

Use of IDHEAS General Methodology to Generalize Operator Performance Data in Halden Experiments for Human Factors and Reliability Analysis Applications

Jing Xing, Ph.D. and Niav M. Hughes, Ph.D.
U.S. Nuclear Regulatory Commission, Washington, DC USA
301-415 (2410; 2362), (Jing.Xing; Niav.Hughes}@nrc.gov

Abstract

The Integrated Human Event Analysis System (IDHEAS), a human reliability analysis method developed by the U.S. Nuclear Regulatory Commission, provides a hierarchical structure to analyze and assess the reliability of human actions. The method is based on cognitive science and is capable of incorporating human performance data to support the estimation of human error probabilities. IDHEAS models human performance in five macrocognitive functions: Detection, Understanding, Decision-making, Action execution, and Teamwork. IDHEAS defines a set of cognitive failure modes for each function to describe the various ways of failing to perform the function. IDHEAS analyzes an event in progressively more detailed levels: event scenario, human actions, critical tasks of the actions, macrocognitive functions and cognitive failure modes of the tasks, and performance influencing factors. This structure provides an intrinsic interface to generalize and integrate various sources of human error data for applications in human factors and human reliability analysis. We analyzed and generalized two Halden Reactor Project operator performance experiments: the 2003-2004 task complexity study and the microtask study on human-system interface evaluation. This paper presents IDHEAS-G structure along with the demonstration of generalizing the experimental results in the studies. The data, once sufficiently populated, can inform applications of human factors and human reliability. Introduction

1. Introduction

The Halden Reactor Project (HRP) has been conducting human performance experiments with nuclear power plant (NPP) simulators as part of their Man Technology Organization (MTO) program for decades. The experimental results provide technical basis for understanding operators, organization, and technologies in NPPs. To highlight a few past experiments of interest, HAMMLAB has been host to thorough lines of investigation of a number of interesting topics including group-view display functionality [1][2][3][4][5], human-automation interaction (which is well summarized in HPR-387[6]), teamwork and work practices [7][8][9][10][11], and staffing strategies for highly automated plants [12]. More recently, some projects are leveraging access to training simulators outside of HAMMLAB as a way of testing operators in their home plant and using their procedures and interface such as in the 2016 investigation of operator response to the failure of a computerized procedure system [13].

Each HRP human performance experiment addresses specific issues regarding new technologies, design concepts, or conduct of operations. To explore the issues, HRP researchers design and perform experiments with specific systems or technologies (e.g., computerized procedures), human-system-interfaces (e.g., large overview displays), or concepts of operation (e.g., reduced control room staffing level). Operation-based scenarios are designed to allow testing the issues. Over the years, HRP have developed many human performance measures, such as operator performance indices, task performance accuracy, task completion time, situational awareness, etc. The experiments use those measures to examine operators' performance as the operators perform required human actions in the scenarios. Researchers draw conclusions about the issues based on the human performance measures and document the experiments in Halden Work Reports (HWRs).

By design, the results of an experiment are applicable to the specific experiment scope, i.e., the systems, human-system-interfaces (HSIs), scenarios, and concepts of operation specified in the study. Yet, the results should be able to provide further insights on general principles about how systems, HSIs, tasks in the scenarios, and concepts of operation may affect human performance. Such insights would allow using the results beyond the experimental scope. Moreover, while individual experiments address specific issues, it is desirable to the HRA and HF communities to learn the state-of-the-art of human performance research from the full mosaic of MTO experiments: What have we learned about human performance and what more do we need to learn?

Given such specific scope and design of each experiment, to gain these insights about the principles and learn the state-of-the-art more broadly, individual experiments need to be generalized in a common framework. Any human performance experiment should be able to be generalized into the common framework, and the framework should be tied with fundamental principles of human performance. This report documents our pilot study of generalizing HRP human performance experiment results using a cognitive framework.

The General Methodology of Integrated Human Event Analysis System (IDHEAS-G), a human reliability analysis (HRA) method developed by the U.S. Nuclear Regulatory Commission (NRC), provides a hierarchical structure to analyze human events [14]. The method is based on a cognitive framework. It models human performance in five macrocognitive functions: *Detection*, *Understanding*, *Decision-making*, *Action execution*, and *Teamwork*. IDHEAS-G defines a set of processors for each function to describe the ways of achieving the function. IDHEAS-G analyzes an event in progressively more detailed levels: event scenario, human actions, critical tasks of the actions, macrocognitive functions and cognitive failure modes (CFMs) of the tasks, and performance influencing factors (PIFs). This structure provides an intrinsic interface to generalize various sources of human performance data to elucidate the general human performance principles underlying the data. Previously, the NRC staff reviewed a large body of human error data from the literature and available human error databases, generalized the data into the IDHEAS-G structure, and used the generalized dataset to inform expert judgment of HEPs for nuclear power plant human actions outside the control rooms. Thus, we chose to use the IDHEAS cognitive structure as a framework to generalize HRP human performance experiment results.

