control room/facility, or should it include the evaluation of individual significant elements of the control room/facility (i.e., operating procedures, computer based procedure system, alarm prioritization rules, navigation rules)? If both, how should the objectives for each be formulated?
- Should ISV be mainly an "acceptance" test to determine if minimum acceptable performance has been achieved? Should ISV include identification of design changes to achieve it? Should it include the identification of areas for improvement beyond minimum acceptance?

Topics For Consideration (2 of 6)

- Should validation be limited to assuring only event mitigation or should it also include event prevention?
- How should probabilistic safety assessment and human reliability analysis be used to guide or support validation?
- Should ISV include validating the results of issues identified during design and preliminary/ design validation tests? How might including this objective affect achieving reasonable confidence? How can this be accomplished in an efficient, repeatable integrated process?

Topics For Consideration (3 of 6)

- Should the scope of validation include the main control room, technical support center, emergency operations facility and local control stations? How does the implementation of a digital control system impact the validation's scope?
- Should the scope of validation include test and maintenance activities even when they do not impact control room activities? How can this be achieved?
- What is the impact on the validation process when applied to multi-unit plants such as small modular reactors? How does one achieve reasonable confidence in one module or a plant consisting of a group of modules?

13

Topics For Consideration (4 of 6)

- What are the potential uses of the results of validation testing in design, ISV, pre-operational testing, plant operations and future modifications?
- How should validation be different for a first of a kind design versus a design based on a predecessor or reference design where the most significant change is the digital I&C platform and HSI?
- How can validation objectives help to match the testing program's needs with the availability of test tools (procedures, trained operators, simulators)? When the tools are not available in a timely manner, what alternatives can be applied to support validation?

14

Topics For Consideration (5 of 6)

- Is there a role for the pre-operational start up testing program in achieving reasonable confidence, in particular, in supporting or augmenting the ISV?
- Can vendor factory acceptance tests be meaningfully applied to the validation program? How?
- What difference must be considered in validation of a new plant as compared to modernizing an existing plant? What should be accounted for in a staged modernization program versus performing the modernization all at one time? How do you set the objective when dealing with a hybrid or interim design? How should impacts on safety be considered in validation when only modifying non-safety systems?

15

Topics For Consideration (6 of 6)

- How should validation be applied when using a vendors standardized I&C/HSI platform in a plant modernization program?

16

Topical Session 2

Rationale for Selecting Measures and Acceptance Criteria

OECD NEA WGHOF Workshop on Integrated System Validation of NPP Control Rooms – Charlotte, NC, February 2015

Challenge Session 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?*

Methods, approaches, resources or rationales used for evaluation and validation of nuclear power plant control room designs: the case of a new build

Draft

Cecilia De la Garza, EDF R&D

Evaluation/validation is an important step of the Human Factors Engineering (HFE) Program in EDF. This presentation will focus on the final validation of a new control room design and will describe the characteristics of an ISV as developed in EDF for a new build.

The context: the control room of a new generation nuclear reactor, involving high automation, with computerized interface and, hybrid procedures (both paper and computerized). The future reactor presents a relevant technological evolution in comparison to the existing NPP in France. R&D has been involved over the last 13 years in the design process and in an iterative process of evaluation/validation.
To summarize, in terms of ISV process, 7 human factors campaigns have been carried out for this new build: three preliminary evaluations/validations from 2002 to 2008, and four final validations from 2009 to 2013, and a last one is planned for 2015 (cf. Fig 1).

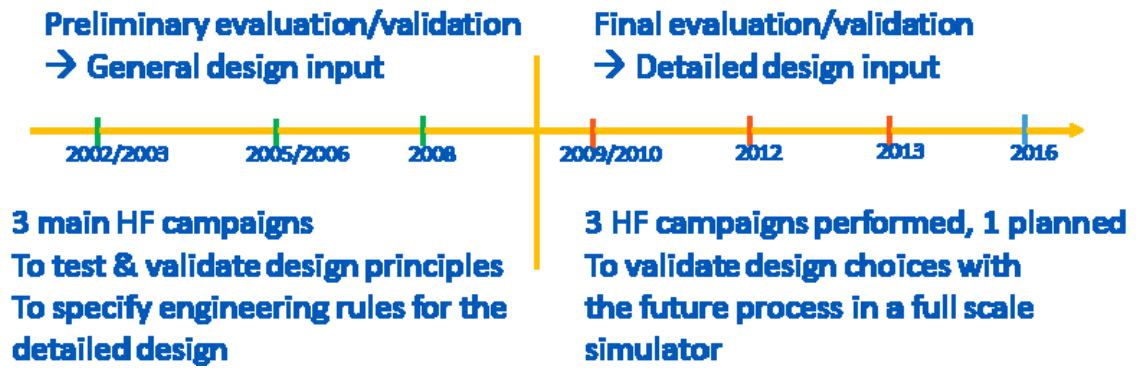


Figure 1 – Recap of preliminary and final evaluation/validation for a new build

Definitions: In the EDF approach, the concept of validation emphasizes its status as a "process" linked to design, and it continues until the commissioning and the first phases of operational feedback.

At EDF R&D the human factors evaluation/validation, could be consistent with ISV as defined in the 0711.

ISV means: to evaluate or to check that the design (interfaces, procedures, work organization, competence) supports the control room function and achieves its functional goals. "…integrated system design (i.e., hardware, software, and personnel elements) meets performance requirements…" in a dynamic mock-up or a full scale simulator.

The methods and results of the evaluations carried out can lead to a sufficient level of confidence to reach the project's objectives (expected performance, operating conditions, acceptability, etc.). Even if neither the design of the control means, nor the design of the full scale simulator is finished. But, each stage of the evaluation process makes it possible to validate choices of design with a sufficient degree of confidence. In theory what is validated at stage N will not be called into question at the time of the campaigns n+1 or the following ones.

Aims: To evaluate all the control means before the first "fuel loading": to evaluate their interactions in different simulations of situations covering normal & emergency operations.

→To determine the safety of the design and prove safe operating before first "fuel loading"

Approaches: a multidisciplinary approach combining Ergonomics and Human Reliability (HR) in final validation in an evaluation/validation process defined in a HFE program

Rationale or model underlying the ISV

From an ergonomic point of view:

→Full scale simulations based on a kind of "task model", a task classification established from an analysis of the existing situations: normal operation in the control room and emergency operation during training on a full scale simulator (work analysis cf. Eg. Vicente, 1999; Tricot et al,1998; Sebillote et al , 1994).
It concerns a description of the main tasks carried out during in situ observations covering a representative sample of work situations of the reactor operating states in normal power: normal and abnormal operations as surveillance – dispatching decreasing/increasing load – alarms monitoring -…), outage, restarting, maintenance tasks, periodic performance tests. And also a task classification in relation with the analysis of emergency operations (diagnosis – stabilisation of the installation – team reconfiguration – safety actions …).
Even if this is not an exhaustive task analysis, this tasks classification is considered representative of the main tasks and it provides input for design and the evaluation/validation process.
This description is not equivalent to the task analysis as defined by the NUREG 01711. It is a macro task analysis. But as it is explained in the NUREG, this task classification will become progessively more detailed over the design cycle.
In particular, in the case of a new build even if many tasks could be anticipated, the introduction of an innovation should imply a new task, different from the tasks carried out in the existing system. So progressively, and in relation with innovation (for instance automation) and the HF evaluation process, new tasks will be identified leading to new inputs for design and/or implying new skills, or having an impact on staffing…

→ "Ecological simulations" are carried out reflecting the characteristics and the constraints of a future work situation
→ Cognitive models are used to support cognitive analysis in decision making, problem solving, (Rasmussen, Amalberti) and to support some methodological choices or protocols for validation. E.g. From HRA: Resilience model developed by Le Bot (see below).

How to specify the scope and objectives of a validation campaign

The control means state & the full scale simulator state could determine the number of observations, the number of scenarios: what is possible to simulate and with what control means? So the scope and the objectives vary from one human factors campaign to another.
But, some relevant points guide the scope and objectives of a validation:

- Technological innovation of the future reactor: the human factors experts established a table listing the technological innovation associated to normal and emergency operation in order to guide validation;
- The state of progress of the design: it could be necessary to validate a specific procedure, the computerized interface in relation with team organization for instance;
- Using national feed-back from the existing NPP, lessons from the past: emergency operation simulation based on a real accident or incident (national and international).

→ To be sure that the team is able to face these kinds of situations is a general scope underlying validation.

System requirements and human Performance requirements

In this part, the approaches of both ergonomics and Human Reliability analysis (HRA) are going to be briefly described in order to highlight how they contribute to the evaluation/validation of the system requirements and the human performance requirements associated to performance criteria previously defined (De la Garza et al., 2013, 2014).

Both are rather "bottom up" approaches than top down, even if top down approaches are useful for some expert judgment, and definition of hypothesis for the protocols preparation.

Measures/ observations

Both approaches share an overall empirical methodology characterized by simulation sessions lasting 3 to 4 hours, observed in situ, followed by a group debriefing with the operating team (2 to 3 hours), prepared in advance by the evaluation team.

Several data collection tools and techniques are combined:

- Paper-and-pencil type collection in situ using a general grid to monitor the activity in progress.
- Audio and video collection from all the workstations.
- The simulator logbooks are kept as they enable analysis after the event, for example to confirm that an action was carried out and the exact time (starting cooling, opening a valve), or to have information on the state of the installation at any given moment (state of steam generators, containment pressure, closed valves).
- Other types of collection using tools according to the needs of the evaluation (e.g. adapting

the Instantaneous Self Assessment tool for analysis of perceived workload).

However, only the general grid for paper-and-pencil type data-collection is also used by the human reliability experts. It is used to record a detailed chronology, decisions made, operating actions, problem-solving, etc. Only ergonomists carry out fine analysis after the event of the cognitive and teamwork activity based on audio recordings, followed for example by activity timelines.

To complete observations and to consolidate understanding of the scenario, decision making and actions of the team, debriefing is conducted at the end of simulation.

