INL/RPT-22-69167

Light Water Reactor Sustainability Program

Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER) Demonstration: Part 1, Empirical Data Collection of Operational Scenarios

September 2022

U.S. Department of Energy

Office of Nuclear Energy

DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

INL/RPT-22-69167

Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER) Demonstration: Part 1, Empirical Data Collection of Operational Scenarios

Jooyoung Park, Ronald L. Boring, Thomas A. Ulrich
Idaho National Laboratory

Taewon Yang, Jonghyun Kim
Chosun University

September 2022

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy

This page intentionally left blank.

ABSTRACT

From within the umbrella of the Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER) framework, this report analyzes human performance differences between professional and student operators when using a simplified simulator (i.e., the Rancor Microworld Simulator). The purpose of this report is to collect and document human performance from an operator-in-the-loop study on the simulator. These human performance data serve as the foundation for modeling and validating HUNTER simulations of operational scenarios. A goal of the HUNTER project is to provide demonstrations of a variety of realistic operating scenarios for nuclear power plants. Data from ten scenarios are collected and documented in this report. These scenarios will subsequently be modeled in HUNTER and documented in a planned future report.

ACKNOWLEDGMENTS

This report presents the findings of a simulator study completed by Chosun University under subcontract from Idaho National Laboratory. The authors wish to thank the Risk-Informed System Analysis (RISA) Pathway of the U.S. Department of Energy's Light Water Reactor Sustainability Program for its support of the research activities presented in this report. In particular, we thank Svetlana Lawrence, pathway lead for RISA, for championing the demonstrations found in this report.

CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS iv

ACRONYMS viii

1. INTRODUCTION 1
   - 1.1 Project Background 1
   - 1.2 Study Background 1

2. EXPERIMENTAL DESIGN 3
   - 2.1 Independent Variables 4
   - 2.2 Experiment Scenarios 4
   - 2.3 Human Performance Measures 6
     - 2.3.1 Workload 6
     - 2.3.2 Situation Awareness 8
     - 2.3.3 Error 8
     - 2.3.4 Time to Completion 8
     - 2.3.5 Eye Movements 10
     - 2.3.6 Number of Manipulations 10
   - 2.4 Apparatus 11
     - 2.4.1 A Representative Simplified Simulator: Rancor 11
     - 2.4.2 Eye-tracking System 11
     - 2.4.3 Video Recorder System 12
     - 2.4.4 Procedures 12
     - 2.4.5 Participants 12
   - 2.5 Training 12
   - 2.6 Data Acquisition 12
   - 2.7 Experimental Procedure 13

3. INVESTIGATION INTO DIFFERENCES IN HUMAN PERFORMANCE 13
   - 3.1 Analysis of Each Human Performance Measure 13
     - 3.1.1 Workload 16
     - 3.1.2 Situation Awareness 16
     - 3.1.3 Error 17
     - 3.1.4 Time to Completion 18
     - 3.1.5 Eye Movements 18
     - 3.1.6 Number of Manipulations 21
   - 3.2 Analysis of the Human Performance Relationship between Operators and Students 22

4. DISCUSSION 26
   - 4.1 Differences in Human Performance Measures 26
   - 4.2 Differences in Task Performance and Workload 27
   - 4.3 Differences in Operator and Student Monitoring and Control of the Simulator 28

5. CONCLUSION 28

FIGURES

Figure 1. The Simplified Human Error Experimental Program (SHEEP) framework. 3

Figure 2. Questionnaire for the MCH rating scale (Gawron, 2019). 7

Figure 3. Questionnaire for the SART rating scale (Taylor, 2017). 9

Figure 4. Example of the procedure format. 10

Figure 5. Example of log data generated from a simulator. 10

Figure 6. The Rancor Microworld Simulator interface. 11

Figure 7. Trend of the overall MCH results. 16

Figure 8. Trend of the overall SART results. 17

Figure 9. Trend of the overall error rate. 17

Figure 10. Trend of the overall average time to complete a task. 18

Figure 11. Overall eye fixation count per task. 19

Figure 12. Overall blink rate. 19

Figure 13. Heatmap for the operator and student groups in non-event scenarios. 20

Figure 14. Heatmap for the operator and student groups in event scenarios. 20

Figure 15. Illustration of how colors are assigned based on eye fixation counts (Niehorster et al., 2020). 21

Figure 16. Trend of the overall number of manipulations per task. 21

Figure 17. Scatter plot of the relationship between the MCH rating and the SART scores, both with the operator and student groups combined and with them separated. 24

Figure 18. Scatter plot of the relationship between the MCH rating and the error, both with the operator and student groups combined and with them separated. 25

Figure 19. Scatter plot of the relationship between the SART scores and the error rates, both with the operator and student groups combined and with them separated. 25

Figure 20. Relationship among time to completion, number of manipulations, and eye movements, both with the operator and student groups combined and with them separated. 26

Figure 21. Two human performance characteristic groups. 27

Figure 22. Relationship between workload and task performance. 28

TABLES

Table 1. Randomized factorial experiment design 4
Table 2. Experiment scenarios, success criteria, and related procedures (Park et al., 2022a) 5
Table 3. Human performance categories and the measures. 6
Table 4. Summary of data acquisition methods, items collected, and human performance measures. 12
Table 5. Average values for the human performance measures. 14
Table 6. ANOVA test result (p-value) 15
Table 7. Correlation analysis for all participants. 22
Table 8. Correlation analysis for the operator group. 23
Table 9. Correlation analysis for the student group. 23

ACRONYMS

ANOVA	analysis of variance
AOI	area of interest
AOP	abnormal operating procedure
CNS	Compact Nuclear Simulator
COBHRA	computation-based human reliability analysis
EOC	error of commission
EOO	error of omission
EOP	emergency operating procedure
HDMI	high-definition multimedia interface
HEP	human error probability
HRA	human reliability analysis
HUNTER	Human Unimodel for Nuclear Technology to Enhance Reliability
HuREX	Human Reliability Data Extraction
INL	Idaho National Laboratory
IRB	Institutional Review Board
KAERI	Korea Atomic Energy Research Institute
LED	light-emitting diode
MCH	Modified Cooper-Harper
NASA	National Aeronautics and Space Administration
NPP	nuclear power plant
NRC	Nuclear Regulatory Commission
OP	operating procedure
p	probability
r	correlation
RISA	Risk-Informed System Analysis
SACADA	Scenario, Authoring, Characterization, and Debriefing Application
SART	Situation Awareness Rating Technique
SGTR	steam generator tube rupture
SHEEP	Simplified Human Error Experimentation Program
TLX	Task Load Index
TO	turbine operator
U.S.	United States
USB	universal serial bus

HUMAN UNIMODEL FOR NUCLEAR TECHNOLOGY TO ENHANCE RELIABILITY (HUNTER) DEMONSTRATION: PART 1, EMPIRICAL DATA COLLECTION OF OPERATIONAL SCENARIOS

1. INTRODUCTION

1.1 Project Background

Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER; Boring et al., 2021) is a framework and software tool for computation-based human reliability analysis (COBHRA). Human reliability analysis (HRA) methods have historically been worksheet based and static, designed to estimate the human error probability (HEP) for fixed accident sequences. In contrast, COBHRA methods like HUNTER are simulation-based, allowing more complex modeling of multiple accident sequences and determination of additional performance measures beyond HEPs, like overall time required for human activities. HUNTER allows consideration of novel contexts like digital control rooms or balance-of-plant activities, providing benefits over historical HRA such as the ability to explore what-if modeling for situations that do not have extensive operating experience.

HUNTER as a software implementation is newly released in 2022 and has to date mainly considered steam generator tube rupture (SGTR) scenarios for validating the HRA models and software code. There is a need to develop additional use cases to refine the functionality of HUNTER and provide demonstrations of how HUNTER may be applied for industry applications. A new feature being implemented in HUNTER is greater interoperability with plant simulators, thereby allowing plant-specific risk modeling within COBHRA. Efforts in 2022 have been focused on interfacing the HUNTER code with the Rancor Microworld Simulator (Ulrich et al., 2017), a simplified simulator of a pressurized water reactor. One goal of Rancor has been the use of a simplified simulator for gathering HRA data using non-licensed reactor operators like students. A study was recently conducted to compare human performance of 20 student and 20 professional reactor operators using Rancor across ten industry-realistic operating scenarios.

The results of the study are presented in this report. Rancor automatically logs objective and subjective performance. The objective performance measures are found in Rancor's data historian, providing time-stamped logs of human actions and plant parameters during the course of the scenarios. The subjective performance measures include measures where the operators report their subjective experience related to operational aspects like workload and situation awareness. These objective and subjective measures are initially used to help understand the course of human activities during operating scenarios, which is then used to create more realistic models of operator actions in HUNTER. Additionally, these measures may separately be used to validate the response of HUNTER in terms of human error rates and timing on tasks. A follow-on report, planned for November, 2022, will discuss the integration of scenarios from this report into HUNTER. The present report summarizes the empirical findings of the operator-in-the-loop study that will be used to inform HUNTER development and modeling.

1.2 Study Background

Human performance data play an important role in the quantification portion of HRA (Park, Jung, & Kim, 2020). To date, a great deal of effort has been made to secure HRA data collected from operational experience (Hallbert et al., 2006; Preischl & Hellmich, 2013, 2016; Sträter, 2000), simulator studies (Chang et al., 2014; Joksimovich, Spurgin, Orvis, Moieni, & Worledge, 1990; Jung, Park, Kim, Choi, & Kim, 2020; Lois, 2009; Park & Jung, 2007; Park, Lee, Jung, & Kim, 2017), and various other sources (Gertman, Gilmore, & Ryan, 1988; Kirwan, Basra, & Taylor-Adams, 1997) in order to provide a technical basis for HEP estimation within HRA. Recent related studies have mainly focused on simulator-based experimental research in which human performance data are collected by having actual operators licensed for nuclear power plants (NPPs) operate a full-scope simulator of a main control room. Currently, the largest of these data collection efforts are being led by the U.S. Nuclear Regulatory Commission (U.S. NRC) and the Korea Atomic Energy Research Institute (KAERI), both of which have collected HRA data from full-scope main control room simulators and actual operators via the Scenario Authoring, Characterization, and Debriefing Application (SACADA; Chang et al., 2014) and Human Reliability Data Extraction (HuREX; Jung et al., 2020) research projects, respectively.

The question remains to what extent it is possible to collect human performance data on nuclear plant operations without the use of full-scope simulators and professional reactor operators (Boring, Ulrich, & Lew, 2018). To complement full-scope simulator studies, Idaho National Laboratory (INL) developed an HRA data collection framework called the Simplified Human Error Experimental Program (SHEEP; Park et al., 2022a; see Figure 1). The SHEEP framework complements full-scope studies by suggesting a way to infer full-scope data based on experimental data collected from students operating simplified simulators, specifically the Rancor Microworld Simulator (Ulrich, 2017) and the Compact Nuclear Simulator (CNS; Kwon, Park, Jung, Lee, & Kim, 1997). The aim of the SHEEP framework is to lower the entry point for collecting useful HRA data by securing large sample sizes at a reasonable cost and amount of labor while also guaranteeing a high degree of freedom when designing experiments. The authors' previous research (Park et al., 2022a) investigated whether data collected from the SHEEP framework could support a representative full-scope study (i.e., the HuREX study). Specifically, student vs. operator errors when using a representative simplified simulator (i.e., Rancor) were incorporated into the HuREX framework and then quantitatively compared with the HuREX error data. Additional details on the SHEEP framework and related previous research are given in Park et al. (2022a). Here the goal is to use data from simplified simulators to inform COBHRA modeling in HUNTER.

This report aims to analyze human performance differences between professional and student operators when using Rancor. This comparison is important for HRA research, because (1) it is an essential effort in understanding the lack of fidelity of the simplified simulators and student operators within the SHEEP study, (2) it enables empirical comparison of what is often treated as the "experience" performance shaping factor in HRA methods, by comparing a relative novice against a highly experienced group; and (3) it determines the generalizability of the student results, given the noted challenges in securing large samples of professional operators for research studies. This report explores a randomized factorial experimental design that features two independent variables: participant type and event class. Scenarios and related procedures pertaining to both normal and abnormal situations were developed in the previous study and then simulated using Rancor. Six human performance measures were considered in the experiment: (1) workload, (2) situation awareness, (3) time to completion, (4) error, (5) eye movements, and (6) number of manipulations. The experiment was conducted using 20 professional reactor operators working at actual NPPs, along with 20 trained students. This report analyzes the resulting experimental data via statistical analysis methods such as analysis of variance (ANOVA) testing and correlation analysis. Lastly, this report discusses the differences in human performance between actual operators and students when using Rancor. These human performance data provide the empirical basis for understanding and modeling the same scenarios in HUNTER.