Two HRP experiments were selected for piloting the generalization. The first one is HWR-758 [15], "The Task Complexity Experiment 2003/2004." The purpose of this experiment was to explore how additional tasks added to base case scenarios affected the operators' performance of the main tasks. The experiment ran five types of complicated scenarios. The second study is HWR-1212 [16], "Evaluation of Design Features in the HAMBO Operator Displays." The purpose of the study was to evaluate whether innovative design features in the HAMMLAB Boiling water reactor (HAMBO) displays show any performance impacts, compared to a conventional presentation of the same information using HRP's microtask methodology. The experiment required operators to retrieve information from static or dynamic displays of several scenarios in the Conventional and Innovative designs. These two studies represented the simulation of very complicated operational scenarios versus the most simple isolated detection tasks. The motivation for this pilot study was threefold: 1) demonstrating generalization of experimental results from these two types of studies into a common cognitive framework, 2) gaining new insights pertaining to general human performance principles from the reported results, and 3) exploring a systematic way to communicate experimental results and interpret the results pertinent to human factors engineering and human reliability analysis.

2. Method

2.1 Overview of IDHEAS Cognitive Framework

The IDHEAS-G cognition framework includes a model of macrocognition that describes the brain processes (i.e., cognitive mechanisms) associated with the success or failure of a task, and a performance influencing factor (PIF) model that describes how various factors affect the success or failure of tasks.

2.1.1 The Macrocognitive Model

The macrocognitive model elucidates the cognitive process of human performance in applied work domains where human tasks are complex and often involve multiple individuals or teams. The model is described as follows:

- Macrocognition consists of five functions: *Detection*, *Understanding*, *Decisionmaking*, *Action Execution*, and *Teamwork*. The first four functions may be performed by an individual, a group or a team, and the *Teamwork* function is performed by multiple groups or teams.
- Any human task is achieved through these functions; complex tasks typically involve all five functions;
- Each macrocognitive function is processed through a series of basic cognitive processors; failure of a cognitive processor could lead to the failure of the macrocognitive function;
- Each processor is reliably achieved through one or more cognitive mechanisms; errors may occur in a cognitive processor if the cognitive mechanisms are challenged;
- PIFs challenge the capacity limits of cognitive mechanisms and can lead to ineffectiveness of the mechanisms.

Table 1 shows the basic cognitive elements for the macrocognitive functions. The cognitive mechanisms are not presented due to the space limits of the paper.

Table 1: Macrocognitive Functions and Their Basic Elements

Detection	Understanding	Decisionmaking	Action execution	Teamwork
D1- Initiate detection - Establish mental model and criteria D2 – Identify and attend to sources of information D3- Perceive, recognize, and classify information D4-Verify the information acquired D5- Retain or communicate the acquired information	U1 - Assess/select data U-2 Select / adapt / develop the mental model U-3 Integrate data with mental model to maintain situational awareness, diagnose problems, or resolve conflicts U-4 Verify, revise, and iterate the understanding U-5 Communicate the understanding	DM1 – Manage the goals DM2 – Adapt a decision model DM3 – Acquire / select information DM4 - Make judgment or plans DM5 - Simulate the decision DM6 - Communicate and authorize the decision	E1 - Assess action plan E2 - Develop / modify action scripts E3 - Synchronize, supervise, and coordinate action implementation E4 - Implement action scripts E5 - Verify and adjust actions	T1 - Establish or adapt teamwork infrastructure T2 - Manage information T3 - Maintain common ground T4 - Manage resources T5 - Plan inter-team collaborative activities T6 - Implement decision/command T7 - Verify, modify, and control the implementation

2.1.2 The Performance Influence Factor Model

PIFs affect cognitive mechanisms and increase the likelihood of macrocognitive function failure. We developed a PIF model that is independent of HRA applications and links to cognitive mechanisms. The model systematically organizes PIFs to minimize inter-dependency or overlapping of the factors. The PIF structure has four layers:

1. PIF category: PIFs are classified into three categories corresponding to characteristics of systems, tasks, and personnel.
2. PIFs: Each category has high-level PIFs describing specific aspects of the systems, tasks, or personnel. Table 2 shows the PIFs within the three categories.

Table 2: Performance influencing factors in IDHEAS-G

System- and environment-related	Task-related PIFs	Personnel-related PIFs
<ul style="list-style-type: none"> • Transparency of systems and instrument & control • Work location accessibility and habitability • Tools and equipment • Environmental factors 	<ul style="list-style-type: none"> • Information availability and reliability • Scenario familiarity • Multi-tasking, interruptions and distractions • Cognitive complexity • Mental fatigue and stress • Physical demands 	<ul style="list-style-type: none"> • Human-system interface (HSI) • Staffing • Training • Procedures, guidelines, instructions • Teamwork factors • Work process

3. PIF attributes: These are the specific traits of a performance influencing factor. A PIF attribute represents a poor PIF state that challenges cognitive mechanisms and increases the likelihood of errors in cognitive processes. Table 3 shows some example attributes of the PIF *Information availability and reliability*.