Ergonomics: an evaluation/validation of the sociotechnical system for the control means design

The ergonomics evaluation aims to analyze interactions between the operating team members and the control means (interfaces, imaging, procedures, and organization). The group point of view is studied in relation to communications, problem-solving and decision-making by the team. This leads to a diagnosis of the current situation which makes it possible to understand the consequences of the actions performed, the difficulties encountered and decisions made by the team, as they interact with the control means. For ergonomics, envisaging types of situations makes it possible to identify invariants, in terms of difficulties and risks as well as reliable individual and group operating methods or those that weaken the sociotechnical system studied. It is then possible to make recommendations aimed at improving man-machine interactions and team operation. Prognosis in ergonomics starts from a diagnosis of the simulation situation and analysis of the appropriateness of the changes proposed in response to the recommendations to solve these problems and respond to the risks identified.

Ergonomists will reconstitute the detailed work activity of each party involved, and especially the cognitive activity relating to decision-making, work load, problem-solving, diagnostics and monitoring (Mérand, et al., 2013). The analysis is done based on the observations and video recordings that make it possible to supplement the chronologies. It may be necessary to listen to a debriefing again to confirm a point. The aim of this analysis is to understand individual operating methods, and how the team works as a whole interacting with the control means and the difficulties and risks identified in the situations, before then being able to give recommendations on how to make each element of the sociotechnical system more reliable.

Human reliability: an evaluation/validation of the sociotechnical system with regard to risks

Human reliability evaluation involves analyzing how the operating system as a whole (team, procedures and interfaces) is capable of managing critical situations, from the point of view of safety and the reliability of the installation (like the MERMOS HRA method (Le Bot, 2007). For this, at EDF, human reliability is inspired on the one hand by functional and reliability engineering approaches (like Probabilistic Safety Assessments) and on the other by contributions from the humanities, in particular distributed cognition (Hutchins, 1995). Thus, the main actions required for safety in the situation considered are defined, and the reliability functions of the operating system are identified: action, inspection (conformity of performance with the actions decided by the

team), review (adapting the actions decided by the team to the situation in progress), and reconfiguration (Le Bot, 2010). Human reliability experts will focus on the results for safety and then go back to the "organizational and human factor" elements that influenced them. In other words, the evaluation is done first by examining the safety results, then the characteristics of the teamwork and also the entire operating system, all these factors are involved and have significant consequences for work safety. This is what is called evaluation of the "proved performance" during the simulation studied. To complete this evaluation which is made starting from the actual consequences to the process, human reliability also considers the "potential performance" of the operating system: that is to say that the important characteristics of the operating system observed during a simulation are transposed into other situations where their impact on safety might be significant, even if during the simulation that characteristic did not have an impact on safety (Pesme et al., 2013). With this in mind, they may be led to consider not only the recurrent risky team behavior, but also rarer behavior if they prove to have a strong impact on safety. Starting from the result makes it possible to bring out the characteristics relating to the most relevant "organizational and human factors" for safety, and thus to propose appropriate and priority recommendations concerning the overall safety of the installation.

To conclude, ergonomists focus their observations on individuals (each operator station, each profession) and team activities, whereas human reliability experts focus on the actions carried out in the process and on key individual and group facts in relation to the operating actions.

Meanwhile, for human reliability, it is a matter of observing all the operating team: the operators, as well as the operations manager and the safety engineer, as it is the entire operating system that is designed to manage situations. What is studied in more detail by human reliability experts is the state of the installation, so that they can understand the situations encountered by operators and the consequences of their operating actions.

In conclusion

EDF approaches for ISV seems to be adapted to an industrial project even if different points have to be investigated.

The ISV as carried out in EDF is an iterative process and is characterized by a dynamic environment supported by simulation taking into account the interface between control means and operation teams. The representativeness of these simulations evolved as the design progressed but the different evaluation steps allow consecutive validations with a good confidence level.

Preliminary validations were supported only by ergonomics, while final validations by both ergonomics and HRA. So for the next steps we will investigate how this double approach could be developed for preliminary validation. And we will continue to develop this kind of integrated evaluation/validation for other kinds of projects.

Bibliography

De la Garza C., Pesme H., Le Bot P. Labarthe J.-P. (2013). Interest of combining two human factor approaches for the evaluation of a future teamwork. In *Safety, Reliability and Risk Analysis:*

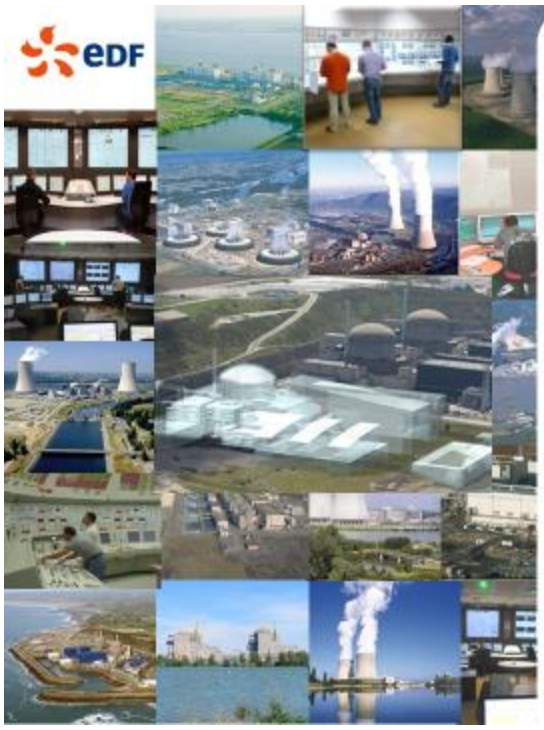
Beyond the Horizon, Steenbergen et al. (Eds.), pp. 523-529. ESREL Conference, 29 Sept – 2 Oct., Amsterdam.

De la Garza C., Pesme H., Le Bot P. Labarthe J-P. (2014). Comparison of two human factors approaches in an integrated evaluation/validation of a future emergency operating organization. ODAM-NES, 18-20 Août, Copenhague.

Le Bot, P., Pesme, H., & Meyer, P. 2007. The CICA concept redefined, for use in the MERMOS method to provide a new organizational reliability model, IEEE/HPRCT, Monterey CA.

Pesme, H., Le Bot, P. & Ruiz, F. 2013. Contribution of Human Reliability to the evaluation of a new Emergency Operating System, ESREL 2013.

Mérand J., Darses F., De la Garza C. (2013). The role of argumentation in the resolution of ambiguous situations: an exploratory study in the field of supervision in the nuclear industry. The 11th International conference on Naturalistic Decision Making. 22 -24 Mai, Marseille.



Workshop on Human Factors Validation of Nuclear Power Plant Control Room Designs

Challenge Session 2

Cecilia DE LA GARZA
Risk Management Departement
Human Factors Group

EDF

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?

Methods and approaches used for evaluation and validation of nuclear power plant control room designs: the case of a new design in EDF

EDF
CHANGER L'ÉNERGIE ENSEMBLE

Context: a project design of a control room for a new reactor (French EPR)

A new generation nuclear build implying,

- High level of automation
- Computerized interface
- Hybrid procedures (paper & computerized)
- New team organization

ISV process for a new build

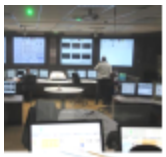
An "iterative process" linked to design, which continues until the commissioning and the first phases of operational feedback

Carried out starting from several evaluations from:

- Control means
- Work organization (team organization)
- Skills and competencies

Aim: to confirm that design supports the control room function and achieves its functional goals

ISV is supported by dynamic an ecological simulation taking into account the interface between control means/operation crew/organization



Recap of preliminary and final evaluation/validation for a new build

Preliminary evaluation/validation
→ General design input

Final evaluation/validation
→ Detailed design input

2002/2003 — 2005/2006 — 2008 — 2009/2010 — 2012 — 2013 — 2016

3 main HF evaluations
To test & validate design principles in dynamic mock-up
To specify engineering rules for the detailed design
To reduce project risk

3 HF evaluations carried out, 1 planned
To validate design choices with the future process in a full scale simulator
To increase simulations representativeness by allowing for more scenarios and for more crews across multiple and complementary evaluations

Characteristics of Simulation for preliminary and final validation

Simulation of different operating situations

Based on a "tasks classification" of the main operating situations

- Established from the analysis of the existing normal operating situations and national & international feed-back in terms of emergency operation
- Covering a representative sample of normal, abnormal and emergency operation
  - E.g. in normal operation: surveillance tasks, alarms monitoring, periodic performance tests, outage tasks…
  - E.g. in emergency operation safety actions in relation with: steam generator tube rupture, loss of the heat sink, loss of electrical power and the critical actions associated

Simulation is a diagnosis and forecasting tool for future operating activity

Supported by scenarios (cf. Labarthe's presentation)

Final validation: simulation in a full scale simulator

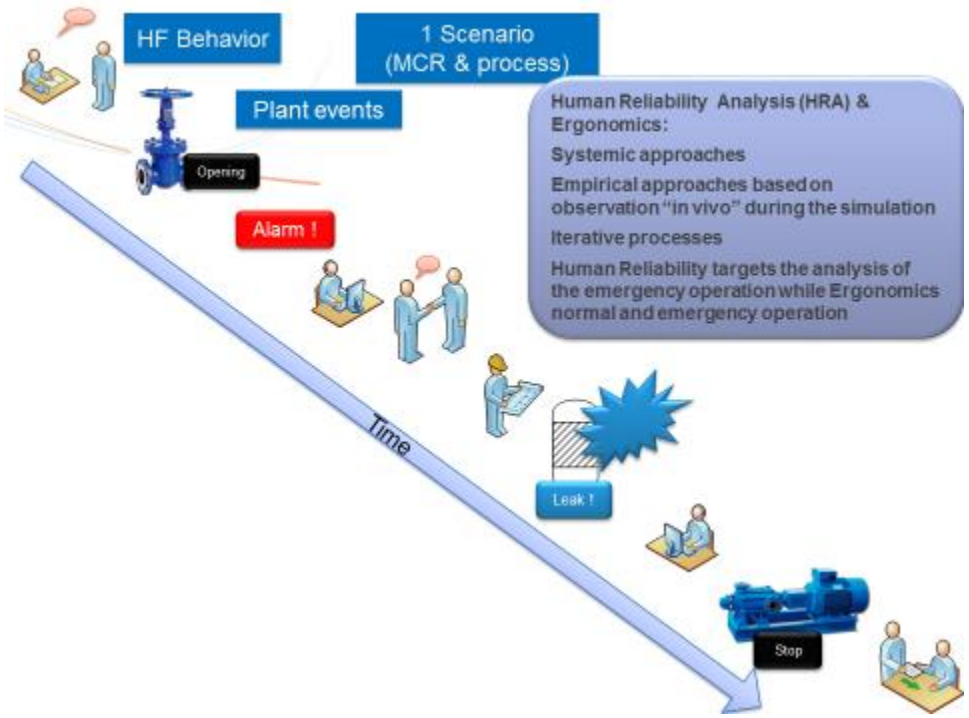
Full scale simulator with EPR process

A multidisciplinary team evaluation: psychology/ergonomics and human reliability analysis