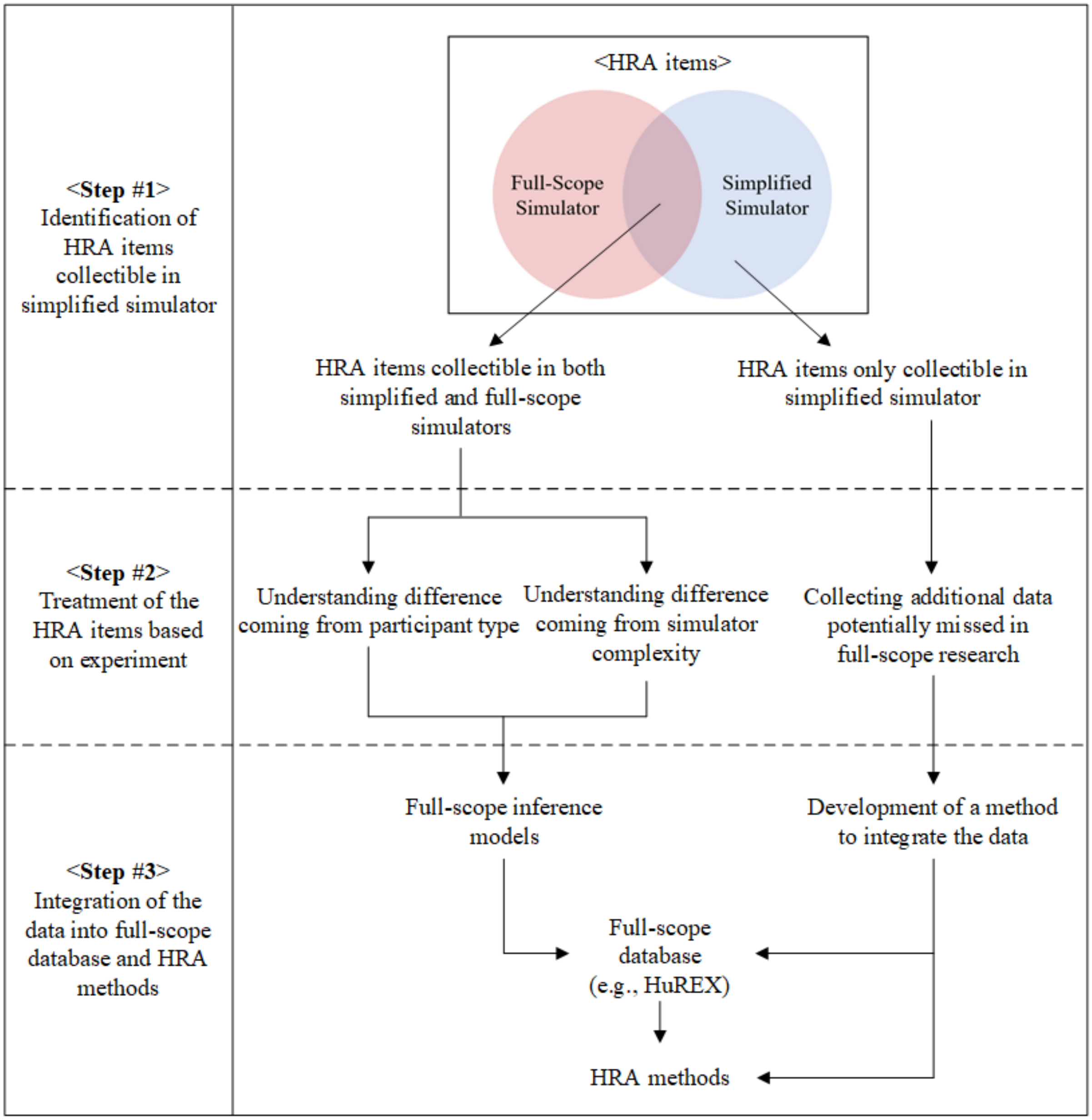


Figure 1. The Simplified Human Error Experimental Program (SHEEP) framework.

2. EXPERIMENTAL DESIGN

In this study, a randomized factorial experimental design (Garcia-Diaz and Phillips,1995) was used to compare the human performance of actual operators and trained students using Rancor. The novice vs. experienced operators are an important consideration for HUNTER in terms of the level of cognitive knowledge required to be modeled to make a functionally realistic COBHRA. The basic experimental design consisted of two independent variables: participant type and event class (see Table 1). In the experiment, each participant was presented three scenarios per event class, in random order. In other words, each participant performed six scenarios. A total of ten scenarios were available.

Table 1. Randomized factorial experiment design

Event Class	Participant Type		Scenario
	Actual Operator	Student	
Non-Event	Human performance measures	Human performance measures	· Fully auto startup (0 to 100%) · Shutdown (100 to 0%) · Startup with manual rod control (0 to 100%) · Startup with manual feedwater flow control (0 to 100%)
Event	Human performance measures	Human performance measures	· Failure of a reactor coolant pump under full-power operation · Failure of a control rod under full-power operation · Failure of a feedwater pump under full-power operation · Abnormal turbine trip under full-power operation · Steam generator tube rupture with an indicator failure · Loss of feedwater pump

2.1 Independent Variables

As mentioned above, this study considered two independent variables (i.e., participant type and event class). Participant type is classified into two groups. The first group consists of licensed operators currently employed at Korean NPPs, while the second consists of students involved in undergraduate or graduate courses at the Department of Nuclear Engineering at Chosun University. As for event class, there are also two groups: non-event and event scenarios. Non-event scenarios are close to the general operations usually performed in normal states such as startup, shutdown, or full-power operation. In these scenarios, participants may not feel as great a burden or as much time pressure in their work as they would in an event scenario. Event scenarios (e.g., abnormal or emergency situations) entail critical actions that should be completed within a limited time window and could positively or negatively affect the future state of the plant.

2.2 Experiment Scenarios

The scenarios and related procedures developed in the authors' previous research (Park et al., 2022a) were applied to the experiment. A list of the ten scenarios, success criteria and related procedures is summarized in Table 2.

Table 2. Experiment scenarios, success criteria, and related procedures (Park et al., 2022a)

No	Name	Description	Procedure	Success Criteria
1	Fully auto start-up (0% to 100%)	Increase reactor power from 0% to 100% in fully automatic mode	· OP-001, Start-up Operation, Steps 1–9	· Reactor power 100% · No reactor trip during the operation
2	Shutdown (100% to 0%)	Shut down the reactor from 100% to 0% in fully automatic mode	· OP-002, Shutdown Operation, Steps 1–8	· Reactor power 0% · No unintended reactor trip during the shutdown
3	Start-up with manual rod control (0% to 100%)	Increase reactor power from 0% to 100% with manual rod control	· OP-003, Manual Rod Control during Start-up, Step 1–9 · OP-004, Manual Rod Control, Step 1	· Reactor power 100% · No reactor trip during the operation
4	Start-up with manual feedwater flow control (0% to 100%)	Increase reactor power from 0% to 100% with manual feedwater control	· OP-005, Manual Feedwater Control during Start-up, Step 1–9 · OP-006, Manual Feedwater Control, Step 1	· Reactor power 100% · No reactor trip during the operation
5	Failure of a reactor coolant pump under full-power operation	According to failure of a reactor coolant pump during full-power operation, it is required to maintain safety functions and cool down the reactor coolant system temperature.	· AOP-001, Rapid Shutdown, Step 1–8	· Diagnosis of an initiating event or failure · Reactor coolant system temperature under 200°C
6	Failure of a control rod under full-power operation	According to failure of control rod during full-power operation, it is required to maintain safety functions and cool down the reactor coolant system temperature.	· AOP-001, Rapid Shutdown, Step 1–8	· Diagnosis of an initiating event or failure · Reactor coolant system temperature under 200°C
7	Failure of a feedwater pump under full-power operation	According to failure of a feedwater pump during full-power operation, it is required to maintain safety functions and cool down the reactor coolant system temperature.	· AOP-001, Rapid Shutdown, Step 1–8	· Diagnosis of an initiating event or failure · Reactor coolant system temperature under 200°C
8	Abnormal turbine trip under full-power operation	According to abnormal turbine trip during full-power operation, it is required to maintain safety functions and cool down the reactor coolant system temperature.	· AOP-001, Rapid Shutdown, Step 1–8	· Diagnosis of an initiating event or failure · Reactor coolant system temperature under 200°C
9	Steam generator tube rupture with an indicator failure	According to steam generator tube rupture, it is required to isolate damaged steam generator, maintain safety functions, and cool down the reactor coolant system temperature.	· EOP-01, Steam Generator Tube Rupture, Step 1–7 · AOP-001, Rapid Shutdown, Step 1–8	· Diagnosis of an initiating event or failure · Isolation of damaged steam generator · Reactor coolant system temperature under 200°C
10	Loss of feedwater pump	According to loss of feedwater pump, it is required to isolate damaged steam generator, maintain safety functions, and cool down the reactor coolant system temperature.	· EOP-02, Loss of Feedwater, Step 1–5 · AOP-001, Rapid Shutdown, Step 1–8	· Diagnosis of an initiating event or failure · Reactor coolant system temperature under 200°C

2.3 Human Performance Measures

In this experiment, data pertaining to six human performance categories—workload, situation awareness, time to completion, error, eye movements, and number of manipulations—were collected for each scenario. Table 3 lists the human performance categories and measures considered in the experiment. Some performance categories feature multiple types of measures, enabling the performance to be compared from different theoretical perspectives. These measure types are described below.

Table 3. Human performance categories and the measures.

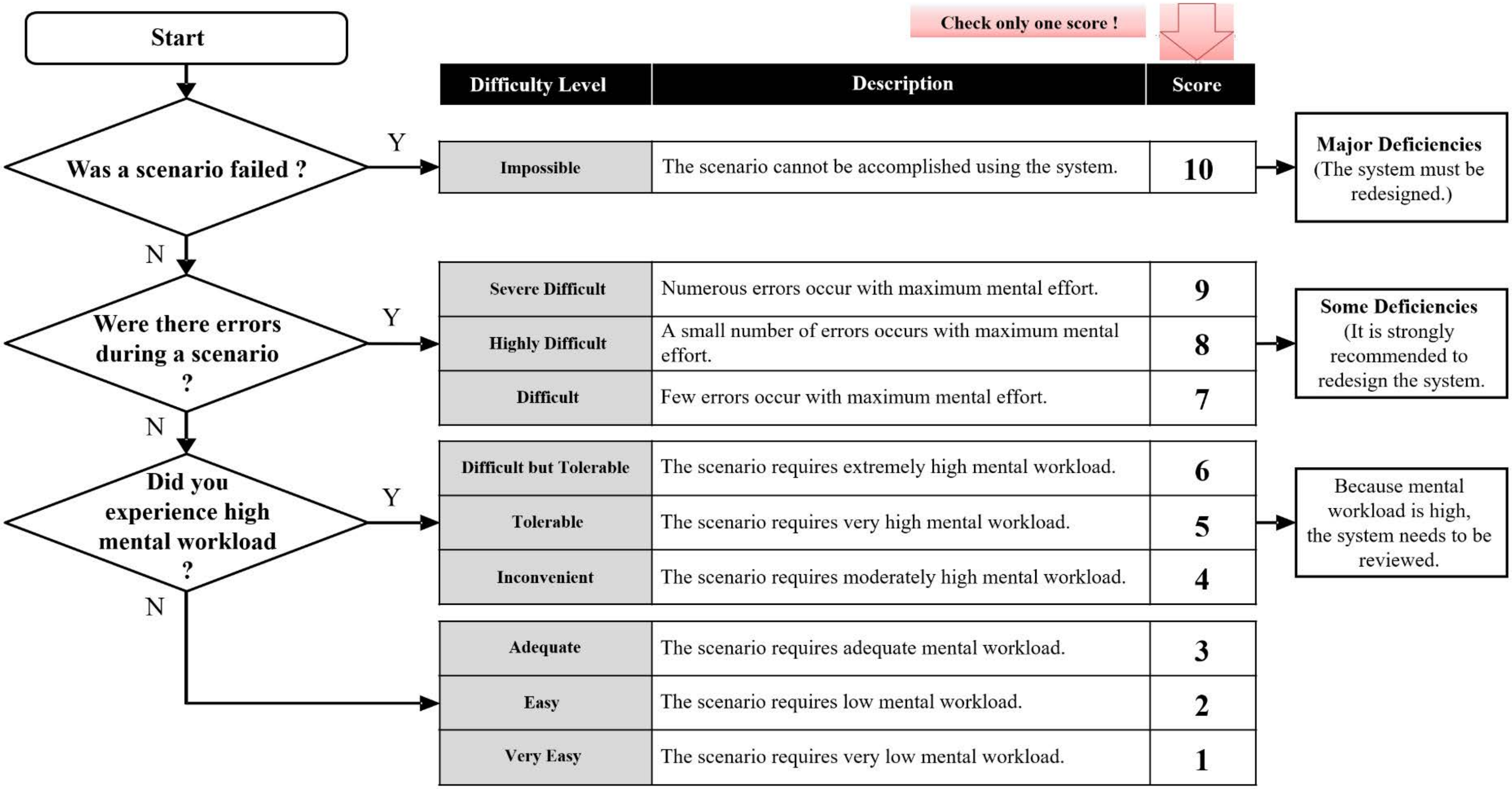
Human Performance Categories	**Human Performance Measures**
Workload	Modified Cooper-Harper (MCH)
Situation Awareness	Situation Awareness Rating Technique (SART)
Error	Error rate
Time to Completion	Average time to complete a step
	Average time to complete an instruction
	Average time to complete a task
Eye Movements	Eye fixation count per task
	Eye fixation duration per task
	Blink rate (i.e., blink count per task)
	Heatmap over area of interest (AOI)
Number of Manipulations	Number of manipulations per task
	Number of manipulations per scenario completion time

2.3.1 Workload

Workload is a human performance measure broadly employed and studied in human factors and reliability research. Representatively, Kim, Jung, and Kim (2014) collected and compared workload scores to experimentally investigate effects from digital human-system interfaces of NPPs. Park and Jung (2006) validated a task complexity measure by using workload scores collected from NPP operators. Park, Jung, and Kim (2020) experimentally analyzed the relationship among human performance measures including workload to handle a challenge of HRA. Cui et al. (2021) measured team workload through an experiment based on simulated maritime operation tasks.