Table 3: Example Attributes of the PIF “Information Availability and Reliability”

Nominal state of <i>Information availability and reliability</i> : Information is needed for personnel to perform tasks. Information is expected to be complete, reliable, unambiguous, and available timely to personnel.
<ul style="list-style-type: none"> • Inadequate updates of information (e.g., a party receives information but fails to inform another party). • Information of different sources is not synchronized. • Conflicts in information

4. Every PIF attribute challenges one or several cognitive mechanisms. IDHEAS-G provides links between PIF attributes and cognitive mechanisms synthesized and inferred from the literature.

2.1.3 Implementation of IDHEAS-G Cognition Framework for Human Reliability Analysis

IDHEAS-G implements the cognition model in the general HRA process, which includes qualitative analysis and HEP quantification. It begins with an event and progressively analyzes more detailed elements of the event: operational narratives and context of the scenarios, human actions in the scenarios, critical tasks in the human actions, macrocognitive functions, cognitive failure modes

(CFMs) of the critical tasks, the PIF states, and human error probabilities. Analysis of these elements is carried out through the following process:

- Scenario analysis: Develop the operational narrative for the scenarios and identify human performance context
- Identification and definition of human actions: Identify the key human actions pertinent to the mission of the event and define the human actions
- Task analysis: Analyze tasks required for the human action and characterize the critical tasks for HEP quantification
- Time uncertainty analysis: Analyze time uncertainties in the human action and quantify the HEP attributing to time uncertainties
- Cognition failure analysis: Identify cognition failure modes of every critical tasks in a human action and estimate the HEP attributing to failures of macrocognitive functions for the critical tasks

2.2 Generalization of HRP human performance experiment results with IDHEAS-G framework

We took the following process to generalize the results in the two selected HWRs:

1. Thoroughly read the HWRs and other related reports (e.g., HRP has several reports on the same topic as that of HWR-758);
2. Apply IDHEAS-G process to analyze the experiment and results, i.e., representing the experiment results in IDHEAS-G elements: operational narrative and context of the scenarios, important human actions, task analysis, macrocognitive functions and applicable failure modes, states of PIFs relevant to operators' tasks, error rates of operators' task performance and other performance measures;
3. Analyze the results to gain insights of general human performance principles with respect to the human reliability analysis and human factors engineering;
4. Document the description of the study (including the HWR abstract) and generalized information in 2) and 3);
5. Verify the analysis with peers.

The piloting results of generalization in this report were verified by the two authors. The next step is to verify our analysis with the HRP researchers who performed the studies. Thus, the information reported in the RESULTS section of this paper is only for the purpose of demonstration.

3. Results

We documented the results in the Tables 4 and 5, one for each of the studies. Each table has three main sections following the abstract of the original paper: documentation of the experimental design (Part I), generalization of the design and results to the IDHEAS-G framework (Part II), and insights and lessons learned gained from the analysis (Part III).

Table 4: Analysis of results of the task complexity experiment 2003/2004

Reference	<p>Halden HWR-758</p> <p>THE TASK COMPLEXITY EXPERIMENT 2003/2004</p> <p>The purpose of this experiment was to explore how additional tasks added to base case scenarios affected the operators' performance of the main tasks. These additional tasks were in different scenario variants intended to cause high time pressure, high information load, and high masking. The experiment was run in Halden Man-Machine Laboratory's Boiling Water Reactor simulator.</p>
-----------	---