Evaluation criteria

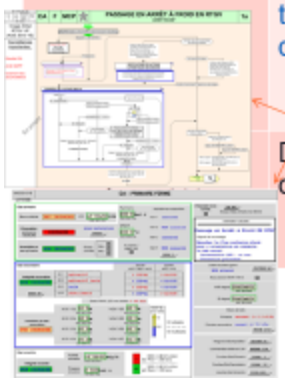
- The operation performance: actions
- Organizational defense Lines: supervision & verification
- Crew performance in interaction with the control means:
  - usability, consistency, usefulness
  - problem solving, decision making…
  - coordination, communication…

EDF

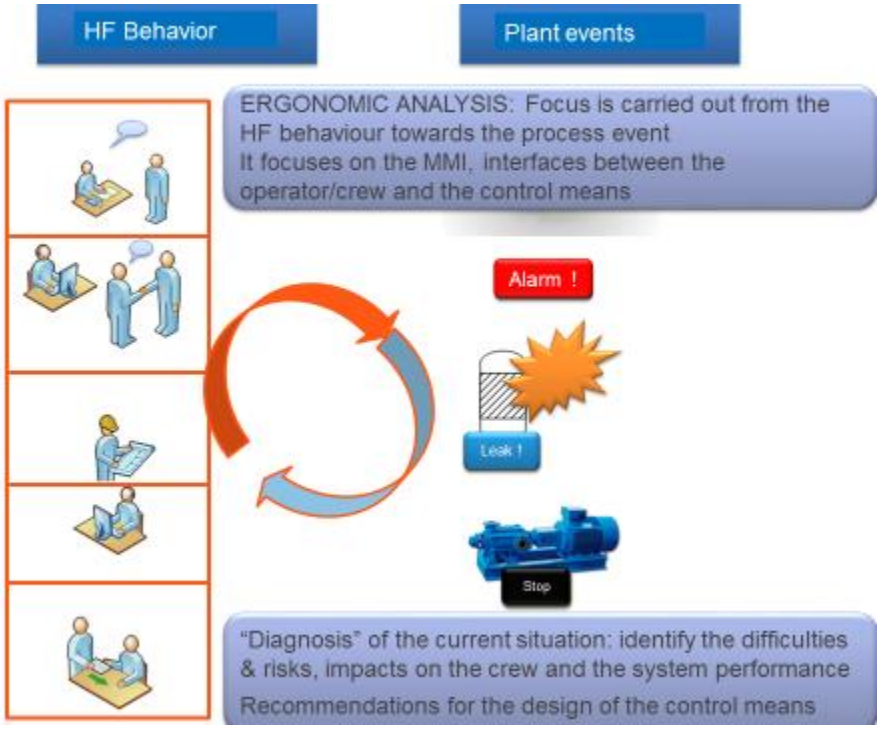


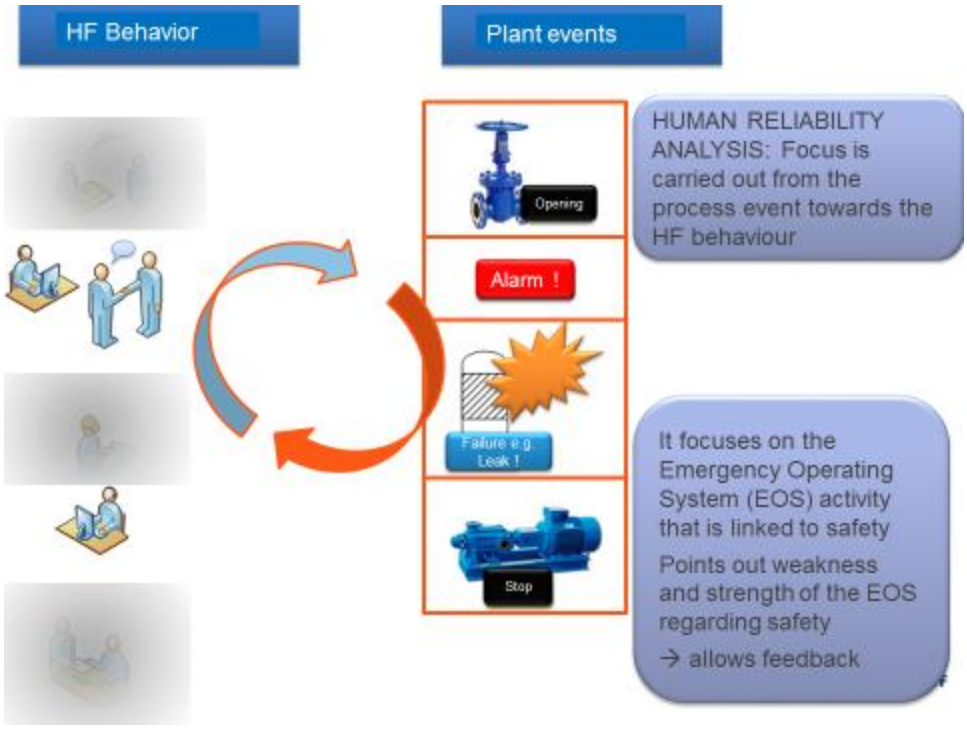
Common and specific methodological points for simulations preparation

Preparation	Ergonomics	Human reliability
Realistic scenarios	Ecological validity: Choice of accident transients and planned failures should cover a set of realistic situations	
	Specific situations to be tested, previously described	Choices of a spectrum of situations as in PSAs. Priority given to the operational dynamics and realism
	Defined according to the objectives of evaluation	Related to the main actions required for safety and to the main safety functions of the EOS



EDF





Common and specific methodological points for simulations data collection

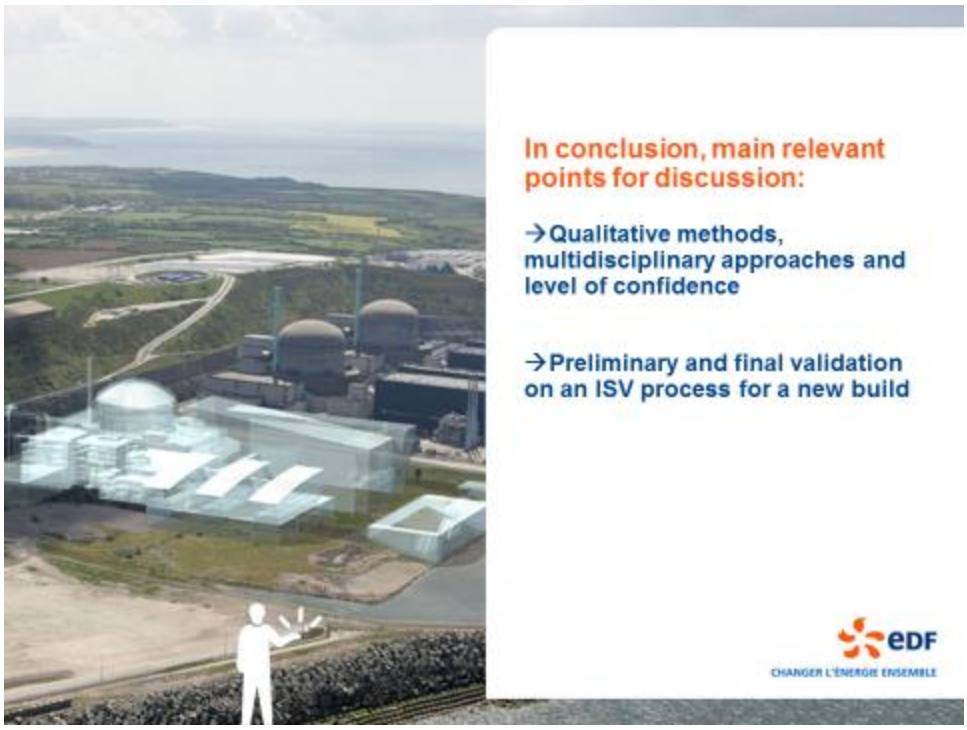
Observation	Ergonomics	Human Reliability
Aim	Articulation of individual and teamwork points of view	Gives priority to teamwork behavior
Observers	General view + According to the team members in the control room	
Notes taking	Chronology, decision making, operation actions, communication, difficulties…	
Video & audio recording	YES to carry out a posteriori detailed analysis	Not essential, can be useful for a particular point
Other data collection	Instrumented tests, such as Instantaneous Self Assessment (workload)	Dynamic process data : curves related to the scenario
	The simulator logbooks to confirm that an action has effectively been taken, exact time, state of the reactor…	
Debriefing with the team	YES	
Interviews	Could be considered to explore a topic	

Common and specific methodological points for simulations analysis

Analyses	Ergonomics	Human Reliability
Objectives and topics analysis differ	Detailed cognitive analysis Analysis of teamwork behavior Understanding of the decision making, work practices & processes	Detailed analysis of safety results and analysis of the human and organizational factors that influenced them (**proven performance**)
The same topic may not have the same meaning	e.g.1 problem with the imagery or with a procedure, human error e.g. 2 workload may have an impact on supervision	Difficulties observed are pointed out only if there are potential consequences on safety, in other contexts (**potential performance**)
	→Analysis of the conditions leading to these situations in order to make recommendations	-Assessment of proven performance -Assessment of potential performance: EOS actions, supervision & control lines linked to safety potential impacts