NASA-Task Load Index (NASA-TLX; Hart and Staveland, 1988) and Modified Cooper-Harper (MCH; Cummings, Myers, & Scott, 2006; Gawron, 2019) are the representative methods for measuring and conducting a subjective mental workload assessment. As the post-test methods, the NASA-TLX and MCH questionnaires are generally filled out by participants to estimate workload scores after experiment scenarios.

To estimate workload, this experiment employed the MCH rating scale (Cummings, Myers, & Scott, 2006; Gawron, 2019), which was originally developed by the aviation industry to estimate pilots' psychological and physical workload, based on responses to a post-scenario questionnaire (see Figure 2). The MCH method has the advantage of suggesting design recommendations based on the rating scale.



Difficulty Level	Description	Score
Impossible	The scenario cannot be accomplished using the system.	10
Severe Difficult	Numerous errors occur with maximum mental effort.	9
Highly Difficult	A small number of errors occurs with maximum mental effort.	8
Difficult	Few errors occur with maximum mental effort.	7
Difficult but Tolerable	The scenario requires extremely high mental workload.	6
Tolerable	The scenario requires very high mental workload.	5
Inconvenient	The scenario requires moderately high mental workload.	4
Adequate	The scenario requires adequate mental workload.	3
Easy	The scenario requires low mental workload.	2
Very Easy	The scenario requires very low mental workload.	1

Figure 2. Questionnaire for the MCH rating scale (Gawron, 2019).

2.3.2 Situation Awareness

Situation awareness is defined as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” (Endsley & Garland, 2000). Having the appropriate situation awareness is one of the key elements to the success of NPP operations (Arigi et al., 2019). NPP operators and personnel should be able to recognize the current situation of the plant and monitor and control the changes throughout the states of a rapidly changing NPP. Like this, situation awareness is more than a simple memory check of what the operator knows or does not know. To date, there have been a lot of efforts to experimentally analyze situation awareness or develop a way to properly measure it (Liu et al., 2021; Hogg et al., 1995; Wei et al., 2013; Marusich et al., 2016; Gugerty and Tirre, 1996; Yan, Yao, and Tran, 2021).

In this study, the Situation Awareness Rating Technique (SART) (Taylor, 2017) was used to estimate participants’ situation awareness levels. Figure 3 shows the SART questionnaire. SART focuses on measuring participants’ knowledge in three areas: demand (questions 1–3), supply (questions 4–7), and understanding (questions 8–10). A composite SART score is calculated via the following formula:

- Overall SART = sum of scores in the understanding area – (sum of scores in the demand area – sum of scores in the supply area).

2.3.3 Error

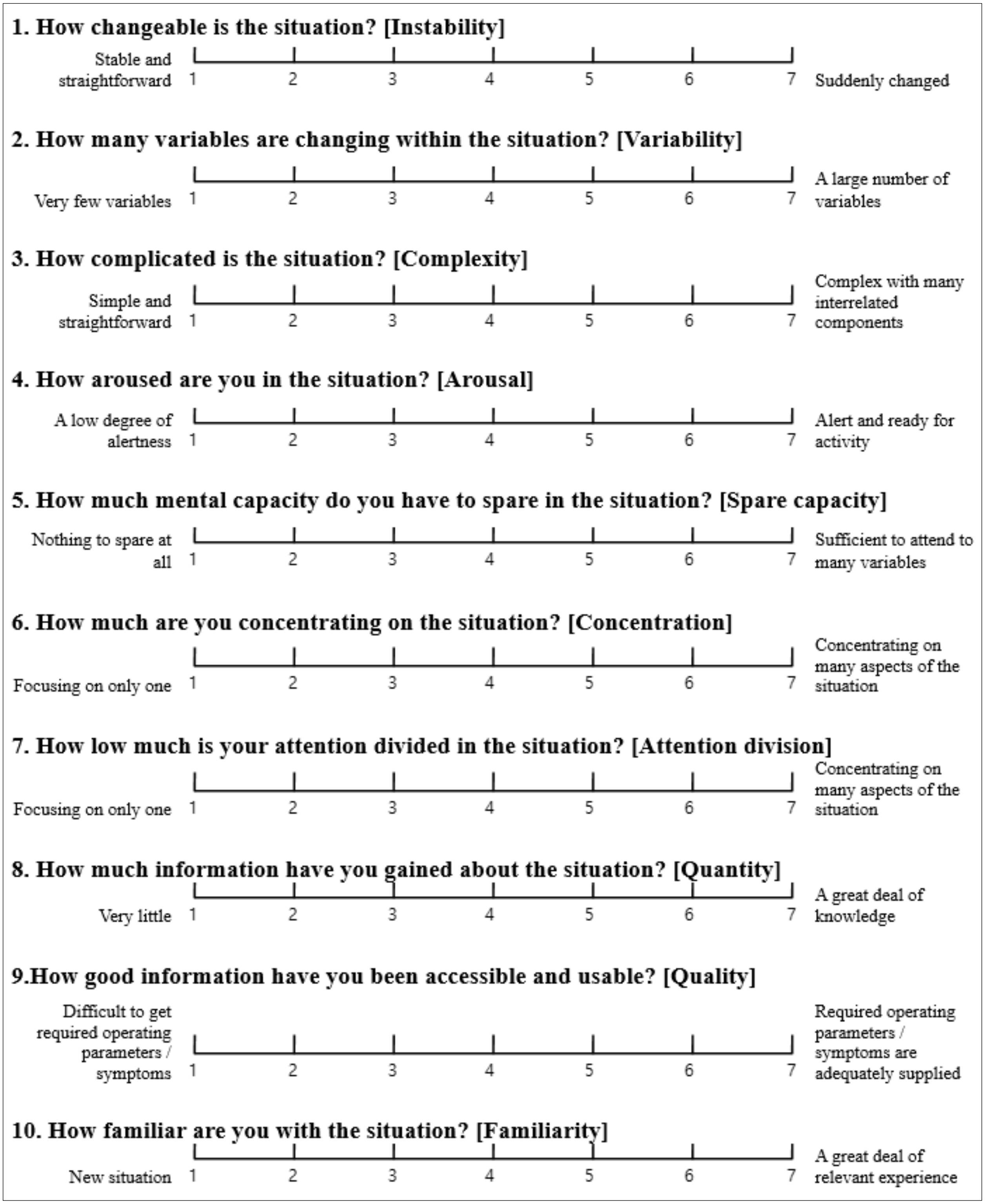
In HRA, there have been a lot of efforts to collect errors from simulator studies as well as produce quantitative HRA data such as basic human error probabilities or performance shaping factor multipliers through statistical analyses (Jang et al., 2013; Kim et al., 2015; Jang et al., 2016; Kim, 2020; Jung et al., 2020). Generally, error is defined as the deviation of task performance from procedures, because most activities in NPPs are performed based on procedures. There are largely two types of errors: Error of Omission (EOC) and Error of Commission (EOC). EOO refers to the error caused by omitting a task, while EOC is an unintended error such as a selection error (e.g., selecting the wrong control) or sequence error (e.g., conducting tasks in the wrong order).

In this study, we used the rules and analysis categories suggested in the HuREX study (Jung et al., 2020). Details on the error analysis are found in (Park et al., 2022a). For the error rate, it is calculated by dividing the number of errors into the number of tasks in a scenario.

2.3.4 Time to Completion

Time to completion is important information in HRA. Most HRA methods use it to estimate HEPs or determine task failure by the time required and time window (Swain & Guttmann, 1983; Parry et al., 1992; Gertman et al., 2005). Accordingly, the time data obtained from experiments has been analyzed in many HRA research projects (Park, Lee, Jung and Kim, 2017; Park, Jung and Kim, 2020; Boring et al., 2021).

This human performance measure encompasses the average time to complete a step, instruction, or task. Each procedure consists of steps that are themselves composed of instructions, which generally include one or more task(s). Figure 4 shows an example of the procedure format. “Perform core cooling using Bypass Valve” is the step, “Adjust the Bypass Valve properly to keep the core temperature below 400°C” is an instruction, and “Open the Bypass Valve by 10.0%” is a task. Accordingly, the average time to complete a step, instruction or task is calculated by dividing the time to complete a scenario into the number of steps, instructions, or tasks in the scenario.



1. How changeable is the situation? [Instability]

Stable and straightforward 1 2 3 4 5 6 7 Suddenly changed

2. How many variables are changing within the situation? [Variability]

Very few variables 1 2 3 4 5 6 7 A large number of variables

3. How complicated is the situation? [Complexity]

Simple and straightforward 1 2 3 4 5 6 7 Complex with many interrelated components

4. How aroused are you in the situation? [Arousal]

A low degree of alertness 1 2 3 4 5 6 7 Alert and ready for activity

5. How much mental capacity do you have to spare in the situation? [Spare capacity]

Nothing to spare at all 1 2 3 4 5 6 7 Sufficient to attend to many variables

6. How much are you concentrating on the situation? [Concentration]

Focusing on only one 1 2 3 4 5 6 7 Concentrating on many aspects of the situation

7. How low much is your attention divided in the situation? [Attention division]

Focusing on only one 1 2 3 4 5 6 7 Concentrating on many aspects of the situation

8. How much information have you gained about the situation? [Quantity]

Very little 1 2 3 4 5 6 7 A great deal of knowledge

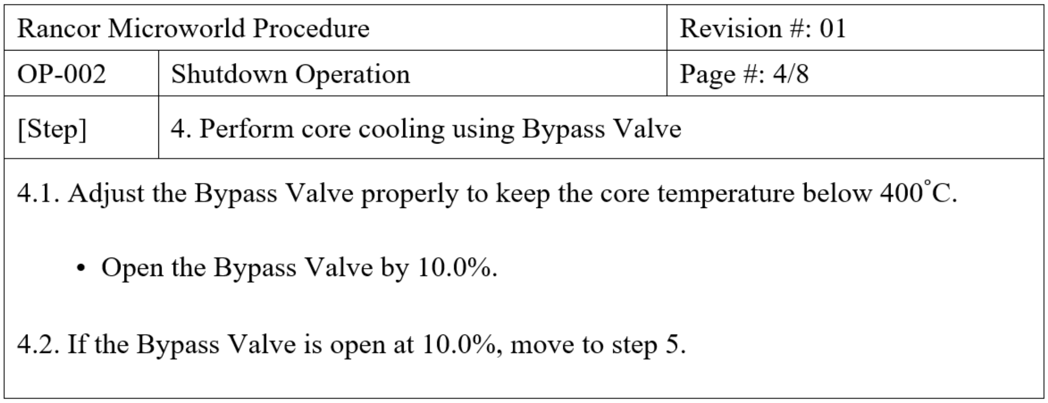
9.How good information have you been accessible and usable? [Quality]

Difficult to get required operating parameters / symptoms 1 2 3 4 5 6 7 Required operating parameters / symptoms are adequately supplied

10. How familiar are you with the situation? [Familiarity]

New situation 1 2 3 4 5 6 7 A great deal of relevant experience

Figure 3. Questionnaire for the SART rating scale (Taylor, 2017).



Rancor Microworld Procedure		Revision #: 01
OP-002	Shutdown Operation	Page #: 4/8
[Step]	4. Perform core cooling using Bypass Valve	
4.1. Adjust the Bypass Valve properly to keep the core temperature below 400˚C. • Open the Bypass Valve by 10.0%. 4.2. If the Bypass Valve is open at 10.0%, move to step 5.		

Figure 4. Example of the procedure format.

2.3.5 Eye Movements

Eye movement has been popularly measured in human factors research because it is easy to estimate through eye-trackers as well as generate useful qualitative and quantitative data required for understanding differences across experimental conditions (Martins and Carvalho, 2015; Dzeng, Lin and Fang, 2016; Takacs and Bus, 2018; Tian et al., 2022).

In this study, eye movement measures such as (1) eye fixation count per task, (2) eye fixation duration per task, (3) blink rate, and (4) heatmap over AOI were estimated using an eye-tracking system (see Section 2.4.2). These are common eye tracking measures for estimating human performance in human factors research.

2.3.6 Number of Manipulations

Number of manipulations refers to how many times the participants manipulated simulator interfaces. Manipulations involve turning pumps, valves, and sliders on/off, mostly via mouse clicks in Rancor. As a distinguishing feature in digital or computerized main control rooms, the manipulations are also called as secondary tasks or interface management tasks. This human performance measure was considered in the lead author’s previous experimental studies (Park, Lee, Jung and Kim, 2017; Park, Jung and Kim, 2020).