	<p>Seven crews participated, each for one week. There were three operators in each crew. Five main types of scenarios and 20 scenario variants were run. The data from the experiment were analyzed by completion time for important actions and by in-depth qualitative analyses of the crews' communications. The results showed that high time pressure decreased some of the crews' performance in the scenarios. When a crew had problems in solving a task for which the time pressure was high, they had even more problems in solving other important tasks. High information load did not affect the operators' performance much and, in general, the crews were very good at selecting the most important tasks in the scenarios. The scenarios that included both high time pressure and high information load resulted in more reduced performance for the crews compared to the scenarios that only included high time pressure. The total amount of tasks to do and information load to attend to seemed to affect the crews' performance. To solve the scenarios with high time pressure well, it was important to have good communication and good allocation of tasks within the crew. Furthermore, the results showed that scenarios with an added complex, masked task created problems for some crews when solving a relatively simple main task. Overall, the results confirmed that complicating, but secondary tasks that are not normally taken into account when modelling the primary tasks in a Probabilistic Risk Analysis (PRA) scenario can adversely affect the performance of the main tasks modelled in the PRA scenario.</p>
Part I. Documentation of Experimental design	
Scenario	Five PRA scenarios, each with four variations
Subject	7 crews
Variables	Base scenario, Base scenario + additional task Base scenario + additional information load Base scenario + additional task + additional information load
Measures	Task completion time <u>Number of Tasks correctly completed</u> Observation of communication Observation of teamwork Individual crews' performance stories
Part II. Generalization of the design and results for Experiment 1: Scenario 4 - Incomplete Scram/Start Of The Boron System	
Scenario analysis	
Scenario definition	In this scenario there is a leakage in the main feedwater system. The leakage leads to isolation of the feedwater system and reactor scram. Normally, at reactor scram, there are 169 control rods inserted into the core within 4 seconds. In this scenario there are 12 nearby control rods that are stuck and the rods will not be inserted by the back-up control rod drive system either. The operators are supposed to start the boron system to reduce the reactor power.
Initial conditions	Production at full power. Ordinary maintenance-period in safety system train C is ongoing. Emergency pumps and diesel-generator in train C are blocked. One of six main cooling water pumps (441 PC1) has a failure. The pump is in operation as usual but the current measurement indicates 0 %. The current measurement is input to the logic for the cooling-function of the Main condenser. The fast transfer function in train D is in manual mode due to a failure. No other known problems.
Initiating event	A failure in the reactor protection system leads to alarm "354 KB701." To reduce the reactor power operators have to start the boron system manually. There are procedures for this situation.
Boundary conditions	At the plant, there is a push button to start the boron system in one of the two trains (train C and D). In the simulator, the push button is placed to the left in the process picture "351 PC1_1 Borinumping". It is also possible to start the system by maneuvering the components one by one.
Event context	Environnemental context: None System context: multiple failures, non-responsive controls Personnel context: Normal crew, good training, procedures are available for the primary task but may not be detailed enough for the secondary tasks Task context: <ul style="list-style-type: none"> ● The primary task and multiple secondary tasks are going on within the same period of time ● Time available is adequate but personnel has time pressure due to multiple failures ● Key information is masked or misleading in some scenario variation
Critical actions	Action A: start the boron system manually
Task analysis of Action A	
Time uncertainty	Adequate time available
Critical tasks	Start the boron system manually

Cognitive activities	<p>Detect the cue: alarm "354 KB701 532 V group B1 malfunction".</p> <p>Diagnosis and decisionmaking: straightforward per alarm response procedures</p> <p>Action execution: start boron system by pressing a button</p> <p>Special requirements for teamwork: None</p> <p>Parallel tasks:</p> <p>Base scenario 4.1 - Several additional tasks take additional time but they are not intermingled multitasks and not necessarily interruptions.</p> <p>Scenario 4.2 – Additional system failure cues are onset before the initiating event; operators need to take care of the secondary tasks.</p> <p>Scenario 4.3 – Additional information load (indications / alarms) require the crew to detect and understand.</p> <p>Scenario 4.4: 4.1 + 4.2 + 4.3</p>
Applicable cognitive failure modes	<p>Failure of detection: D4 - Key alarm "354 KB701 532 V group B1 malfunction" not attended to</p> <p>Failure of decision: DM2 - Incorrectly prioritize competing goals in Scenarios 4.2, 4.3, and 4.4 (not performing the critical task first)</p> <p>Failure of execution – E3 – Action not initiated</p> <p>Failure of execution – E4 - Failure of simple action (start the boron system)</p>
PIF states	<p>All scenarios:</p> <p>Distraction: Secondary tasks may distract crews - Low impact</p> <p>Competing goals: Crews may choose to attend to secondary tasks first - Low impact</p> <p>Scenario 4.2: Time pressure: due to added tasks - Low impact</p> <p>Scenario 4.3: Distraction: Added information load causes more distraction - Low to moderate impact</p> <p>Scenario 4.4: Added tasks and information load caused i) time pressure, ii) more distraction, iii) possible mental fatigue - low to moderate impact</p>
likelihood of failure	<p>Scenario 4.1 – All CFMs are highly unlikely</p> <p>Scenario 4.2 - All CFMs are highly unlikely</p> <p>Scenario 4.3 - All CFMs are highly unlikely</p> <p>Scenario 4.4 - DM2 possible – Crew may attend to other tasks first; E3 possible – Crew may forget to initiate the action execution of the primary task due to distraction</p>
Experimental Results	
Number of incorrectness and CFMs	<p>0/28 – Seven crews in four scenarios all correctly started boron system</p> <p>There were errors at the CFM level but the errors were recovered.</p> <p>1/28 D4- Key alarm "354 KB701 532 V group B1 malfunction" not attended to in Scenario 4.1. The shift supervisor did not seem to know that some control rods other than those that belong to scram group C1 were not inserted. Later he realized that by asking the reactor.</p> <p>3/28 DM2 – Incorrectly prioritize competing goals in Scenario 4.4. After feedwater isolation, the shift supervisor first suggested that the reactor operator should open a scram valve in the scram system (354), then said that the reactor operator should start the boron system.</p> <p>1/28 E4 _incorrectly executed simple action in Scenario 4.4. The shift supervisor said that the reactor operator should try to start the boron system. The shift supervisor tried several times to start the boron pump on the man-machine interface (Instead of starting the boron system, he tried to start the boron pump and he did not manage to start the boron system). The shift supervisor asked the reactor operator to look at what he did wrong on the boron system, and the reactor operator started the boron system in train C.</p>
# of failures in non-critical actions and added tasks	<p>5/14 E4 –Incorrectly execute a simple action in Scenario 4.2 and 4.4. Crews did not close the open pressure relief valve (314 VA2). they forgot to open the pressure control valve (314 VA17) before they closed the isolation valve (314 VA23).</p> <p>4/20 E4 – Fail to open scram valve in group B1 (354 VB1). Two crews lacked knowledge for doing this task and they failed the task in all the scenarios.</p> <p>1/28 E4 – Fail to Insert the 531 detectors" in Scenario 4.2</p>
Part III. Performance stories and insights	
Depend-ency	<p>Two of the crews that did not close the pressure relief valve (314 VA2) did not start enough auxiliary feedwater pumps to keep the level in the reactor tank.</p>
Error of commiss-ion	<p>The crew monitored that the low-pressure injection pumps started to pump water to the reactor tank. Then the reactor operator said that they had to be careful that they did not dilute all the boron in the reactor tank and asked if they should plan to stop the low-pressure injections pumps. The shift manager called and discussed the situation with the plant management (simulated) and they agreed to stop the low-pressure injection pumps and to stop the boron in one train, so that they could keep that boron tank in reserve until they had got control over the level in the reactor tank. (One out of two boron tanks are enough to make the reactor under-critical). The reactor operator then stopped the low-pressure injection pumps when the level in the reactor tank was about 6 meters in the reactor tank. Both of the crew who got low level in the reactor tank activated the TB switch to get fast</p>