Results interpretation and recommandation

conclusions	Ergonomics	Human Reliability
From diagnosis to forecast HED	Guided by the difficulties & the risks observed	Justified and prioritized from a safety point of view
Recommendations	Classified according to topics, and discussed with the design teams, analyzed in terms of feasibility, possible solutions…	Recommendation assessed regarding the spectrum of situations (as in PSA)



Thank you for your attention

Challenge Session 2

Rationale for Selecting Measures and Acceptance Criteria

P. Ø. Braarud,
OECD Halden Reactor Project

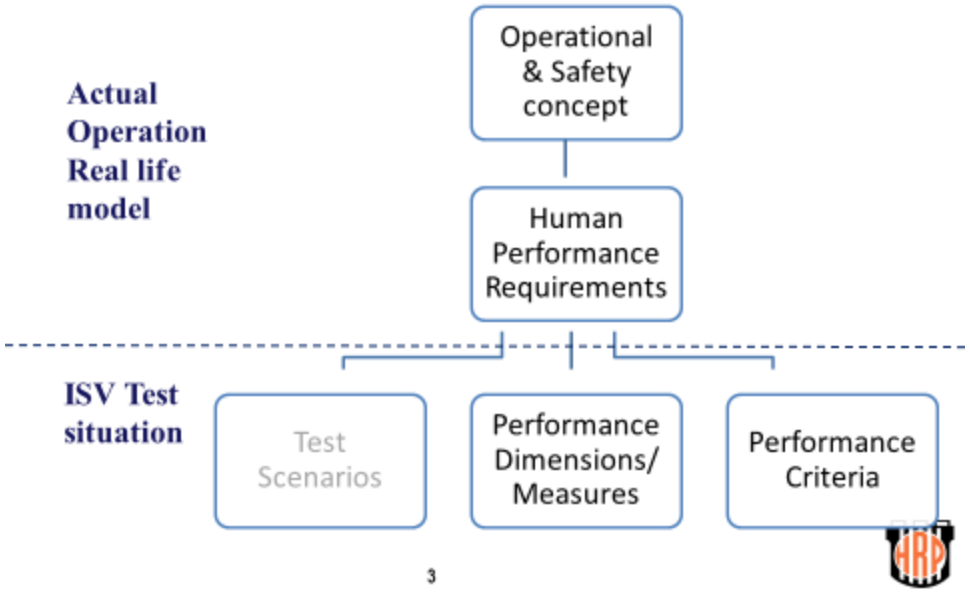
Examples of ISV Challenges

O'Hara (This workshop):

- What aspects of human-system performance should be measured
- What specific metrics should be used
  - What is the psychometric suitability of measures being used, such as construct validity and inter-rater reliability
- How should acceptance criteria be determined (especially for cognitive measures and for opinions and comments from observers and operators)
- How should pass/fail measures be identified
- The explicit linkage of performance measures to higher-level mission characteristics
- Adding teamwork-specific objectives

2

Concepts for discussion



3

Human Performance Requirements

- «Requirements» - Technical?
  - E.g., «Reactor Vessel pressure, Core temperature, Readiness of safety trains and safety functions, radioactive outlet to environment….
- *Human Performance* Requirements ?
  - (a) process control actions and plant monitoring
  - (b) requirements of how work should be performed
  - (c) the performance support that should be provided from the control room means/tools.
- "Requirement challenge" in Human Factors, e.g., Harwood, K. (1993)
  - Difficult to specify in sufficient detail for testing
  - Requirements influences each other (trade-offs), are situation specific, often require deep understanding of the actual work.

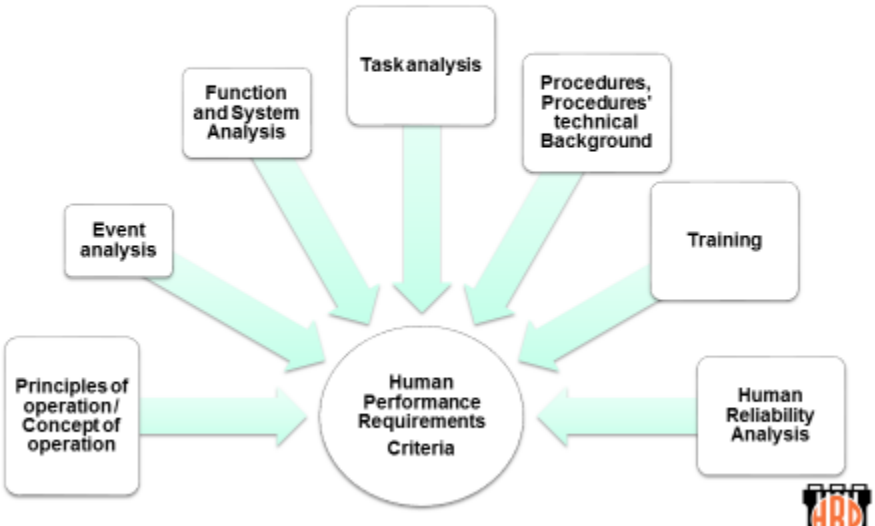
4

ISV Test Performance Criteria

- Criteria apply to the Test
  - IF criteria are satisfied we believe real life requirement can be fulfilled.
- ISV Test Criteria - performance issues identified if criteria not met
- Generally two types of criteria:
  - Reference system based (Benchmark, Norm)
  - Requirement based

5

Sources for clarification of Requirements and specification of Criteria



6

Overall framework

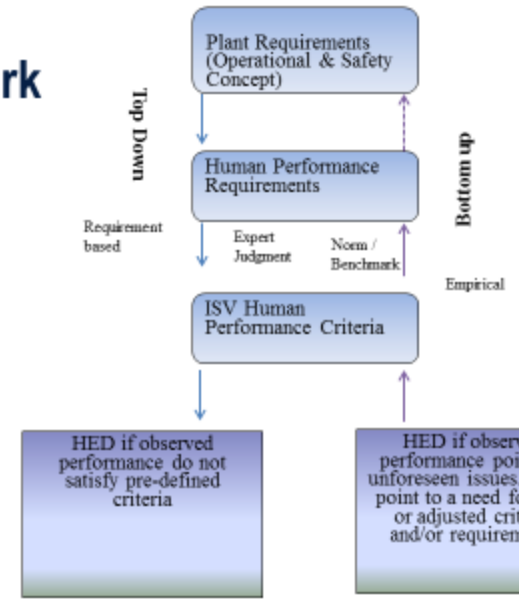


Figure from NPIC&HMIT 2015 paper: An Overall Framework for the Definition of Requirements, Criteria and Human Engineering Discrepancies for Control Room Validation

Human Performance Measures

- Broad literature available
- Characteristics
  - Reliability, Objectivity, Non-intrusive,……
- **Main Question: Validity**
  - The relationship with Human performance of the real world operation concept and safety concept
  - Validity Important but often difficult to evaluate
  - In some cases a tradeoff between reliability, objectivity and validity
  - Assumption: Each plant unique case, the plant specific operational and safety concept are important
  - Literature provides Frameworks that need to be finally defined / adjusted, «populated» with plant specific content – not at all a trivial task

8

Topics

- How case / context specific is performance based evaluation of a control room
  - What standards and guides can apply regarding measures and criteria? At what level, high level or detailed specifications?
  - Role of «Generic Measures»
- «Human Performance Requirements» as part of Plant's Operational & Safety concept ?
  - How to define and identify for ISV testing?
- Generalization
  - Solving given scenario (e.g., process goals of given scenario)
  - Potential or Capability beyond given test scenario

Topical Session 3

Construction of Validation Test Scenarios, the Test Design, and their Relationship to the Validation Framework

2/5/2015

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

Challenges in Defining Validation Scenarios: Searching for Edge Cases & Complicating Situational Factors

Emilie M. Roth
Roth Cognitive Engineering

Abstract

Human performance in real-world settings is fundamentally a function of individual and team cognitive and collaborative factors; situational complexities in the unfolding events; and the attributes of the available support 'artifacts' (e.g., displays, procedures, decision-aids). Design of validation scenarios necessarily requires consideration of the interplay across all three (the cognitive triad), and how particular confluences across the three elements of the cognitive triad may lead to performance vulnerabilities (Patter & Rousseau, 2010; Roth & Eggleston, 2010; Roth, Gualtieri, Elm & Potter, 2002).

Actual accidents often involve a confluence of complicating situational factors that challenge cognitive and collaborative performance (Feary & Roth, 2014). The recent Fukushima Daiichi nuclear power plant accident is a clear case in point (National Academy of Science, 2014), but there have been others in the Nuclear Industry, including the H. B. Robinson Fire in 2010. The challenge for design & validation of complex systems is to sample situations that exhibit similar 'confluences' of complicating factors – that represent *Edge Cases.* Edge cases are demanding decision-making situations that constitute the 'edge' of a human machine system that may create performance vulnerability (O'Hara presentation).

There is a need for better analysis and modeling tools to define edge cases to include in validations so as to achieve reasonable confidence that the human machine system will operate resiliently in the face of complicating situational factors that challenge cognitive and team processes. It is important to incorporate scenarios that go beyond routine 'textbook' cases, sampling realistically demanding conditions that challenge cognitive and collaborative processes. This is especially important when evaluating the ability of operators to perform critical human actions, for example critical actions called out in probabilistic risk assessments. Critical human actions are likely to be straightforward to accomplish in a 'textbook' scenario, but more challenging under conditions that stress monitoring, attention allocation, situation assessment, goal prioritization, communication and coordination and/or response planning processes.

Searching for edge cases will require identifying challenges at the intersection of characteristics of the people (e.g., knowledge, skills, biases), the situations, and the technology. Operational experience reviews, lessons learned from other industries, cognitive task analyses and generic

lists of 'complicating situational factors' can all feed into the definition of edge cases to include in a validation.

One promising direction for identification of edge cases is to leverage lists of generic 'complicating situational factors' to guide systematic search. Patterson, Roth & Woods (2010) present one such list of complicating situational factors that can be used as 'seeds' for identifying domain complexities to incorporate in test scenarios. These factors are loosely organized around core macrocogntive (cognitive and collaborative) functions: detection/monitoring; sense-making/situation assessment; planning/action formulation; information sharing/communication/coordination; and attention/workload management. This list represents the most recent embodiment of a working effort to capture characteristics that pose challenges to macrocognitive processes in a domain-independent fashion. It represents a starting point, rather than an end-point in characterizing elements of domain complexity.