Manipulations were generally counted by referencing the log data generated from the simulator logs (see Figure 5). This study used the number of manipulations per task and per scenario completion time as the representative measures.

```
{"EventType":"Simulation", "Tag":"Start", "Time":0, "RX":98.2692294477296, "MW":54.5491744984957}
{"EventType":"Alarm", "Tag":"AtmosDumpActive", "State":"Cleared", "Unit":1, "Time":1.0093952, "RX":98.2286081374419, "MW":54.5491744984957}
{"EventType":"Alarm", "Tag":"CoreHighTemp", "State":"Alarmed", "Unit":1, "Time":2.0098807, "RX":98.2286081374419, "MW":54.840732182516}
{"EventType":"ControlAction", "Tag":"SGSGAin", "Mode":"manual", "Value":"0.461748397052678", "Time":2.0098807, "RX":98.0600762918027, "MW":54.7726059086581}
{"EventType":"ControlAction", "Tag":"SGSGBin", "Mode":"manual", "Value":"0.461748397052678", "Time":2.0098807, "RX":98.0600762918027, "MW":54.7726059086581}
{"EventType":"ControlAction", "Tag":"SGSGAin", "Mode":"manual", "Value":"0.461176073053509", "Time":4.0130494, "RX":98.9049985703981, "MW":54.8454880458299}
{"EventType":"ControlAction", "Tag":"SGSGBin", "Mode":"manual", "Value":"0.461176073053509", "Time":4.0130494, "RX":98.9049985703981, "MW":54.8454880458299}
{"EventType":"ControlAction", "Tag":"SGSGAin", "Mode":"manual", "Value":"0.46095038168767", "Time":6.0079151, "RX":98.9794085666886, "MW":54.9782853881572}
{"EventType":"ControlAction", "Tag":"SGSGBin", "Mode":"manual", "Value":"0.46095038168767", "Time":6.0079151, "RX":98.9794085666886, "MW":54.9782853881572}
{"EventType":"ControlAction", "Tag":"SGSGAin", "Mode":"manual", "Value":"0.461876577398047", "Time":8.0051575, "RX":99.0464760967634, "MW":55.1216107106604}
{"EventType":"ControlAction", "Tag":"SGSGBin", "Mode":"manual", "Value":"0.461876577398047", "Time":8.0051575, "RX":99.0464760967634, "MW":55.1216107106604}
{"EventType":"ControlAction", "Tag":"SGSGAin", "Mode":"manual", "Value":"0.462045661847654", "Time":10.0004206, "RX":99.1073924451002, "MW":55.2599619518473}
{"EventType":"ControlAction", "Tag":"SGSGBin", "Mode":"manual", "Value":"0.462045661847654", "Time":10.0004206, "RX":99.1073924451002, "MW":55.2599619518473}
{"EventType":"ControlAction", "Tag":"SGSGAin", "Mode":"manual", "Value":"0.461287502157667", "Time":12.0006931, "RX":99.1637561689002, "MW":55.3498612416287}
{"EventType":"ControlAction", "Tag":"SGSGBin", "Mode":"manual", "Value":"0.461287502157667", "Time":12.0006931, "RX":99.1637561689002, "MW":55.3498612416287}
{"EventType":"ControlAction", "Tag":"SGSGAin", "Mode":"manual", "Value":"0.462277544506157", "Time":13.9937138, "RX":99.2153265615288, "MW":55.3921688214245}
```

Figure 5. Example of log data generated from a simulator.

2.4 Apparatus

The experiment in this study used Rancor, an eye-tracking system, a video recording system, and a procedure system.

2.4.1 A Representative Simplified Simulator: Rancor

Rancor is a representative simplified simulator developed by INL and University of Idaho (Ulrich, 2017). It has been used to examine theoretical and practical design concepts and to provide a graphical user interface that enables researchers to create generic process control systems. Although Rancor has simplified systems, procedures, and scenarios, it characterizes major operational strategies in NPPs. Figure 6 shows the simulator interface. The current version is designed to be operated by one participant at a time.

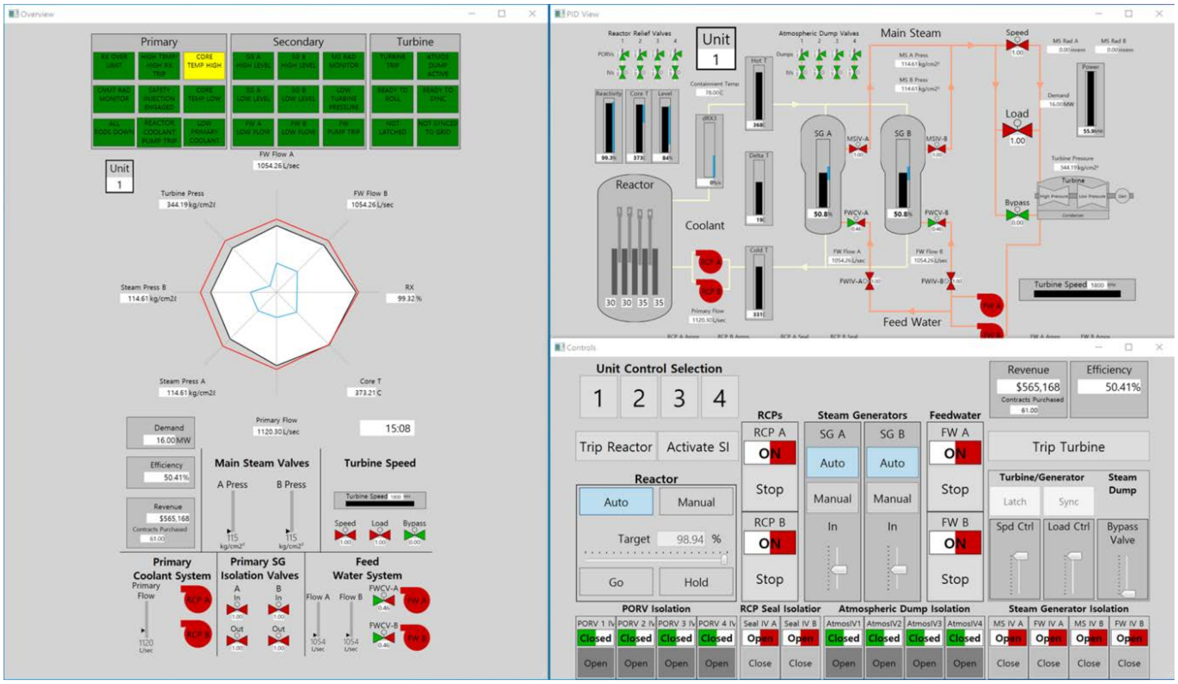


Figure 6. The Rancor Microworld Simulator interface.

2.4.2 Eye-tracking System

The Tobii Pro Glasses 2 eye-tracking system (Niehorster, Hessels, & Benjamins, 2020) was used to acquire eye movement data in this experiment. This system consists of:

- *A wearable eye tracker* – a head unit containing a high-definition scene camera, eye-tracking sensors, infrared illuminators, a micro high-definition multimedia interface (HDMI) connector, a cable guide, a protective lens, and a nose pad.
- *A recording unit* – connected to the head unit via an HDMI cable and consisting of a standard-definition memory card, battery compartment, HDMI/USB connectors, power/connection status light-emitting diodes (LEDs), and a belt clip.
- *Controller software* – used for managing participants, controlling the eye tracker, and viewing both real-time and recorded eye-tracking data.
- *Analysis software* – used for other eye-tracking analyses besides qualitative analyses of eye-tracking session replays.

2.4.3 Video Recorder System

A video recording system was set up to trace the actions of each participant. The records were mainly used for counting the number of tasks performed, analyzing and counting errors, and estimating the time to completion for each step. Two video cameras were placed on tripods at strategic positions around the participant. The first camera primarily captured the Rancor interface throughout the experiment, while the second one primarily captured the procedures being implemented by the participant during the experiment. In addition, the audio was saved to track each participant's progression through the procedures or scenarios.

2.4.4 Procedures

Paper-based procedures were used in the experiment. The procedures were modeled based on Westinghouse pressurized water reactor procedures, but tailored and simplified for Rancor. Participants used paper-based procedures when operating Rancor. The procedures guided participants on how to perform specific tasks and how to respond to alarm conditions.

2.4.5 Participants

A total of 40 participants (20 operators and 20 students) participated in the experiment. Most of the operators were licensed professional operators currently employed at NPPs. They were all operators on shift (i.e., a shift supervisor, shift technical advisor, reactor operator, or turbine operator [TO]) or instructors at the training center—save for one operator who lacked a current license. However, this individual was an acting TO with 14 years of experience in NPP operations. Thus, the knowledge and experience of the acting-but-unlicensed TO were considered equivalent to that of a licensed operator. As for the students, they were all undergraduate seniors or graduate students from the Department of Nuclear Engineering at Chosun University. They were knowledgeable about NPP systems and operations, having completed a significant portion of their coursework, which included courses such as "Introduction to Nuclear Engineering," "Reactor Theory," "Reactor Control," and "Simulator Operation." Each participant completed an informed consent according to an approved Institutional Review Board (IRB) in place at Chosun University and Idaho National Laboratory.

2.5 Training

The training material prepared for each participant included the reason for the experiment, a description of Rancor and its systems, potential scenarios, questionnaires, and practice sessions with Rancor. The students underwent a test scenario to provide further assurance of their proficiency with the system and to boost their confidence. Training for each student lasted about 3 hours, while training for each operator lasted approximately 2 hours.

2.6 Data Acquisition

Table 4 summarizes the data acquisition methods and collectible items pertaining to the human performance data collection.

Table 4. Summary of data acquisition methods, items collected, and human performance measures.

Method	Items Collected	Human Performance Measures
Questionnaire	General information on each participant, MCH workload ratings (see Figure 2), and SART scores (see Figure 3)	Situation awareness, and workload
Eye-Tracker	Eye fixation counts, eye fixation duration, blink counts, and heatmap over AOIs	Eye movements
Video/Audio Recording	Number of errors, and time to completion for each step, instruction or task	Error, and time to completion
Simulator Log	Number of manipulations	Number of manipulations

2.7 Experimental Procedure

The experimental procedure consists of three steps as below.

- *Step #1: Eye-tracker calibration and recording equipment check* – Staff needs to calibrate the eye-tracker and check if video and audio recording equipment is ready to record.
- *Step #2: Experiment* – Participants run six scenarios, while staff keeps monitoring if the eye-tracker and recordings work well. Note that only six scenarios of the ten total presented in Table 2 are used for each participant. Three randomly chosen normal operational and three randomly chosen abnormal operational scenarios are selected.
- *Step #3: Data acquisition* – Staff collects and saves the video and audio recordings, the simulator log, and the eye-tracker data, while participants fill out the MCH and SART questionnaires and then give them to the staff.

3. INVESTIGATION INTO DIFFERENCES IN HUMAN PERFORMANCE

This section primarily explores the analytical results regarding which differences in the human performance measures can be traced back to the two independent variables (i.e., participant type and event class), as well as the human performance measures that correspond to participant type (i.e., operator or student). Two statistical analysis methods (i.e., ANOVA testing and correlation analysis) were applied to investigate these differences.

3.1 Analysis of Each Human Performance Measure

Table 5 summarizes the average values of those human performance measures that depend on participant type and event class. An ANOVA test was performed on each human performance measure in order to assess the amount of variability between the group means (in the context of variation within groups), thus revealing whether these differences in means are statistically significant. Table 6 summarizes the results of the ANOVA tests (p-values). The significance of the mean differences from the ANOVA tests is also marked in Table 5.

The overall ANOVA results reveal that several human performance measures significantly differed depending on participant type and event class. The MCH scores, SART scores, error rates, average time to complete an instruction, and blink rates were all statistically different when comparing the operator and student groups. All measures for time to completion, eye movements (excluding blink rate), and number of manipulations showed statistical differences between the non-event and event scenario classes. In addition, the measures for time to completion and eye movements (excluding blink rate) revealed statistically significant interactions between participant type and event class. Details on these results are presented in the following subsections.

Table 5. Average values for the human performance measures.

Human Performance Characteristic	Measure	Average Value	Independent Variable			
			Participant Type		Event Class	
			Student	Operator	Non-event	Event
Workload	MCH	3.633	4.150**	3.117**	3.775	3.492
Situation Awareness	SART	19.192	18.175*	20.208*	19.308	19.075
Error	Error rate [#/task]	0.008	0.009*	0.006*	0.007	0.008
Time to Completion	Average time to complete a step [sec]	34.381	34.149	34.610	47.486**	21.055**
	Average time to complete an instruction [sec]	5.941	6.219*	5.662*	6.607**	5.275**
	Average time to complete a task [sec]	4.965	5.146	4.785	5.875**	4.055**
Eye Movements	Eye fixation count per task [#]	4.564	4.632	4.495	5.672**	3.455**
	Eye fixation duration per task [sec]	1.604	1.623	1.584	2.109**	1.098**
	Blink rate [#/sec]	0.338	0.397**	0.280**	0.326	0.350
Number of Manipulations	Number of manipulations per task [#]	0.358	0.362	0.355	0.461**	0.256**
	Number of manipulations per scenario completion time [#/sec]	0.076	0.073	0.079	0.083**	0.069**

*Shows a statistical difference within the 95% confidence level ($p < 0.05$).