	depressurization and they also stopped the low pressure injection system so they did not dilute the boron too much. Both of these actions were very critical and important for safety.
Change of conduct of operation	In the scenarios with high time pressure there were many failures and the crews had relatively short time to solve them. The crews who solved these scenarios well managed to divide the tasks in the scenario; the shift supervisor worked on some of the tasks and delegated other tasks to the reactor operator. The shift supervisor also divided the tasks between the reactor and the turbine operator. This changed from the three-way-communication. The safety impact of the change is not assessed here.
Work process	In those scenarios with high time pressure, the work practice with first checks did not work well. The reactor operator had many failures to work on, and also the shift supervisor did some of them. There did not seem to be time to do first checks and to report the first checks within 10 minutes. A safety analysis should be performed to assess the work process of handling the situation like this.

Table 2: Analysis of microtask evaluation of innovative versus conventional design features

Reference	Halden HWR-1212 Evaluation of design features in the HAMBO operator displays. A current challenge within the nuclear industry is to assess the impact of computerized interfaces on human performance. The purpose of this study was to evaluate whether innovative design features in the HAMmlab BOiling water reactor (HAMBO) displays show any performance impacts, compared to a conventional presentation of the same information: Is there added value in having innovative design features in addition to the numerical information? We used a within-subject experimental approach where nine participants responded to the same blocks of questions in two conditions: with innovative features such as mini-trends, pie-charts and bar graphs; and a control condition where the process information was presented through numerical information only. Overall, the performance results showed that the participants were more accurate in the innovative condition and showed equivalent response times in both conditions. Pie charts, mini-trends and pictorial elements eliminated the need to recall values from memory such as nominal and previous process parameters. Bar graphs were advantageous for checking the status of multiple components or systems. For questions that required verification of parameters, there were no differences in accuracy between the conditions, but the operators tended to answer quicker in the conventional displays, suggesting that the innovative features might have acted as a distractor when they were not actively relevant for the question. Eye tracking analysis of the questions with the largest differences between conditions showed that dwell times and fixation counts tended to be lower in the innovative condition. The heat maps also suggested more focused attention in the innovative condition. These findings suggest that the participants were able to locate and identify relevant information more effectively in the innovative displays. The analysis of the operators' confidence ratings showed a tendency for overconfidence in incorrect responses that is interesting to explore in future studies.
Part I. Documentation of Experimental design	
Scenario	No real scenario. Operators individually observes the simulator displays and answer questionnaire
Methodology	Microtask
Subject	Nine Swedish operators
Variables	Conventional display (Numerical information only) vs. Innovative Display (Mini trends, Pie Charts, Pictorial Elements, and Other (e.g., bar graph, trend diagrams)) Static screen shoots vs. 5-min dynamic scenario observation Types of questions in the questionnaire (type of cognitive task = 1) assessing trends 2) multiple parameters 3) check values)
Measures	Accuracy of answering questionnaire Response time of answering questionnaire Eye tracking
Part II. Generalization of the design and results for questions on static and dynamic scenarios.	
Scenario analysis	Not applicable
Task analysis	
Time uncertainties	Adequate time with time pressure ("Answer the questions as accurate as possible and as quick as possible")
Critical tasks	Answer every question in the questionnaire
Cognitive activities	Acquire information Note: The task is to acquire information even for the dynamic scenarios because the operators had no task when the scenario was playing.