Another important lessons 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, prioritizing, and decision-making, largely shifts from the control room to emergency response centers. In the case of the U. S. this is likely to be the technical support center (TSC). Little attention has been paid to the adequacy of the information and communication systems in the TSC for supporting the kind of high level situation awareness, planning, decision-making, and coordinating that needs to go on in the TSC and between the TSC, the control room and individuals in other response centers as well as in the field. More attention in the design of the validation, and especially in design of the validation scenarios is needed to the activities that occur beyond the control room, so as to achieve reasonable confidence in the ability of the broader emergency response organization to cope with and mitigate accidents. Finally, 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 organization to respond resiliently.

The requirements on design of scenarios called for here raises a number of pragmatic challenges in the design and execution of validation studies that will need to be tackled. These include overcoming limitations in existing simulators, and dealing with the practical limits in the time available to conduct validation. While these pragmatic constraints are real challenges, it is believed they are, and need to be, surmountable.

References:

Feary, M. and Roth, E. (2014). Some challenges in the design of human-automation interaction for safety critical systems. In Proceedings of the 5th International conference on Applied Human Factors and Ergonomics AHFE 2014, Krakow, Poland.

National Academy of Sciences (2014). Lessons learned from the Fukushima Nuclear Accident for Improving Safety of U. S. Nuclear Plants. The National Academy of Sciences Press

Patterson, E. S., Roth, E. M., & Woods, D. D. (2010). Facets of complexity in situated work. In E. S. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams.* Farnhaum, UK: Ashgate Publishing.

Potter, S. & Rousseau, R. (2010). Evaluating the resilience of a human-computer decision-making team: A methodology for decision-centered testing. In E. S. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams.* Farnhaum, UK: Ashgate Publishing.

Roth, E. M. & Eggleston, R. G. (2010) Forging new evaluation paradigms: Beyond statistical generalization. In E. S. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams.* Farnhaum, UK: Ashgate Publishing.

Roth, E. M., Gualtieri, J. W., Elm, W. C. & Potter, S. S. (2002). Scenario development for decision support system evaluation. In Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting (pp. 357-361). Santa Monica, CA: Human Factors and Ergonomics Society.

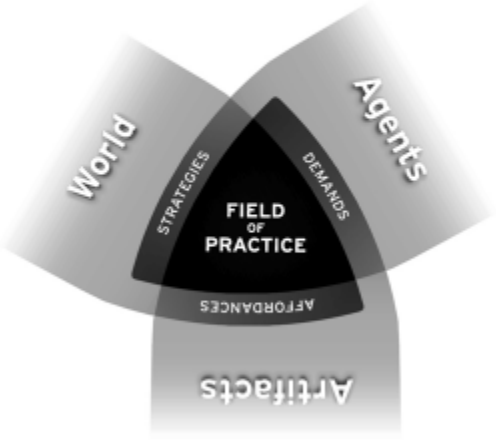
CHALLENGES IN DEFINING VALIDATION SCENARIOS:

SEARCHING FOR EDGE CASES & COMPLICATING SITUATIONAL FACTORS

Emilie Roth
Roth Cognitive Engineering

The Cognitive Systems 'TRIAD' is Fundamental to Analysis, Design & Evaluation of Complex Systems

- Human performance is determined by:
  - Cognitive & Collaborative Factors
  - Situational Complexities
  - Attributes of the available support 'artifacts':
    - Displays
    - Procedures
    - Decision-aids



Need to consider all three when defining validation scenarios

Challenges in Defining Validation Scenarios: Identifying & Incorporating 'Complicating Factors'

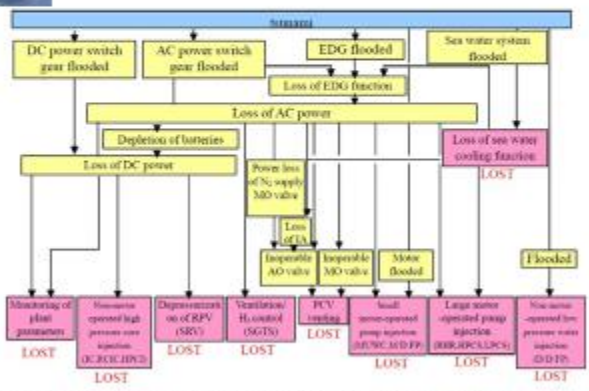
3

- Actual accidents often involve a confluence of complicating factors:
  - Fukushima Daiichi Nuclear Power Plant Accident
- Challenge for design & validation of complex systems – is to sample situations that exhibit similar 'confluences' (***Edge Cases***):
  - demanding decision-making situations that constitute the "edge" of a human machine system that may represent potential weaknesses (*O'Hara presentation*).

Fukushima Daiichi Accident

4

March 11, 2011 Tsunami height: approx 13 m



Loss of critical functions to prevent core damage and mitigate impacts

Created Complex Unanticipated Demands on Operators

5

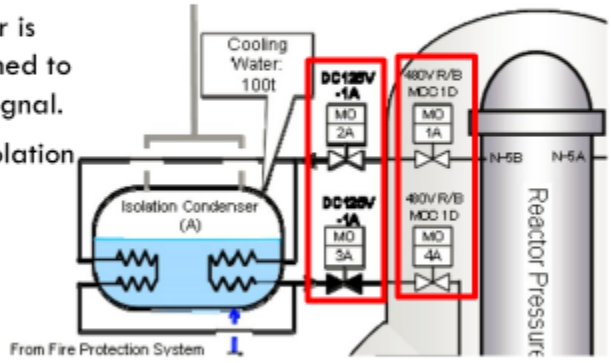
- Total loss of sensor indications
- Dark, high radiation
- Available Procedures inapplicable
- Preplanned mitigation strategies inapplicable
- Limited ability to communicate between control room, Emergency Response Center,and operators in the field
- Need to assess, prioritize, and cope with problems in multiple units simultaneously

Isolation Condenser (IC) 'Fail-Safe Logic' had Unintended Negative Consequences

6

IC Isolation Logic:

- When DC control power is lost the system is designed to generate an isolation signal.
- The motor-operated Isolation valves are:
  - DC-driven outside containment,
  - AC-driven inside containment



When the Tsunami hit, AC and DC power was lost, the valve position depended on relative timing of loss of control and valve drive power and could be open, closed, or in partial position.

Cognitive and Collaborative Challenges

- **Diagnostic Challenge**: Could not tell whether the IC system was functioning or not
- **Control Challenge**: Valves could not be re-opened until AC power was recovered
- **Response Planning Challenge**: Needed to identify and implement an alternative source of cooling using fire engines.
- **Communication and coordination challenges:** Among Emergency Response Center, Control Room, and operators in the field
- **Competing Demands:** Needed to prioritize demands for attention and resources across the multiple units.

While this may seem an extreme case, there have been other less severe events that similarly challenged cognitive and collaborative performance – e.g., H.B. Robinson Fire (3/28/2010)

Implications for Design of Validation Scenarios - I

8

- Need better analysis and modeling tools to define ***Edge Cases*** that exhibit representative 'confluences' of cognitive and collaborative complexity:
  - go beyond routine 'textbook' cases
  - sample realistically demanding situations that challenge cognitive and collaborative processes.
- Especially important when examining critical human actions identified from PRA
- Operational experience reviews, lessons from other industries, cognitive task analyses, and generic lists of complicating situational factors, can all feed this activity

Implications for Design of Validation Scenarios - II

9

- Need to expand beyond the Control Room – to the Technical Support Center and the larger emergency response organization
- Need to extend to Severe Accident scenarios
- Need to validate the ability of the new design to support performance under situations that we cannot fully anticipate
  - Characterizing cognitive and collaborative challenges at a 'generic' level is thus key.

Potential Paths Ahead

10

- Searching for 'edge cases' will likely require identifying challenges at the intersection of – the people, the situation, and the technology
- Defining generic '***complicating situational factors***' may provide a leverage to guide search.

Generic Complicating Situational Factors

11

- We developed a list that can be used as 'seeds' for identifying domain complexities to incorporate in test scenarios (Patterson, Roth & Woods, 2010)
- These factors are loosely organized around core Macrocognitive (cognitive and collaborative) functions:
  - detection/monitoring;
  - sense-making/situation assessment;
  - planning/action formulation;
  - information sharing/communication/coordination and
  - attention/workload management.
- Most recent embodiment of a working effort to capture characteristics that pose challenges to macrocognitive processes in a domain independent fashion.

Example Complicating Situational Factors

Macro-Cognitive Function	Complicating Situational Factor
Detecting / Noticing	Data overload Missing Information Misleading information
Diagnosing / 'Sense-Making'	Ambiguous cues 'Mismatch with expectations based on training / mental models
Planning/Deciding	Competing Goals Mismatch with procedures
Communication / Coordination	Multiple competing demands Mismatch with organizational structure

Summary of Challenges

13

- Need better analysis and modeling tools to identify ***Edge Cases*** that exhibit representative 'confluences' of cognitive and collaborative complexity:
  - go beyond routine 'textbook' cases
  - sample realistically demanding situations – representative of actual accidents and incidents experienced
  - expand beyond the Control Room – to the TSC and the larger emergency response organization
  - extend to Severe Accident scenarios
- Need to tackle pragmatic challenges: limitations in simulators; limited time available to conduct validation.
- Recent advances in "Macrocognition" methods and metrics – particularly generic lists of complicating situational factors -- may provide some ways ahead (e.g., Patterson & Miller, 2010).

References

14

- National Academy of Sciences (2014). Lessons learned from the Fukushima Nuclear Accident for Improving Safety of U. S. Nuclear Plants. The National Academy of Sciences Press
- Patterson, E. S., Roth, E. M., & Woods, D. D. (2010). Facets of complexity in situated work. In E. S. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams.* Farnhaum, UK: Ashgate Publishing.
- Potter, S. & Rousseau, R. (2010). Evaluating the resilience of a human-computer decision-making team: A methodology for decision-centered testing. In E. S. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams.* Farnhaum, UK: Ashgate Publishing.
- Roth, E. M. & Eggleston, R. G. (2010) Forging new evaluation paradigms: Beyond statistical generalization. In E. S. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams.* Farnhaum, UK: Ashgate Publishing.
- Roth, E. M., Gualtieri, J. W., Elm, W. C. & Potter, S. S. (2002). Scenario development for decision support system evaluation. In Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting (pp. 357-361). Santa Monica, CA: Human Factors and Ergonomics Society.