**Shows a statistical difference within the 99% confidence level ($p < 0.01$).

Table 6. ANOVA test result (p-value).

Human Performance Characteristic	**Human Performance Measure**	**Independent Variable**		
		Participant Type	**Event Class**	**Participant Type * Event Class**
Workload	MCH	0.001**	0.350	0.700
Situation Awareness	SART	0.027*	0.798	0.597
Error	Error rate	0.046*	0.430	0.686
Time to Completion	Average time to complete a step	0.801	0.000**	0.002**
	Average time to complete an instruction	0.047*	0.000**	0.000**
	Average time to complete a task	0.148	0.000**	0.000**
Eye Movements	Eye fixation count per task	0.706	0.000**	0.016*
	Eye fixation duration per task	0.787	0.000**	0.005**
	Blink rate	0.008**	0.595	0.863
Number of Manipulations	Number of manipulations per task	0.727	0.000**	0.053
	Number of manipulations per scenario completion time	0.222	0.001**	0.139

* Showed a statistical difference (within the 95% confidence level) with respect to the independent variable ($p < 0.05$).

**Showed a statistical difference (within the 99% confidence level) with respect to the independent variable ($p < 0.01$).

3.1.1 Workload

As depicted in Table 6, the MCH scores showed a significant difference in regard to participant type. As seen in Table 5, the student group had a higher average MCH score than the operator group: 4.150 vs. 3.117, respectively. Based on the questionnaire for the MCH rating scale (see Figure 2), the student group's average workload approximately corresponded to the "inconvenient" level of difficulty (indicating a moderately high mental workload), whereas the operator group's average workload was close to the "adequate" difficulty level (indicating an adequate mental workload). Thus, the students' higher MCH scores indicated a heavier workload. Event class type, however, yielded no statistical difference in terms of MCH scores.

Figure 7 shows the trend of the overall MCH results for the student and operator groups during both non-event and event scenarios, indicating that the student group experienced a higher workload than the operator group, regardless of event class.

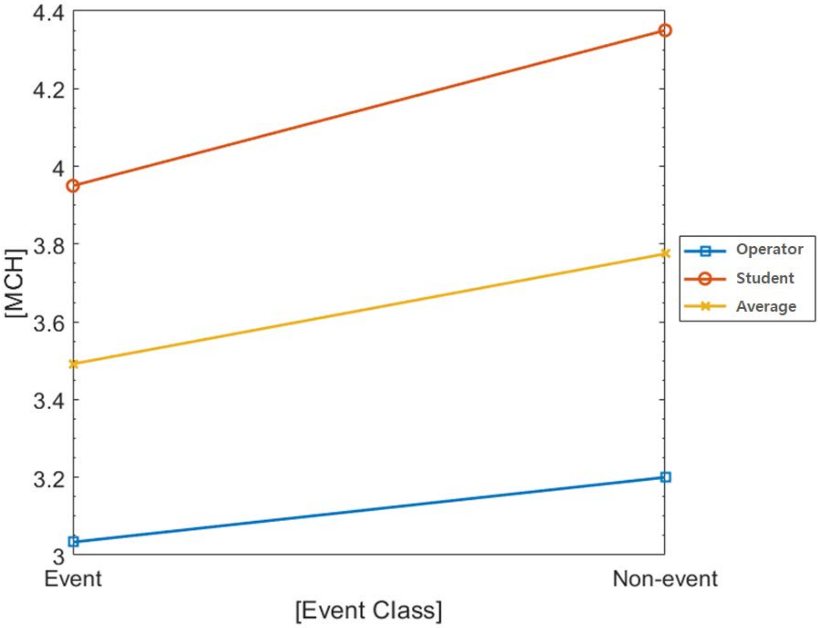


Figure 7. Trend of the overall MCH results.

3.1.2 Situation Awareness

The SART scores only showed a statistical difference in regard to participant type, as shown in the ANOVA test results (see Table 6). The operator group reported higher values than the student group: 20.208 vs. 18.175, respectively (see Table 5). As per the formula (introduced in Section 2.3.2) for calculating overall SART scores, a higher SART score indicates better situation awareness.

Figure 8 represents the trend of the overall SART results for the student and operator groups during both non-event and event scenarios, indicating that the operator group had better situation awareness than the student group, regardless of event class.

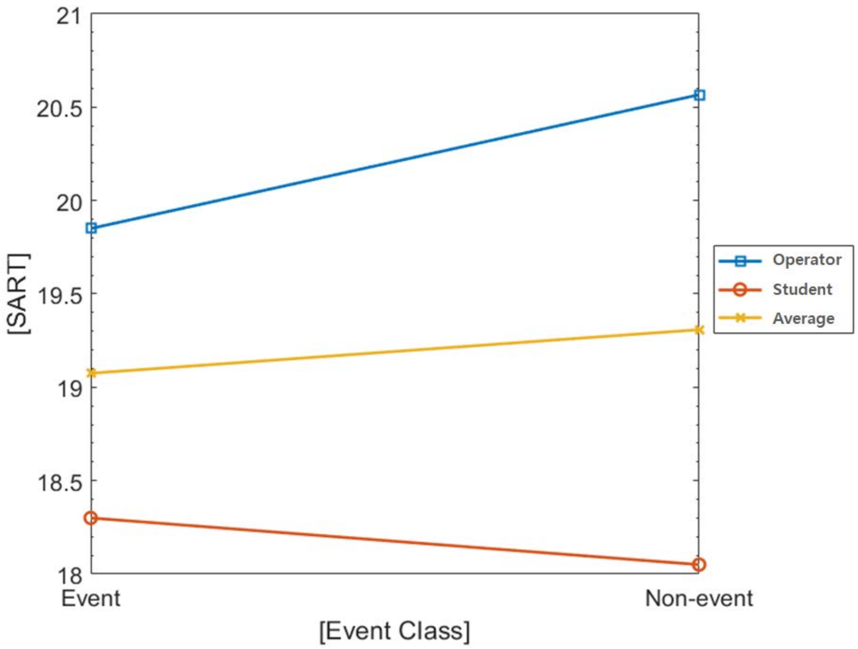


Figure 8. Trend of the overall SART results.

3.1.3 Error

The error rate statistically differed in regard to participant type, but not in regard to event class (see Table 6). The error rate for the student group was 1.5x higher than for the operator group: 9.0e-3 vs. 6.0e-3, respectively (see Table 5). Figure 9 shows the trend of the overall error rates for the student and operator groups during both non-event and event scenarios, demonstrating that the operator group achieved better performance (i.e., lower error rates) than the student group, regardless of event class.

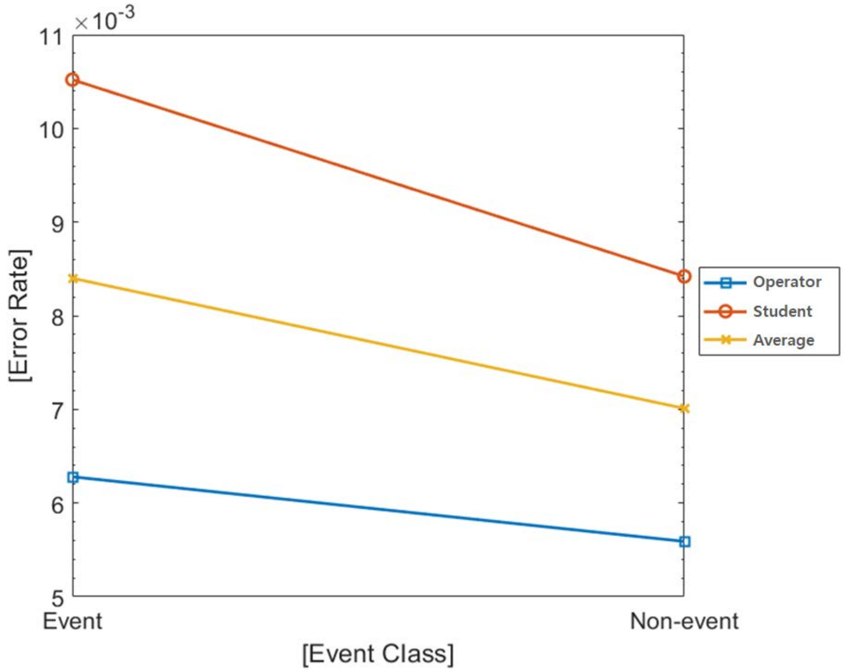


Figure 9. Trend of the overall error rate.

3.1.4 Time to Completion

All human performance measures for time to completion showed significant differences in regard to event class, though there was no statistical difference in regard to participant type, apart from average time to complete an instruction (see Table 6). If we look at the average time to complete a task across both event classes, event scenarios reflected lower average times than non-event scenarios: 4.055 vs. 5.875 seconds, respectively (see Table 5).

ANOVA testing revealed that the interaction between participant type and event class is indeed statistically significant. Figure 10 shows the trend of the overall average time to complete a task for the student and operator groups during both non-event and event scenarios. As shown in Table 5, the student group generally took longer than the operator group to complete a task: 5.146 vs. 4.785 seconds, respectively. However, Figure 10 shows that the operators took longer than the students in non-event scenarios: 6.183 vs. 5.567 seconds, respectively.

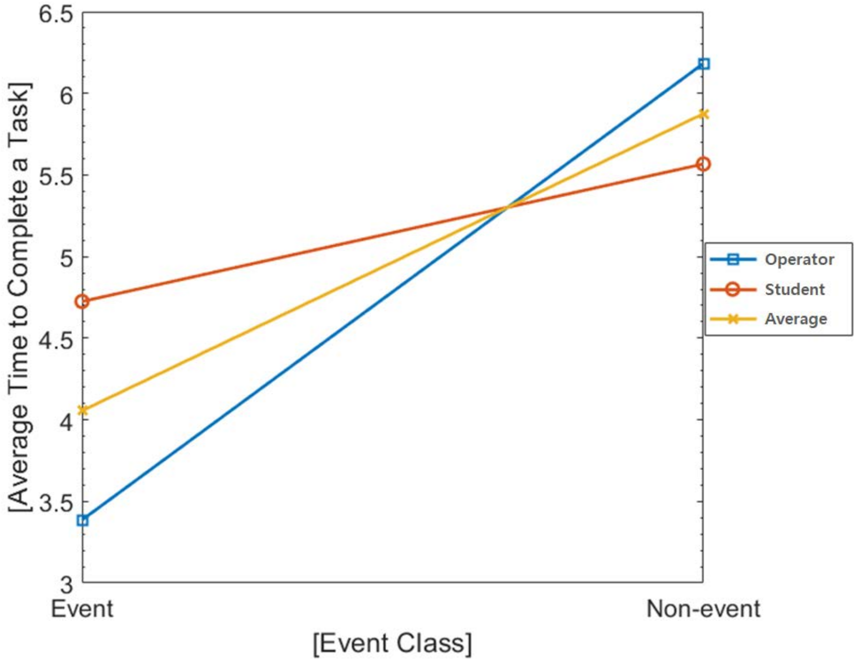


Figure 10. Trend of the overall average time to complete a task.

3.1.5 Eye Movements

The ANOVA results for eye movements are categorized into two different types. First, eye fixation count and duration per task revealed statistically significant differences in regard to event class but not participant type (see Table 6). For event scenarios, the average eye fixation count per task is lower than for non-event scenarios: 3.455 vs. 5.672, respectively (see Table 5). There are also interactions between participant type and event class. Figure 11 shows the trend of the overall eye fixation count per task for both the student and operator groups during non-event and event scenarios. Although the student group exhibited more fixations than the operator group (4.632 vs. 4.495, respectively [see Table 5]), Figure 11 indicates that in non-event scenarios, the operator group had higher fixation counts than the student group: 6.042 vs. 5.3025, respectively. Furthermore, the blink rate only showed a significant difference with respect to participant type (see Table 6). The blink rate for the student group was higher than for the operator group: 0.397 vs. 0.280, respectively (see Table 5). Figure 12 shows the trend of the overall blink rate for the student and operator groups during both non-event and event scenarios, revealing that the student group had a higher blink rate than the operator group, regardless of event class.

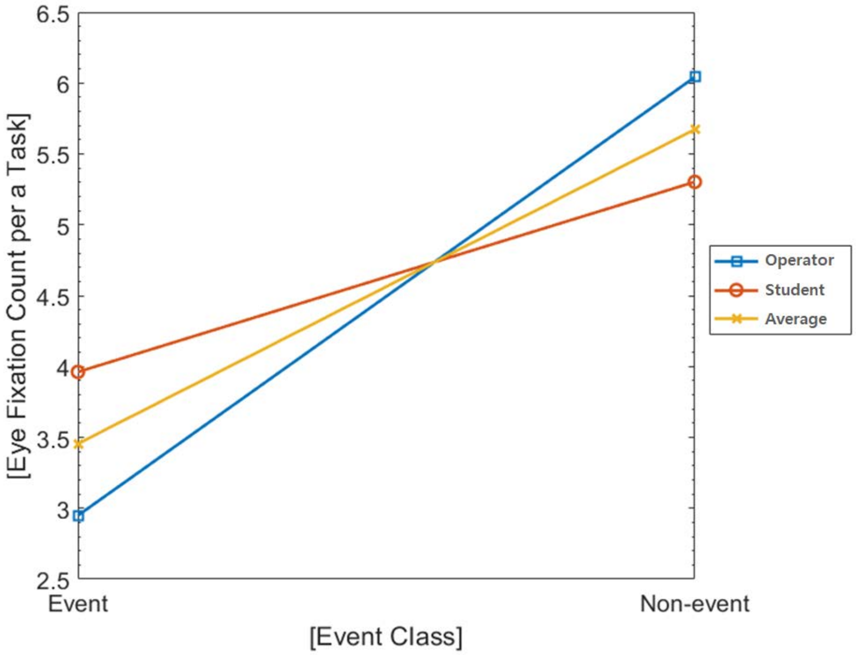


Figure 11. Overall eye fixation count per task.