Applicable failure modes	Failure of detection: D1 – Incorrect or no mental model - Not applicable. The question asked established the mental model of what to detect D2 – Attend to wrong sources – Applicable D3 – Incorrect perception, recognition, or classification – Applicable D4 – Failure of verification, peer-checking, or supervision – Applicable D5 – Failure of retaining or communicating detected information – Not applicable		
PIF states		Conventional	Innovative
	Work process	No peer-checking, No supervision, Poor verification (because the instruction “as quick as possible”) No feedback	Same as Conventional
	Information	Incomplete information for trending questions with conventional display	Nominal
	HSI	Nominal to low impact - Parameters needed for a task are spatially distributed - Saliency may be less than nominal	Nominal to low impact - Presentation of parameters is complicated and needs interpretation - More clutter

Prediction on likelihood of failure		Conventional	Innovative
	D2	Unlikely	Unlikely
	D3	- Low for most questions – Should be the same as Innovative because HSI is nominal. - Very high or infeasible for questions about trending or other questions that required retrospective memory; Operators would not pay attention to trending even with dynamic scenarios because they did not have the task when the information was available, and the trending information was not available when they performed the detection tasks.	- Low for most questions. - High for verifying the parameters because of clutter and complicated graphic representation.
	D4	Moderate likelihood of errors due to lack of peer-checking/supervision and possibly not much verification	Moderate likelihood of errors due to lack of peer-checking/supervision and possibly not much verification
	Pt	Very low because of adequate time	

Experimental Results

% Correctness	<table border="1"> <thead> <tr> <th>Type of questions</th> <th>Conventional Static</th> <th>Innovative Static</th> <th>Conventional Dynamic</th> <th>Innovative Dynamic</th> </tr> </thead> <tbody> <tr> <td>Trending</td> <td>0.67</td> <td>0.87</td> <td></td> <td></td> </tr> <tr> <td>Multiple parameters</td> <td>0.78</td> <td>0.89</td> <td></td> <td></td> </tr> <tr> <td>Check values</td> <td>0.94</td> <td>0.88</td> <td></td> <td></td> </tr> <tr> <td>Overall</td> <td></td> <td></td> <td>0.86</td> <td>0.91</td> </tr> </tbody> </table>					Type of questions	Conventional Static	Innovative Static	Conventional Dynamic	Innovative Dynamic	Trending	0.67	0.87			Multiple parameters	0.78	0.89			Check values	0.94	0.88			Overall			0.86	0.91
	Type of questions	Conventional Static	Innovative Static	Conventional Dynamic	Innovative Dynamic																									
	Trending	0.67	0.87																											
	Multiple parameters	0.78	0.89																											
	Check values	0.94	0.88																											
	Overall			0.86	0.91																									
	<p>Table 11 in HWR-1212 shows the questions that revealed the lowest accuracy in the dynamic scenarios.</p> <table border="1"> <thead> <tr> <th></th> <th>Question</th> <th></th> <th></th> <th>Conv</th> <th>Innov</th> </tr> </thead> <tbody> <tr> <td>12</td> <td>The pressure in the RPV has exceeded 7.3 MPa.</td> <td>323</td> <td>Pictorial</td> <td>0.44</td> <td>0.67</td> </tr> <tr> <td>22</td> <td>The highest temperature in the reactor containment was 58 °C.</td> <td>323</td> <td>Pictorial</td> <td>0.22</td> <td>0.89</td> </tr> <tr> <td>35</td> <td>The condensate flow 462 KB301 has increased.</td> <td>462</td> <td>Mini-trend</td> <td>0.11</td> <td>1</td> </tr> </tbody> </table>						Question			Conv	Innov	12	The pressure in the RPV has exceeded 7.3 MPa.	323	Pictorial	0.44	0.67	22	The highest temperature in the reactor containment was 58 °C.	323	Pictorial	0.22	0.89	35	The condensate flow 462 KB301 has increased.	462	Mini-trend	0.11	1	
		Question			Conv	Innov																								
	12	The pressure in the RPV has exceeded 7.3 MPa.	323	Pictorial	0.44	0.67																								
	22	The highest temperature in the reactor containment was 58 °C.	323	Pictorial	0.22	0.89																								
35	The condensate flow 462 KB301 has increased.	462	Mini-trend	0.11	1																									