15

Backup Slides

Example of Complicating Factors (1)

16

Factor	Description	Cognitive/Collaborative Functions Impacted
Data overload (Needle in a haystack)	Problems that need to be detected and addressed are buried in a large amount of potentially relevant information.	• detection/monitoring • attention/workload management
Signal-noise relationship (False alarms)	Detecting a signal from background noise is difficult because the signal is close to the noise distribution. This is particularly challenging when there is a high false alarm rate or there are negative consequences for acting on false alarms. Information is prone to be discounted if the indicators are perceived to be unreliable or have a high false alarm rate.	• detection/monitoring • sense-making /situation assessment
Missing information	Information that is needed for an accurate assessment is missing (e.g., due to lack of sensors or failed sensors, lack of system update, lack of informants on the ground; or poor communication).	• detection/monitoring • sense-making /situation assessment
Ambiguous Cues	There are multiple, alternative, explanation for the pattern of symptoms observed.	• sense-making /situation assessment • planning/action formulation

Example of Complicating Factors (2)

Factor	Description	Cognitive/Collaborative Functions Impacted
Misleading Information/"Garden path" problems	Initial information suggests the wrong hypothesis (based on strong but incorrect evidence).	• sense-making /situation assessment; • planning/action formulation;
Uncertain Information	The accuracy of the information cannot be definitely ascertained	• sense-making /situation assessment; • planning/action formulation;
Complex or counterintuitive dynamics	A process changes over time in a complex difficult to predict manner making it difficult to develop an appropriate mental model and to anticipate/project the impact of changes over time	• sense-making /situation assessment; • planning/action formulation;
Multiple simultaneous 'influences'.	There are multiple independent 'influences' that are simultaneously present and in combination explain the observed evidence. There exists an alternative 'single influence' explanation for the evidence that appears more parsimonious but turns out to be false.	• sense-making /situation assessment; • planning/action formulation;

Example of Complicating Factors (3)

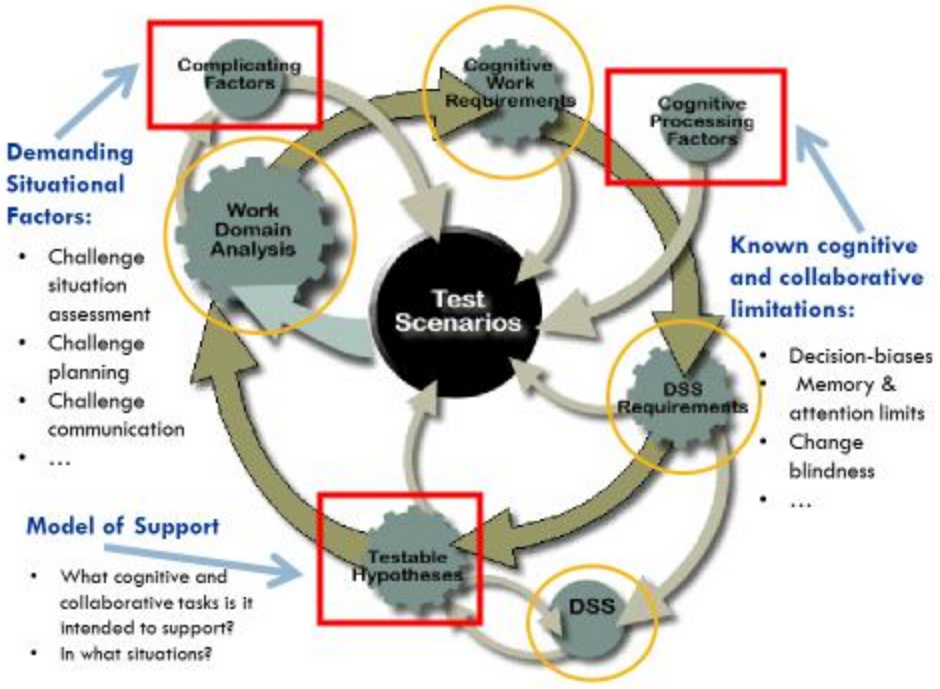
Factor	Description	Cognitive/Collaborative Functions Impacted
Distributed Information across individuals/organizations	Information distributed across participants and/or roles is required for recognition of a coherent pattern.	• sense-making /situation assessment • Coordinating
Distributed information over time (overturning updates)	Situations that require integrating information over time periods	• Sense-making/situation assessment
Hidden coupling (many-to-many mapping; effects at a distance; cascading effects)	A cascade of secondary effects can make it difficult to connect observed symptoms to the originating source and can complicate response planning.	• Sense-making/situation assessment • Planning/action formulation
Multiple simultaneous constraints	Problems where there are multiple constraints that need to be simultaneously satisfied.	• planning/action formulation
Goal Conflict situations	Situations where there are multiple conflicting goals that need to be balanced	• planning/action formulation
Unintended effects (managing side effects)	An action can have unintended secondary effects that need to be recognized and managed.	• planning/action formulation; • Coordinating

Example of Complicating Factors (4)

Factor	Description	Cognitive/Collaborative Functions Impacted
No available predefined plan or procedure (previously unanticipated situation)	An unfamiliar situation for which no predefined plan or procedure is available	• planning/action formulation
Incomplete or inadequate guidance or procedure.	The guidance provided is incomplete or suboptimal.	• planning/action formulation
Mismatch between predefined plans or procedures and the situation confronted (wrong plan)	The situation deviated from the assumptions underlying the plan/procedure and if followed verbatim will not achieve the desired effect.	• planning/action formulation
Sequential Interdependencies among multiple individuals (coordination demands)	Tasks involving sequential dependencies across multiple individuals requiring communication and coordination	• planning/action formulation • coordination
Interacting tasks across multiple individuals (coordination demands)	Interacting tasks across multiple individuals requiring tight communication and coordination.	• planning/action formulation • coordination
Decreased access to team members (remote teams)	Team members are physically distant, distant in time or have reduced richness of the communication medium.	• planning/action formulation • coordination

Example of Complicating Factors (5)

Factor	Description	Cognitive/Collaborative Functions Impacted
Workload	Requirements for multiple mental or physical actions that need to be accomplished within a limited period of time.	• detection/monitoring; • sense-making /situation assessment; • planning/action formulation; • attention/workload management
Attention demands (Attention bottlenecks)	Requirements for rapid attention shift.	• detection/monitoring; • sense-making /situation assessment; • planning/action formulation; • attention/workload management
Demands on prospective memory	Requirements to perform an activity in the future for which there is no strong memory cue under high attention and working memory load conditions.	• planning/action formulation; • attention/workload management
Interruptions (Memory bottlenecks)	Interruptions make it easy to forget to do unresolved tasks and prioritize tasks appropriately.	• planning/action formulation • attention/workload management



Checklist in Evaluating Validation Scenarios

- What is the model of support being tested? Do the validation scenarios provide opportunities to exercise and test this model of support?
- Do the scenarios capture the range of complicating factors that arise in the actual operational environment so as to assess extent and boundaries of effectiveness of support?
- Have the performance issues (e.g., potential vulnerabilities, biases, errors and breakdown points) that can impact the decisions and related cognitive and collaborative activities of interest been identified? Do the test scenarios create opportunities to assess the impact of the decision aid on these potential performance deficiencies?
- Have situations/probes/target events been embedded in the scenarios so as to create the opportunity for cognitive and collaborative activities of interest to be exercised in an *observable* manner?

Workshop on Human Factors Validation of Nuclear Power Plant Control Room Designs

Topical Session 3

Jean-Paul LABARTHE
Louis GRAGLIA

EDF R&D
Risk Management Department
Human Factors Group

Topical Session 3

Construction of Validation Test Scenarios, the Test Design and their Relationship to the Validation Framework

Methods and approaches used for scenarios development for the ISV program of a new reactor: the EDF utility experience

Summary

1. ISV program definition impacts for the construction of validation test scenarios
2. Scenario definition
3. Participant sampling and number of crews
4. Conclusion / discussion

1- ISV program definition impacts for the construction of validation test scenarios

A reasonable level of confidence must involve:

- Evaluation criteria and a robust evaluation methodology
- Simulations representative of the future operating context (scenarios, teams, operating resources)
- A recognised multi-disciplinary evaluation team, independent of the stakeholders

1- ISV program definition impacts for the construction of validation test scenarios

- Objectives of the programme
- Topics of analysis: coverage of all validation topics (organisation, HMI, procedure, etc.) throughout the different evaluations
- Operating situations: Covering the life cycle of operations: normal (start-up, shutdown, maintenance, etc.), emergency situations, etc.
- According to the progress of the project: different design stages in terms of operating resources and simulations,
- …
- ISV program validated with the regulator and different stakeholders
  – Before launching the first campaign: reduces the risks for the instruction

2- Scenario definition

Identification of operating activities to simulate

- Depending on the objectives of the test campaign and the scope defined in the ISV programme
- Primary criteria for selecting situations:
  – Frequency of use (routine NPP situations, etc.)
  – Operating issues/criticality (safety, production, etc.)
  – Innovations
  – Operating complexity (e.g. accumulation of breakdowns, problem solving, etc.)
  – Variety (SE used, types of procedure, usage contexts, etc.)
- Identification, for each situation, objectives in terms of evaluation (topics / cases covered) required on driving performance expected, observable for the collection of data

Multi-disciplinary development process

- Expertise in safety, operation (operator and designer), human factors (ergonomics and HRA), training: to define and analyse the scenarios

2- Scenario definition

Grouping operating activities in realistic scenarios

- 1 scenario = realistic grouping of operating activities, for several hours (3 to 4 hours, or 8 hours to simulate an entire shift)
- Combination of simple scenarios, including an accumulation of breakdowns, and more complex scenarios (accumulation of thermal-hydraulic accidents and fire or loss of safety functions), always plausible (expert advice), possibly based on feedback from operations