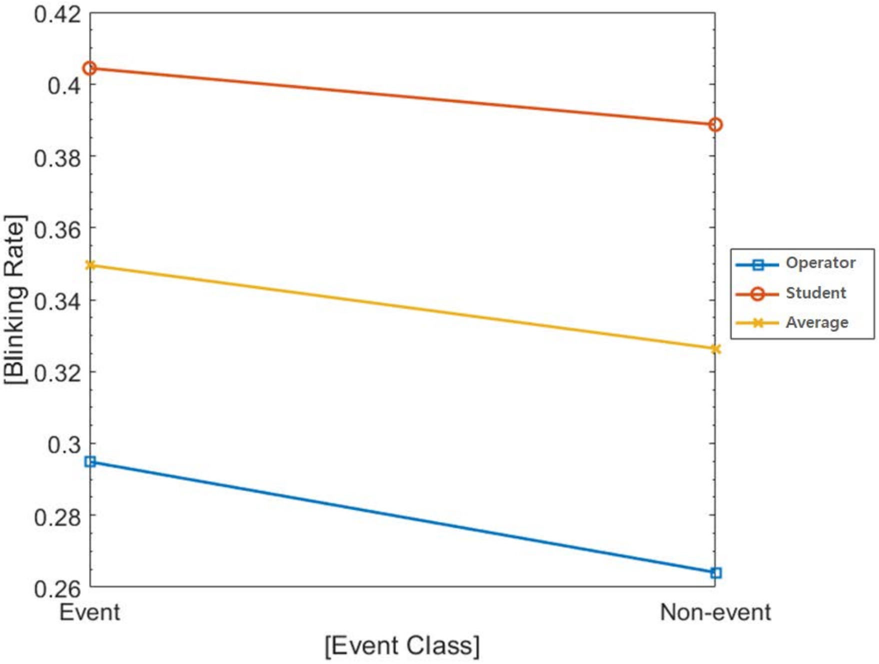


Figure 12. Overall blink rate.

This study also collected heatmap information for AOIs via the eye-tracking system. Figure 13 and Figure 14 show heatmaps for the operator and student groups during non-event and event scenarios, respectively. The eye fixation counts are used in assigning colors to the heatmap, as shown in Figure 15. The red areas are those that received the highest number of fixations, whereas the green areas are those that received the least number of fixations.

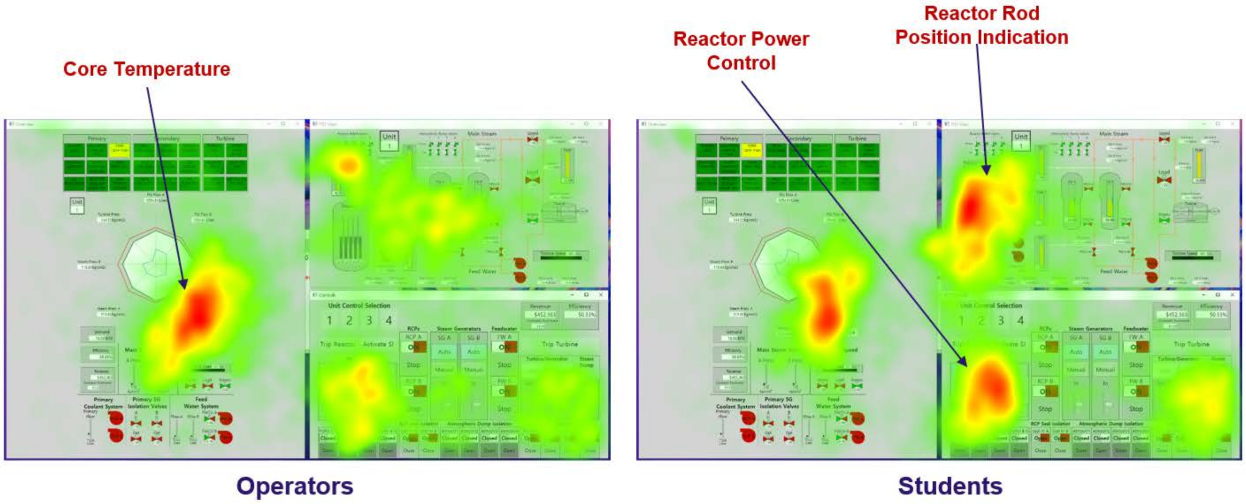


Figure 13. Heatmap for the operator and student groups in non-event scenarios.

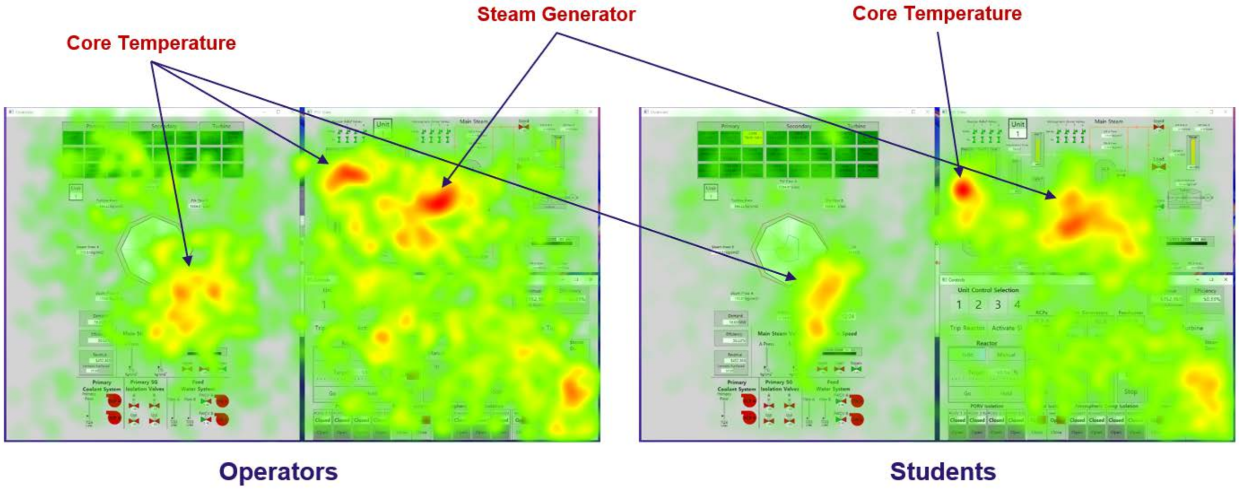


Figure 14. Heatmap for the operator and student groups in event scenarios.

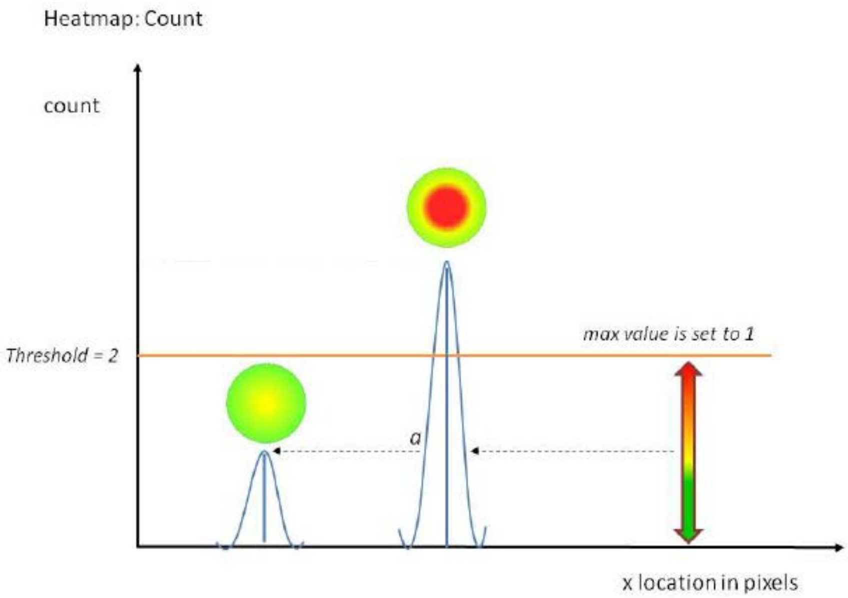


Figure 15. Illustration of how colors are assigned based on eye fixation counts (Niehorster et al., 2020).

3.1.6 Number of Manipulations

All human performance measures for number of manipulations showed significant differences in regard to event class, but no statistical difference regarding participant type (see Table 6). Comparing the number of manipulations per task showed that the non-event scenario group performed more manipulations per task than the event scenario group: 0.461 vs. 0.256, respectively (see Table 5). Figure 16 shows the trend of the overall number of manipulations per task for the student and operator groups during both non-event and event scenarios, indicating that non-event scenarios require more manipulations than event scenarios, regardless of participant type.

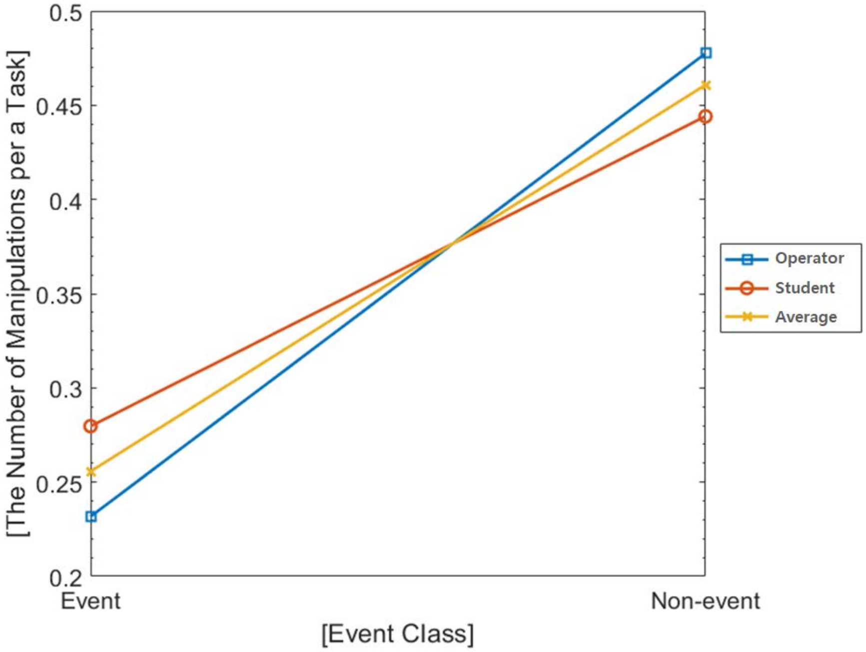


Figure 16. Trend of the overall number of manipulations per task.

3.2 Analysis of the Human Performance Relationship between Operators and Students

Through correlation analysis, this section investigates how human performance measures correlate with each another, and how they correlate within each individual participant group. The seven human performance measures, which represented the six human performance characteristics, were correlated in the manner described below. For eye movements, eye fixation counts per task and blink rate were selected, since they had different characteristics in the ANOVA test, as detailed in Section 3.1.5.

- Workload: MCH
- Situation Awareness: SART
- Error: error rate
- Time to Completion: average time to complete a task
- Eye Movements (two measures): eye fixation count per task and blink rate
- Number of Manipulations: number of manipulations per task.

Table 7 shows the correlation matrix of coefficient values for the seven human performance measures, collapsed across participant type to include both operators and students. The grey-colored cells indicate a correlation coefficient above 0.3 (i.e., moderate or strong relationship), thus denoting statistical significance. For all participants, the relationships between (1) MCH and SART, (2) MCH and error rate, (3) time to completion and eye fixation count per task, (4) number of manipulations and time to completion per task, and (5) number of manipulations and eye fixation count per task all show moderate or strong correlations.

Table 7. Correlation analysis for all participants.

	MCH	**SART**	**Error rate**	**Time to completion per task**	**Eye fixation count per task**	**Blink rate**	**Number of manipulations per task**
MCH	1						
SART	-0.548**	1					
Error rate	0.435**	-0.247**	1				
Time to completion per task	0.097	-0.069	-0.015	1			
Eye fixation count per task	0.040	-0.009	0.006	0.743**	1		
Blink rate	0.117	-0.097	0.028	0.134*	-0.158*	1	
Number of manipulations per task	-0.122	0.128*	-0.145*	0.481**	0.477**	0.045	1

*Shows a statistical difference within the 95% confidence level ($p < 0.05$).