	40	The level in 462 TD1 has been constant	462	Mini-trend	0.11	0.89
Response time	Overall, response time for Innovative is several seconds shorter than that for Conventional.					
Eye tracking	Eye tracking analysis of the questions with the largest differences between conditions showed that dwell times and fixation counts tended to be lower in the innovative condition. The heat maps also suggested more focused attention in the innovative condition. These findings suggest that the participants were able to locate and identify relevant information more effectively in the innovative displays.					
Part III. Performance stories and insights						
Innovative HSI	Innovative HSI design features have the potential of improving human performance by reducing the negative impact of some PIFs; they also have the potential of deteriorating human performance by increasing the negative impact of some other PIFs. For example, the graphic features improve the HSI attribute that related parameters are spatially distributed. Yet, they introduce the new HSI attribute that parameter representations are complicated and need interpretation.					
Conventional HSI	While the Innovative HSI resulted in higher accuracy for questions asking for trending (.87) and assessing multiple parameters (.89), the highest accuracy reported was 0.94 for checking parameter values with Conventional, but the accuracy for Innovative was still in the high range (.88). This result suggests that when the cognitive activity required is checking straight values without additional cognitive manipulations, the numerical information only condition is sufficient and possibly even preferred. Caution when applying this interpretation is suggested, however, in that when performing a real scenario, simply checking a value in isolation (i.e., not in combination with other cognitive activities) is rare. Thought should be given when determining how indications will be used so as to avoid unnecessary duplication or use of conventional HSI which may neglect to acknowledge the necessity for additional cognitive manipulations.					
ISV	While the mean accuracy is 0.86 for Conventional and 0.91 for Innovative, Table 11 shows that the accuracies for some questions are very poor with Conventional, and the accuracies for Question 12 are poor for both Conventional and Innovative. The result implies that some safety issues can be hidden in the high average accuracy.					
ISV	The conclusion of the study is that operators performed better with Innovative than Conventional as indicated by the mean accuracy and response time. However, being a few second faster in the response time does not make significant difference in operators' job performance. For accuracy, removing the questions in Table 11 where the accuracies for 4-5 questions are extremely low would result in roughly equal accuracies for Conventional and Innovative. Those questions were infeasible for Conventional because they all asked for the information that no longer existed when operators performed the detection tasks. In other words, when the data (accuracies for individual questions) fall in bimodal or multi-modal distributions, simply averaging them can be misleading.					
Work process / conduct of operation	The highest accuracy reported was 0.94 for checking parameter values with Conventional. This is still too high compared to the error rates when operators perform real scenarios as a crew. The only non-nominal performance influencing factor here was no peer-checking/supervision/feedback. The result may provide evidence for the effect of peer-checking/supervision/feedback.					
Confidence Ratings	The results of the confidence ratings indicated an overconfidence in incorrect responses. As noted in the HWR, the phenomenon is established in cognitive psychology [17], however, it is as of yet unexplored in the context of nuclear operators in the nuclear power plant main control rooms (NPP MCRs) (See [18] for previous Halden work using self-rating bias measurement). The phenomenon demonstrated could be considered from a methodological standpoint for researchers doing research in the NPP MCRs as it has possibility to influence results and conclusions. It should be further explored and/or considered when choosing measures and methodologies for studying expert operators in the nuclear domain. Furthermore, as mentioned in the HWR, confidence ratings can help us to better understand decision making [19] and again, from a methodological standpoint, might aid in understanding differences between novice and expert operators [20]					
Eye Tracking	Eye tracking analysis of the questions with the largest differences between conditions showed that dwell times and fixation counts tended to be lower in the innovative condition. The heat maps also suggested more focused attention in the innovative condition. These findings suggest that the participants were able to locate and identify relevant information more effectively in the innovative displays.					
Eye Tracking/Met hodology	"We also explored the usefulness of eye tracking methodologies within the context of interface assessment for nuclear process control. The findings showed that eye metrics could mediate the interpretation of specific performance outcomes by highlighting					

	monitoring patterns and styles during the tasks. The operators considered the eye-tracking glasses as comfortable throughout the study and reported that it did not interfere with their tasks, which is a relevant advantage of the equipment.”
Microtask/Methodology	“Used the microtask method as a procedure to evaluate HSI design features in the HAMBO operator displays for the first time. The method was flexible enough to enable the data collection and analysis according to multiple variables such as operator roles, types of tasks, questions, design features, or displays. The data collection was efficient, with a short session providing a significant amount of data.”

4. Conclusions

IDHEAS-G is a general HRA methodology built on a cognitive framework. Its layered structure allows for generalization of human performance data of different formats and various levels of detail. This pilot study demonstrated the feasibility of using the IDHEAS-G cognitive framework to generalize the experimental results of very complicated, full scenarios run by crews as well as the results of very simple microtasks performed by individual operators. Moreover, the generalization of the results in a cognitive framework allows for identifying insights of general human performance principles that are applicable beyond the specific experiment scope. The generalization provides a systematic understanding of human performance by elucidating contextual factors, challenging human performance, delineating how personnel perform tasks for the required human actions, and revealing how and why personnel may fail to perform required actions as represented through the applicable cognitive failure modes and the associated performance influencing factors. These together enhance the technical basis to improve system design, new technologies, concepts of operation, procedures, and training in NPP control room operations.