Number of scenarios to achieve during a ISV

- Depending on what must be evaluated:
  - Project constraints (schedule and resources):
  - Aim for a minimum number of scenarios to evaluate each evaluation theme at least once (to be subsequently multiplied by the number of teams who will act out this scenario)
  - 5 to 10 scenarios per campaign depending on the objectives of the evaluation/validation (targeted vs. overall)

3- Participant sampling and number of crews

Profile

- Depends whether or not the personnel are recruited on-site
- Profiles differed in preliminary and final ISV (mixed)

Training

- Specific to the tests (scope of the scenarios) to limit training biases

➔ need to develop dedicated training courses prior to every validation, without revealing the scenarios: significant logistical challenges

Number of crews

- Availability constraints
- All teams act out the same scenarios (enables comparisons)
- At least 3 teams to limit "individual"/crew factors

Safeguarding the tests

Number of scenarios

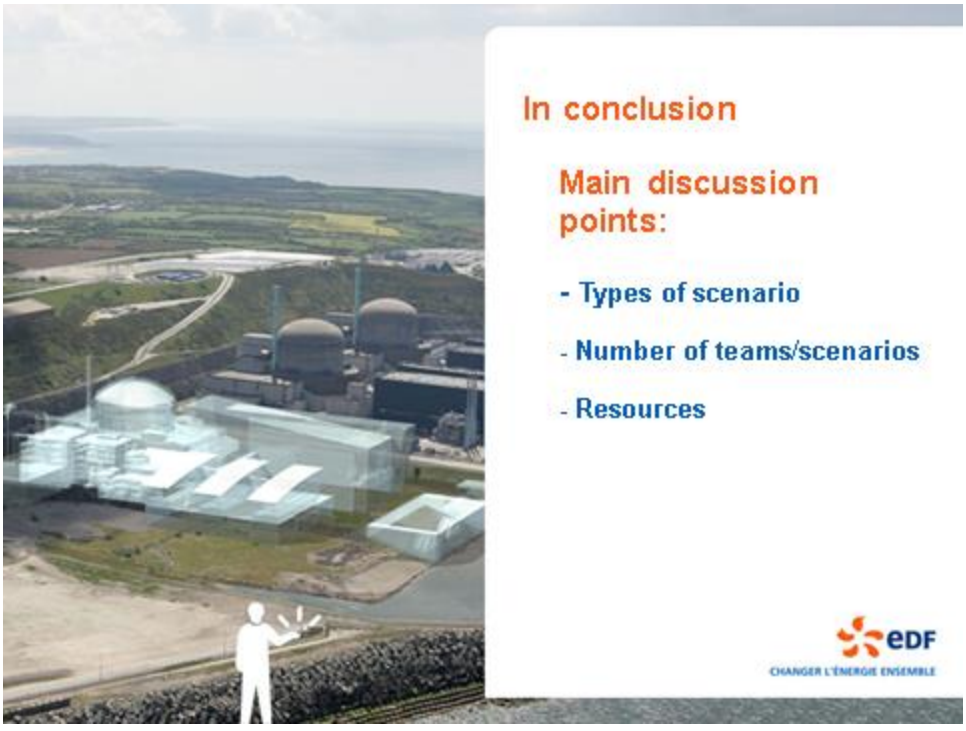
- 1 or 2 extra scenarios can be established to deal with a difficulty (technical problem on the simulator, documents no longer available, etc.)

Validation of the scenarios

- Designer's involvement in the definition of the scenarios: validate the availability of operating resources (simulator, procedures, images, etc.)
- Technical validation on the simulator: make sure the simulator is adapted to the operating resources
- Dry runs: also refine the evaluation process

Schedule

- Allow for at least one extra day so that a scenario can be replayed in the event of a problem => reserve resources accordingly (simulator, teams, etc.)



Thank you for your attention

Topical Session 4
Analysis of Validation Results

TOPICAL SESSION 4

Gyrd Skraaning Jr
OECD Halden Reactor Project
PO Box 173, NO-1751 Halden
gyrds@hrp.no

Abstract of Challenge Presentation

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.

The challenge presentation will argue that this methodology 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 organized, weighted and judged together from multiple angles to reach a conclusion on whether the control room is, and will remain acceptable. This process is similar to a trial court or a safety case, 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. In other words, 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. Thus, it will be argued in the presentation 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. 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, prioritized and compared to reach a trustworthy decision on acceptability.

The workshop will be challenged to suggest alternative data analysis and interpretation strategies to determine the final validation conclusion. Following the reasoning above, a separate acceptability analysis stage of the validation is one possible way forward (see the White Paper; *Integrated System Validation (ISV): The Acceptability Analysis Process*).

Gyrd Skraaning jr
OECD Halden Reactor Project

1

Prescriptive validation methods

- Process-based validation where safety is justified by following a prescribed validation procedure
- Presumptions
  - human performance requirements can be fully understood, pre-defined and specified
  - correct execution of prescribed process results in clear and convincing validation evidence
  - procedure generalizes across validation contexts
    - e.g., types of plants, operational concepts, analog-hybrid-digital control environments, modernizations-new builds, levels of automation

2

Hypothesis

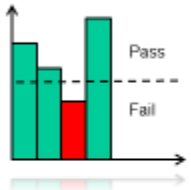
Prescriptive validation methods fail to provide "…the statistical and logical bases for determining that performance of the integrated system is, and will be acceptable" (NUREG-0711 rev3, p.93)

Current Guidance

appeals to a simple psychometric logic

develop acceptance criteria for human performance measures

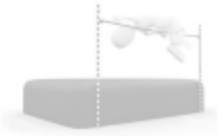
collect human performance data



check if observed performance scores satisfy acceptance criteria

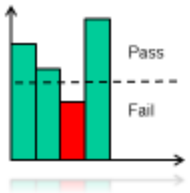
accept or reject design solution

Data analysis and interpretation



develop acceptance criteria for human performance measures

collect human performance data



check if observed performance scores satisfy acceptance criteria

accept or reject design solution

5

Data aggregation

- Data on many levels of aggregation
  - crews
  - operator roles
  - individual operators
  - responsible operators
  - scenario types
  - scenarios
  - scenario periods

6

Types of performance measures

- Pass/fail vs. diagnostic measures
- Plant performance
- Task performance
  - task completion
  - response times
- Team performance
- Cognitive performance
  - situation awareness
  - workload

7

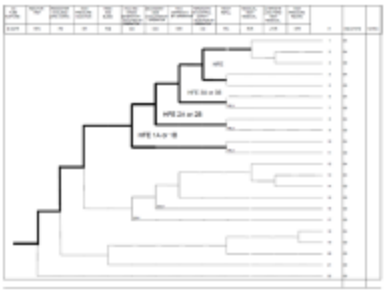
Objective of measurement

- Express the safety of human performance in the sampled scenarios
- Capture the robustness and generalizability of the observed performance
  - Anticipate the safety of future operation
  - «Resilience indicators»

8

Graded approach

- Prioritize importance of findings according to identified criteria, such as
  - novelty of tested design features
  - impact on operational safety



9

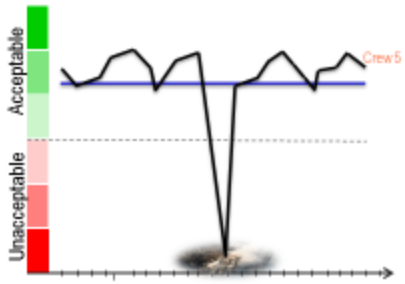
Analysis and interpretation

- Huge and complex data set
  - separate spurious effects from systematic patterns
  - understand contra-intuitive findings
  - handle conflicting evidence

10

Minimum vs. typical performance

- Evaluate the acceptability of
  - the lowest performance scores
  - average performance



11

Adverse effects

- Discover unanticipated adverse effects of the new control room design (if any)
- Examples
  - loss of team transparency
  - out-of-the loop effects
  - alienation

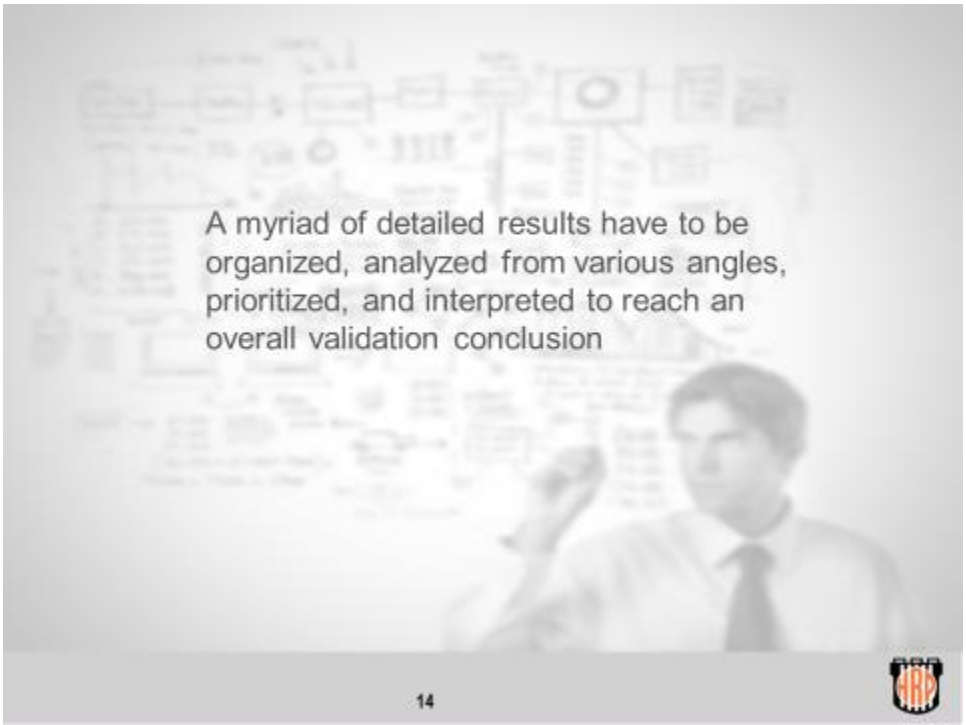
"About your anti-dandruff shampoo."