**Shows a statistical difference within the 99% confidence level ($p < 0.01$).

Table 8. Correlation analysis for the operator group.

	MCH	**SART**	**Error rate**	**Time to completion per task**	**Eye fixation count per task**	**Blink rate**	**Number of manipulations per task**
MCH	1						
SART	-0.375**	1					
Error rate	0.168	0.028	1				
Time to completion per task	0.004	0.025	-0.070	1			
Eye fixation count per task	-0.005	0.127	-0.053	0.780**	1		
Blink rate	0.024	-0.121	-0.109	0.194*	-0.058	1	
Number of manipulations per task	-0.048	0.095	-0.086	0.605**	0.571**	-0.010	1

*Shows a statistical difference within the 95% confidence level ($p < 0.05$).

**Shows a statistical difference within the 99% confidence level ($p < 0.01$).

Table 9. Correlation analysis for the student group.

	MCH	**SART**	**Error rate**	**Time to completion per task**	**Eye fixation count per task**	**Blink rate**	**Number of manipulations per task**
MCH	1						
SART	-0.629**	1					
Error rate	0.485**	-0.362**	1				
Time to completion per task	0.171	-0.184*	0.004	1			
Eye fixation count per task	0.096	-0.131	-0.003	0.606**	1		
Blink rate	0.107	-0.044	0.053	0.053	-0.209**	1	
Number of manipulations per task	-0.190*	0.169	-0.202*	0.274**	0.486**	0.088	1

*Shows a statistical difference within the 95% confidence level ($p < 0.05$).

**Shows a statistical difference within the 99% confidence level ($p < 0.01$).

The results of the correlation analysis specifically conducted for operators are shown in Table 8. Again, the grey cells reflect moderate or strong correlations at a statistically significant level. For the operator group, the significant correlations were between (1) MCH and SART, (2) time to completion and eye fixation count per task, (3) time to completion and number of manipulations per task, and (4) eye fixation count and number of manipulations per task.

Table 9 shows the results of the correlation analysis conducted for the students, with the highlighted cells showing moderate or strong correlations between (1) MCH, SART, and the error rate; (2) time to completion per task and eye fixation count per task; and (3) eye fixation count per task and number of manipulations per task.

Figure 17 shows scatter plots of the relationship between the MCH rating and the SART scores, both with the operator and student groups combined and with them separated. Both participant types show significant negative linear correlations for MCH and SART, meaning that situation awareness generally decreases with increased workload. The plots in which the student group was separated out from the operator group show that the correlation for the student group was stronger than for the operator group.

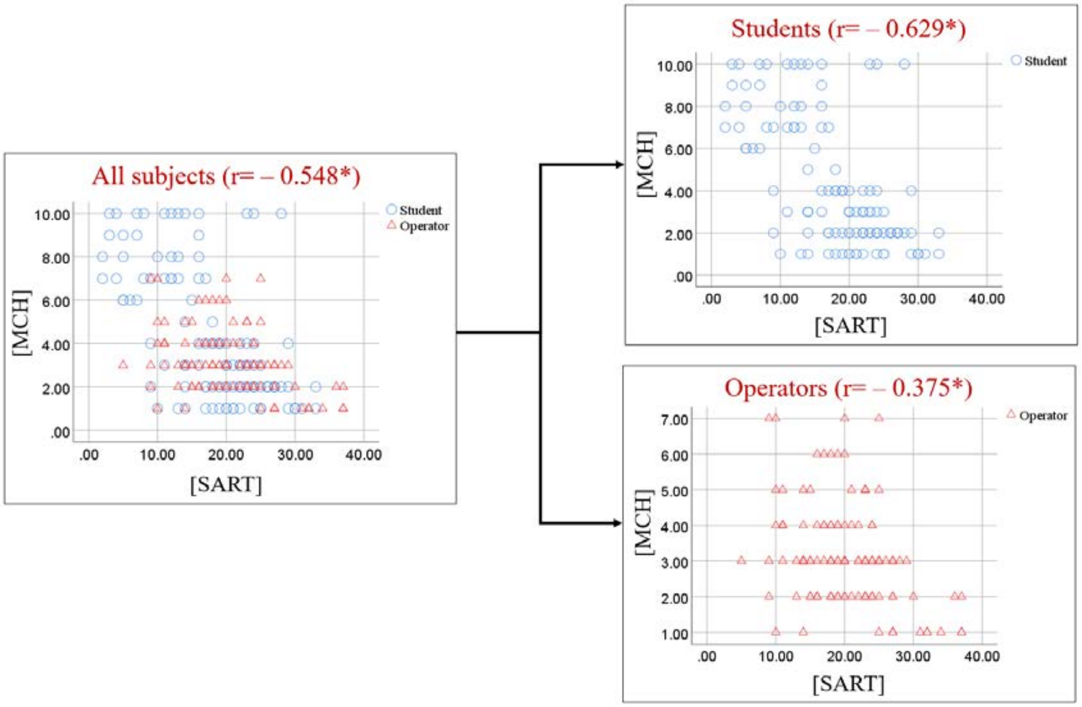


Figure 17. Scatter plot of the relationship between the MCH rating and the SART scores, both with the operator and student groups combined and with them separated.

Another interesting result is depicted in Figure 18. Collectively, both participant types show moderately positive linear correlations for error rates and workload. However, when separated out, only the student group shows a statistically significant positive correlation between error rate and MCH score; the operator group does not.

Figure 19 also shows a highlight from the correlation analysis. Collectively, both participant types show small negative linear correlations between error rates and SART scores. However, when separated out, the student group shows a moderate linear correlation between error rate and SART value, while no statistically significant correlation exists between operators' error rates and SART values.

Figure 20 shows the relationship among time to completion, number of manipulations, and eye movements, along with the correlation coefficients. Both participant types indicate moderate or strong positive linear correlations among the human performance measures. On the other hand, when the groups are separated, the correlation coefficient between time to completion and number of manipulations per task in the student group greatly decreases.

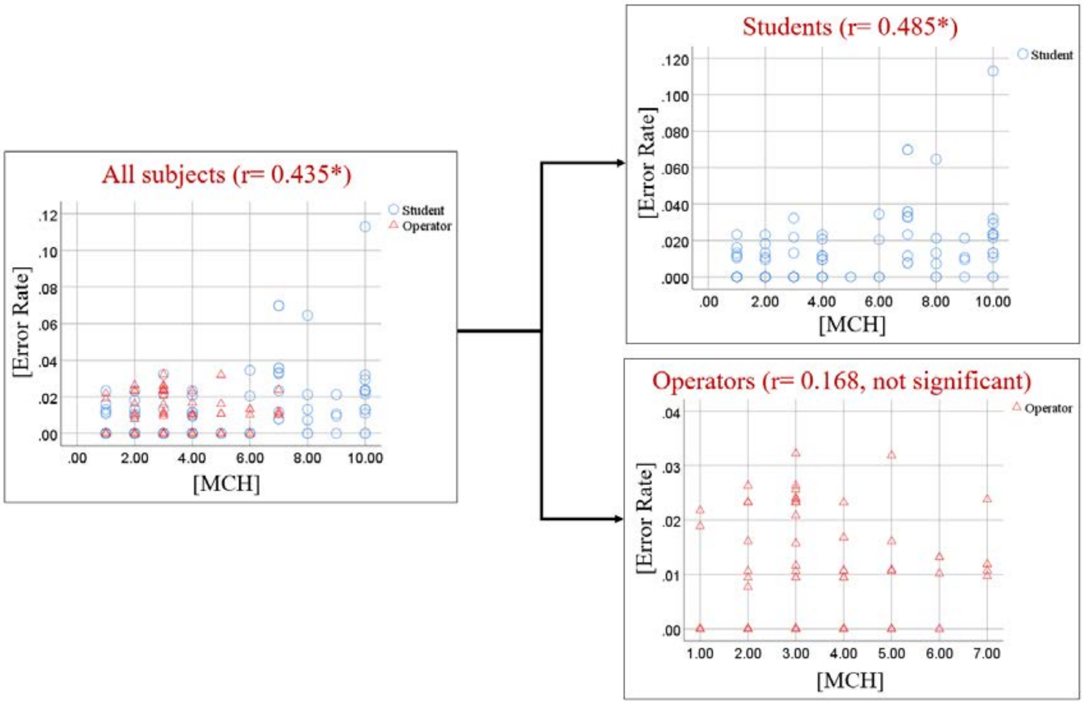


Figure 18. Scatter plot of the relationship between the MCH rating and the error, both with the operator and student groups combined and with them separated.

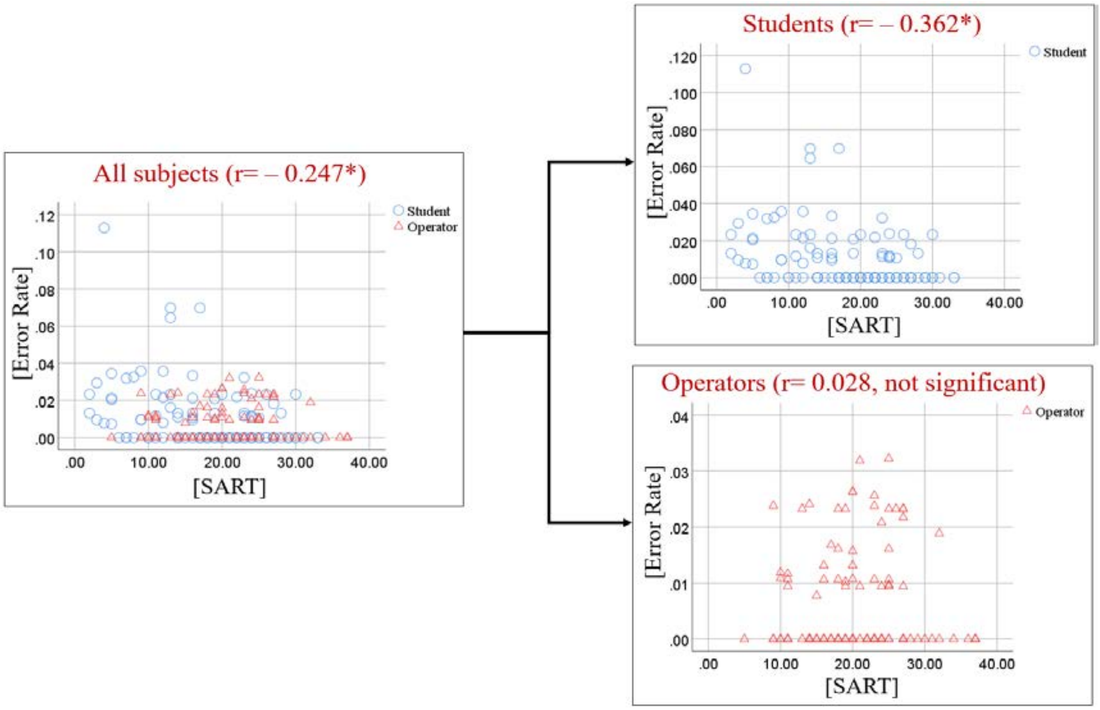


Figure 19. Scatter plot of the relationship between the SART scores and the error rates, both with the operator and student groups combined and with them separated.

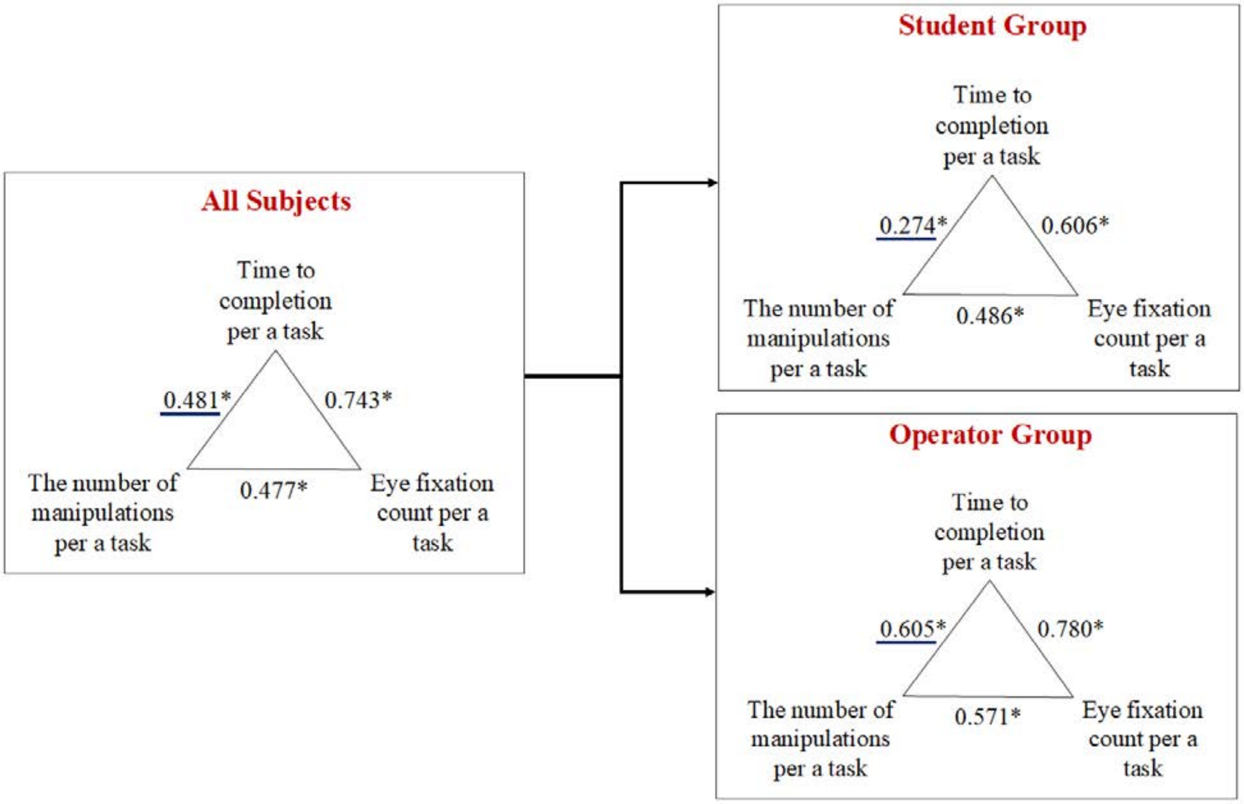


Figure 20. Relationship among time to completion, number of manipulations, and eye movements, both with the operator and student groups combined and with them separated.