5. Disclaimer

This paper presents a research project conducted by the staff in the U.S. Nuclear Regulatory Commission. It does not represent an NRC official position. Although the NRC staff may suggest a course or action in the paper, these suggestions are not legally binding and the regulated community may use other approaches to satisfy regulatory requirements.

6. References

- [1] Hurlen, L., Skraanning, G., Myers, B., Carlsson, H., and Jamieson, G. (2014). The Plant Panel: Feasibility Study of an Interactive Large Screen Concept for Process Monitoring and Operation. (HWR-1129). OECD Halden Reactor Project, Halden, Norway.
- [2] Skraanning, G., Hurlen, L., Le Darz, P. & Jamieson, G. (2016) Feasibility Study of an Interactive Large Screen Concept for Automated Plant Startup. (HWR-1179). OECD Halden Reactor Project, Halden, Norway.
- [3] Massaiu, S. and Holmgren, L. (2016). Preliminary Results from the 2013 Resilient Procedure Use Study with Swedish Operators. (HWR-1122). OECD Halden Reactor Project, Norway.
- [4] Massaiu, S. and Holmgren, L. (2017). The 2013 Resilient Procedure Use Study with Swedish Operators: Final Results. (HWR-1216). OECD Halden Reactor Project, Halden, Norway.
- [5] Eitriheim, M. Svengren, H., and Fernandes, A. (2017). Evaluation of the Human System Interface Concept for Near-Term Applications. (HWR-1211). OECD Halden Reactor Project, Norway.
- [6] Skraanning, G. and Jamieson, G. (2017). Twenty Years of HRP Research on Human Automation Interaction: Insights on Automation Transparency and Levels of Automation. (HPR-387). OECD Halden Reactor Project, Halden, Norway.

- [7] Strand, S., Kaarstad, M., Svengren, H., Karlsson, T. & Nihlwing, C. (2010). Work Practices 2009 HAMMLAB Study: Team Transparency in Near-Future Computer-Based Control Rooms. (HWR-952). OECD Halden Reactor Project, Halden, Norway
- [8] Kaarstad, M., Strand, S. & Nihlwing, C. (2012). Work Practices and New Technologies – iPad as a Tool for the Shift Supervisor to monitor Process Information. (HWR-996.) OECD Halden Reactor Project, Halden, Norway.
- [9] Kaarstad, M., Strand, S. & Nihlwing, C., Holmgren, L. & Berntsson, O. (2014). Control Room and Field Operator Collaboration – Use of a Handheld Tool. (HWR-1124). OECD Halden Reactor Project, Halden, Norway
- [10] Skranning, G. (2016). A Reanalysis of the Work Practice Experiments in HAMMLAB (2009-2014). (HWR-1194). OECD Halden Reactor Project, Halden, Norway.
- [11] Skjerve, A.B.M., Nihlwing, C., Nystad, E. (2008). Lessons learned from the extended teamwork study. (HWR-867). Halden, Norway: OECD Halden Reactor Project.
- [12] Eitrheim, M. H., Skraaning, G, Lau, N., Karlsson, T., Nihlwing, C., Hoffmann, M., & Farbrod, J. E. (2010). *Staffing Strategies in highly automated future plants: Results from the 2009 HAMMLAB Experiment*. (HWR-938). OECD Halden Reactor Project, Halden, Norway
- [13] Taylor, C., Hildebrandt, M., Hughes, N. & McDonald, R. (2016). Operator response to failure of a computerized procedure system. Results from a training simulator study. (HWR-1198). OECD Halden Reactor Project, Halden, Norway.
- [14] US Nuclear Regulatory Commission (2019). The general methodology of an integrated humanevent analysis system (IDHEAS-G) , NUREG-2198, 2019 (In preparation)
- [15] Laumann, K., Braarud, P.O., Svengren, H. (2005). The Task Complexity Experiment 2003/2004. (HWR-758). OECD Halden Reactor Project, Halden, Norway.
- [16] Eitrheim, M. H., Fernandes, A., and Svengren H. (2017). Evaluation of design features in the HAMBO operator displays. (HWR-1212). OECD Halden Reactor Project, Halden, Norway.
- [17] Joshua Klayman, Jack, B. Soll, Claudia Gonzalez-Vallejo, Sema Barlas. 1999. Overconfidence: It depends on How, What, and Whom You Ask. *Organizational Behavior and Human Decision Processes*, 79(3), 216-247.
- [18] Massaiu, S., Skjerve, A.B.M., Skraaning Jr., G., Strand, S., Wærø, I. (2004). Studying Human-Automation Interactions: Methodological Lessons Learned from the Human-Centred Automation Experiments 1997-2001. (HWR-760). *OECD Halden Reactor Project*.
- [19] Roger Ratcliff, Jeffrey J. Starns. 2013. Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, 120(3), 697–719. <http://doi.org/10.1037/a0033152>.
- [20] Mark T. Spence, Merrie Brucks. 1997. The Moderating Effects of Problem Characteristics on Experts' and Novices' Judgments. *Journal of Marketing Research*, 34(2), 233.