12

Take weaknesses into account

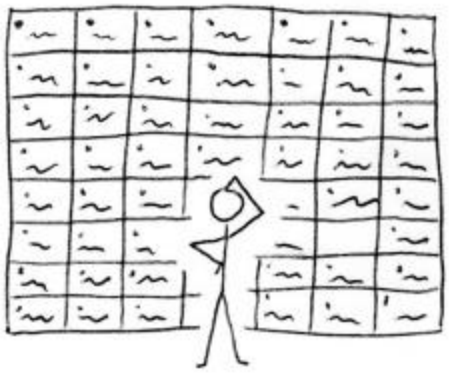
- Consider practical and principal weaknesses of the validation approach, e.g.
  - meaningfulness of benchmarking
  - reliability of expert rating
- Interpret findings in light of general methodological vulnerabilities
  - simulator fidelity, scenario representativeness, participant training, statistical conclusion validity etc. (see NUREG-0711 rev3, pp. 85-93)

13



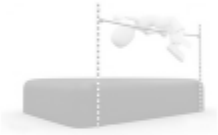
14

Real life



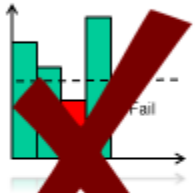
15

Data analysis and interpretation



develop acceptance criteria for human performance measures

collect human performance data



check if observed performance scores satisfy acceptance criteria

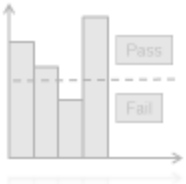
accept or reject design solution

16

Acceptance criteria

develop acceptance criteria for human performance measures

collect human performance data



check if observed performance scores satisfy acceptance criteria

accept or reject design solution

17

A priori acceptance criteria

- Validation outcome cannot bias the criteria
- Impossible to adjust the acceptance criteria when you know the performance results
- Preferred from a theoretical point of view

18

Real world

- Operators may use the new control environment in unexpected ways
- Design solutions can have unanticipated effects on risk taking, teamwork, role allocation, work processes etc.
- Judgment of new technologies and its use is biased by the past

19

This "telephone" has too many shortcomings to be seriously considered as a means of communication

- Memo at Western Union, 1878

The cinema is little more than a fad. It's canned drama. What audience really want to see is flesh and blood on the stage

– Charlie Chaplin, 1916

People will soon get tired of staring at a plywood box every night

- Darryl Zanuck, 20th Century Fox, 1946

There is no reason anyone would want a computer in their home

- Digital Equipment Corporation (DEC) president, chairman and founder Ken Olsen, 1977

20

Thus

- *A priori* acceptance criteria seem unrealistic and insufficient

UK92480 treatment of heart disease

Unsuccessful medical trails in the 1990s

New situation

develop acceptance criteria for human performance measures

collect human performance data

check observed performance scores satisfy acceptance criteria

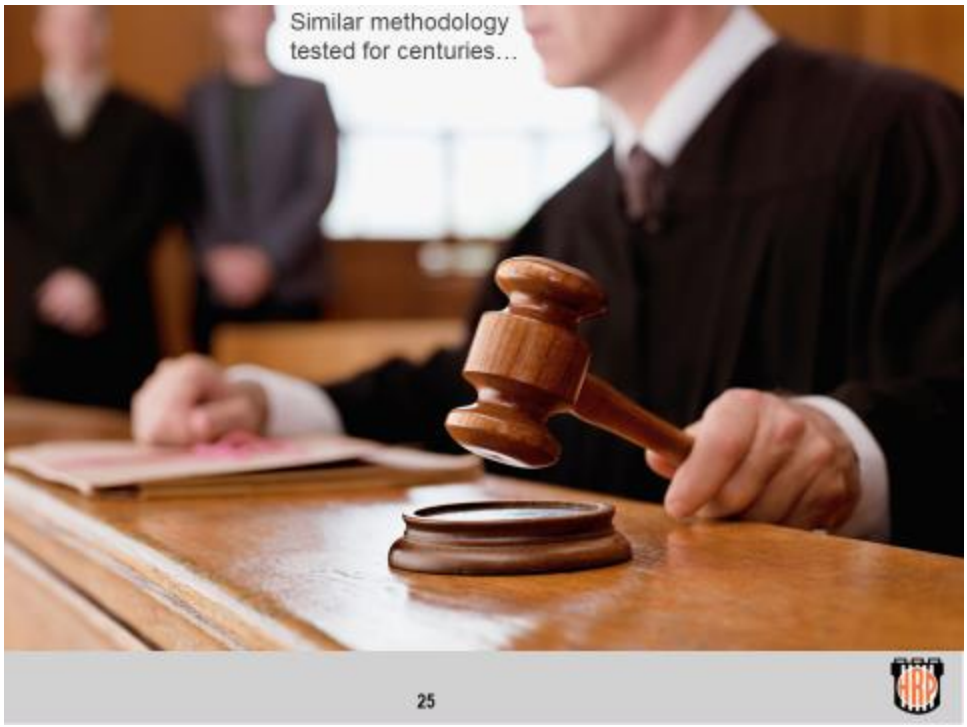
accept or reject design solution

How can we establish reasonable confidence in the validation without following a prescribed procedure?

Safety Case

- Safety is justified by structured arguments supported by evidence
- Used for complex decision making where formal proof is impossible
- Present a coherent argument for safety based on a complex body of evidence
- Reasonable confidence given by
  - rational and systematic interpretation of evidence
  - not by compliance to a prescribed validation process

24


Similar methodology
tested for centuries…
25

Beyond Skepticism: Induction, Falsification, and Reasonable Confidence in MCR Validation

Robert B. Fuld

Experts Workshop on HF Validation of NPP Control Room Designs & Modifications

Charlotte, NC

21 Feb 2015

Dimension of Confirmatory Processes (Fuld, 1997)

PROCESS	VERIFICATION	VALIDATION
Empirical Goal	Truth	Strength
System Model	Closed	Open
Reliability Tests	Redundancy/repetition	Diversity/replication
Breadth of Observation	Population of events	Sample of events
Object of Analysis	Decomposed element	Interacting aggregate
Temporal focus	Past / historical	Future / predictive
Questions	Well-defined	Ill-defined
Answers	Categorical	Incremental
Confidence	Strong	Weak
Reasoning	Deductive	Inductive
Judgment	Rule-based	Knowledge-based
Conclusions	Symmetrical	Asymmetrical
Complete	Yes	No

Engineering validation is necessary, because:

1. Adequate strength of the aggregate system must be demonstrated.
2. The future/system/reality remains open.
3. The event sample is very limited.
4. The relevant questions cannot all be defined.
5. We need reasonable confidence in successful validation outcomes to place systems in-service.

We should not be skeptical:

1. If we prepare sufficiently (e.g. per 0711).
2. If we correct problems found during the process and progress incrementally to a refined state.
3. If we accept the simulator modeling of safety parameters.
4. If we identify a representative test set to rigorously challenge the system on expected (and other) operating conditions.
5. If we pass or fail on safety criteria, which are objective, conservative, and independently established & confirmed.

These things give grounds for reasonable confidence in passing outcomes, as obtained.

Falsification (Popper, 1953)

1. Successful validation
   - The null result is the 'pass' of all safety criteria. We should have grounds for reasonable confidence in corroboration of this null.
   - If testing is successful we learn little – we already had grounds for reasonable confidence - but we confirm what we already thought.
   - Successful validation is inductive.
2. Failed validation
   - The alternate result is the 'failure' of some safety criteria. This is taken as an objective fact, given the simulation model.
   - If testing fails, we are surprised - but we learn something.
   - Failed validation is deductive (*modus tollens*).

Passing only corroborates adequacy of the current design (which may still fail someday), while failing 'proves' that the current design is not yet adequate.

Pass/Fail vs. Diagnostic Criteria

1. Pass/Fail criteria are primarily objective.
2. Diagnostic criteria (DCs) are primarily subjective.
3. Pass/Fail criteria should not depend on DCs for decisions of acceptability or retest.
4. Pass/Fail criteria should be as few as reasonably possible.
5. DCs should be as diverse as reasonably possible.

More on Diagnostic Criteria (DCs)

1. DCs are only states of belief. Their impact on safety is not clear, less so if they diverge from Pass/Fail criteria.
2. DCs may help explain why a particular trial failed.
3. DCs may be the source of HEDs, so an acceptable design may still need to address some DC results for resolution.
4. DCs may be the subject of varied analyses as are deemed worthwhile, but DCs themselves are not a basis for passing or failing a trial.

Convergent Results Analysis

1. Signal Detection Theory (SDT) offers a model.
2. Physical safety limits (Pass/Fail criteria) on measurable plant parameters define the 'state of the world' or signal/noise rates (i.e., unsafe/safe design).
3. Comments and subjective ratings (i.e., DCs) define the 'state of belief' or response rates.
4. The two may be compared to determine Hit (F/V), Miss (F/N), False Alarm (P/V), and Correct Rejection (P/N) rates for the DCs.

Risk-important Human Actions (RIHAs)

1. Are a convenient sample of HAs (e.g. for timing).
2. Give insight on accuracy of PRA estimates.
3. Are rarely related to safety.
4. May not be significant to plant risk or to acceptable PRA results.
5. May change entirely with the PRA.
6. May invite negative training (like TMI).

Thus, RIHAs should be used only as DCs, and not as Pass/Fail criteria.

Repetitions for Validation

1. Minimizing unnecessary repetitions is a key means to economize validation exercises.
2. Repetition is performed (in validation) to ensure that an obtained result is repeatable and not unique.
3. A repeatable result may be shown with as few as two trials.
4. Repetition is not performed to increase confidence in validation outcomes.

Summary

1. Successful engineering validation is necessary prior to plant service.
2. Falsification is the appropriate model for engineering validation.
3. Reasonable confidence in successful validation outcomes is justified by performing appropriate preliminaries.
4. Failure of the system should be based on safety limits (i.e., Pass/Fail criteria).
5. Diagnostic criteria and RIHAs should be used as sources for analysis, insight and HEDs.
6. Convergent results may be assessed by SDT.
7. Unnecessary repetitions should be minimized.

Questions?