In addition, human performance measures pertaining to workload, situation awareness, and error do not significantly correlate with measures of time to completion, eye movements, and number of manipulations, whether in regard to all participants, only the operator group, or only the student group. In particular, the blink rate does not meaningfully correlate with any other human performance measures.

4. DISCUSSION

This section discusses the insights derived from the analysis results introduced in Section 3, covering the differences between the operator and student groups in regard to (1) human performance characteristics, (2) task performance (depending on workload level), and (3) simulator operation.

4.1 Differences in Human Performance Measures

In this study, the six human performance characteristics and their corresponding measures are largely classified into two groups, as shown in Figure 21. Group 1 (subjectivity-oriented measure group) is comprised of human performance characteristics that are products of cognitive activities. These include workload, situation awareness, and error. On the other hand, Group 2 (task-performance-oriented measure group) is comprised of performance characteristics directly estimated from human-system collaboration activities during a given scenario. These include time to completion, eye movements, and number of manipulations. Ultimately, this study shows that minimal correlations exist between the human performance characteristics in Group 1 and those in Group 2 as shown in Table 7.

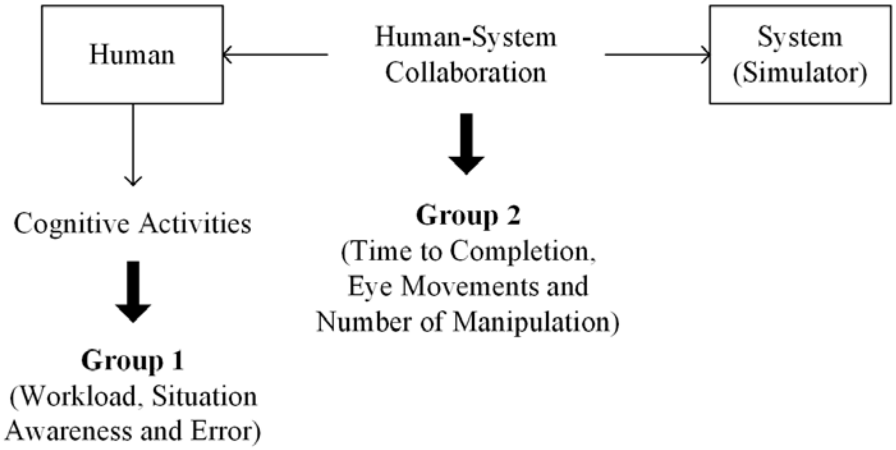


Figure 21. Two human performance characteristic groups.

Differences associated with participant type may correlate with Group 1 rather than Group 2. In this study, we basically assume that actual operators (as opposed to students) are generally much more familiar with NPP systems and operations based on their extensive hands-on experience and regular training, while basic abilities such as recognizing an audible alarm or activating a pump may not vary significantly between operators and students. According to the ANOVA results (see Table 6), the human performance characteristics in Group 1 reflected significant differences, whereas blink rate and average time to complete an instruction were all the measures in Group 2 that significantly differed based on participant type. In fact, Group 1 highly relates to experience, whereas Group 2 aligns more closely with the basic human abilities. This finding may indicate that experience has a definite influence on cognitive performance, though the differences between operators and students in terms of basic human abilities are less significant at least in the Rancor environment.

4.2 Differences in Task Performance and Workload

The major outcome covered in the previous section can be compared against prior research into workload and task performance, as illustrated by the three regions (i.e., “Bored,” “Pressured,” and “Optimal”) shown in Figure 22 (Gore, 2018; Swain & Guttmann, 1983). The first region (“Bored”) indicates that task performance at a very low workload level is not optimal, since workers are insufficiently motivated to remain adequately engaged in their tasks. The second region (“Pressured”) generally results in drastically deteriorated human performance as a result of extremely high workload, causing increased error rates and resultant operational delays. Lastly, the “Optimal” region promotes superior task performance via an adequate level of stress. The present study indicates that operators perform better than students in terms of workload, situation awareness, and error. Furthermore, correlations were found between various human performance characteristics, such as between workload and situation awareness, error rate and workload, and error rate and situation awareness. Among these relationships, the error rate for students correlates with their workload and situation awareness, whereas the error rate for operators does not. From these results, it can be inferred that the students and operators found themselves in the “Pressured” and “Optimal” regions, respectively.

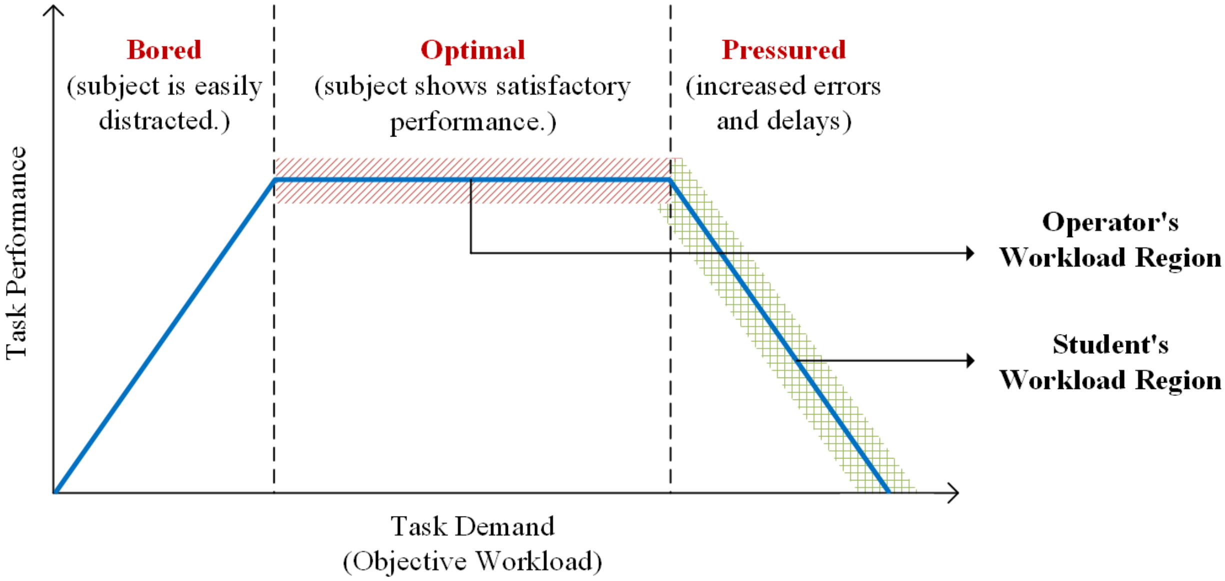


Figure 22. Relationship between workload and task performance.

4.3 Differences in Operator and Student Monitoring and Control of the Simulator

Operational performance within the simulator environment differed in accordance with participant type, as reflected in the heatmaps shown in Figure 13 and Figure 14. In non-event scenarios (Figure 13), operators tended to concentrate on the simulator interface’s core temperature display instead of on other parameters or components, whereas students frequently focused on not only the core temperature but also the reactor rod position indication and reactor power control. In non-event scenarios, reactor power is both the most important parameter and the underlying purpose of the operation. It seems that operators are more likely to concentrate on the most important parameters within the context of a given scenario, being knowledgeable as to what specific information is needed from each window of the simulator interface. Because students are unfamiliar with the basic structure of the simulator and its interface, they may continually seek out options in the control window, as well as changes implemented due to a particular manipulation.

In event scenarios (Figure 14), operators tended to monitor the interface more than students did, despite the students spending more time gathering information. Figure 10 shows that the students’ time to completion and number of manipulations is less than for the operators. This may mean that operators, generally speaking, use their time to complete manipulations, whereas students may devote less time to manipulations but spend longer on tasks such as gathering information (e.g., looking for a valve or indicator).

5. CONCLUSION

This report documents human performance across multiple scenarios for experienced and inexperienced operators. A randomized factorial experimental design was developed featuring two independent variables: participant type and event class. Six human performance measures were considered in the experiment: (1) workload, (2) situation awareness, (3) time to completion, (4) error, (5) eye movements, and (6) number of manipulations. The experiment was conducted using 20 professional reactor operators working at actual NPPs, along with 20 trained students. ANOVA testing and correlation analysis was performed on the experimental data, and the differences in human performance between actual operators and students when using Rancor were discussed. While it is indeed possible to use a simplified simulator with students to collect human performance data related to nuclear operations, there are differences noted between students and operators.

This study is part of an ongoing effort to understand the role of fidelity in the use of simplified simulators and student operators within the SHEEP framework. This report focused on comprehensive tendencies that students and operators have when using Rancor, while observing that some results are difficult to interpret exactly or different from the expected ones. For this reason, our research team is continuously collecting and analyzing student/operator human performance data by using a less simplified simulator (i.e., the Compact Nuclear Simulator [Kwon et al., 1997]) by following the SHEEP framework. In the future, a more detailed human performance analysis will be conducted to better understand human performance differences arising from participant type (i.e., student vs. operator) and simulator complexity (i.e., the more simplified simulator vs. the less simplified one) (Yang, Park, Boring et al., 2022). Then, full-scope inference models will be developed by defining qualitative and quantitative differences in the participant type and simulator complexity (Park et al., 2022b).

For the present purposes, the data collected using Rancor provide an ideal use case for understanding human performance across industry-relevant scenarios that are modeled in NPP HRAs. The mapping of operator performance across multiple factors provides the basis for building scenario models in HUNTER and benchmarking the results of HUNTER analysis outputs against actual student and operator performance. The next planned report will detail the modeling or the ten scenarios from this empirical study in HUNTER.

REFERENCES

Arigi, A. M., Kim, G., Park, J., & Kim, J. (2019). “Human and organizational factors for multi-unit probabilistic safety assessment: Identification and characterization for the Korean case,” *Nuclear Engineering and Technology*, 51(1), 104-115.

Boring, R., Ulrich, T., & Lew, R. (2018). Parts and wholes: Scenarios and simulators for human performance studies. *Advances in Intelligent Systems and Computing, 778,* 116-127.

Boring et al., 2021. An Adaptable Toolkit for Dynamic Human Reliability Analysis: Progress Toward HUNTER 2, Idaho National Laboratory, INL/EXT-21-64525.

Chang, Y. J., Bley, D., Criscione, L., Kirwan, B., Mosleh, A., Madary, T., et al. 2014. "The SACADA database for human reliability and human performance," *Reliability Engineering & System Safety*, 125:117-133. 10.1016/j.ress.2013.07.014.

Cui et al. (2021). “Measurements of team workload: A time pressure and scenario complexity study for maritime operation tasks,” *International Journal of Industrial Ergonomics*, 83, 103110.

Cummings, M., Myers, K., and Scott, S. D. 2006. *Modified Cooper Harper evaluation tool for unmanned vehicle displays.* Paper presented at the Proceedings of UVS Canada: Conference on Unmanned Vehicle Systems Canada.

Dzeng, R. J., Lin, C. T., & Fang, Y. C. (2016). Using eye-tracker to compare search patterns between experienced and novice workers for site hazard identification. *Safety science*, 82, 56-67.

Endsley, M. R., and Garland, D. J. (2000). *Situation awareness analysis and measurement*: CRC Press.

Garcia-Diaz, A., Phillips, D. T. (1995). Principles of experimental design and analysis.

Gawron, V. J. 2019. *Human Performance, Workload, and Situational Awareness Measures Handbook, -2-Volume Set*: CRC Press.

Gertman, D., Gilmore, W., and Ryan, T. 1988. "NUCLARR and human reliability: data sources and data profile," Paper presented at the Conference Record for 1988 IEEE Fourth Conference on Human Factors and Power Plants.

Gertman, D., Blackman, H., Marble, J., Byers, J., & Smith, C. (2005). The SPAR-H human reliability analysis method. US Nuclear Regulatory Commission, 230(4), 35.

Gore, B. F. 2018. "Workload and fatigue" *Space Safety and Human Performance,* pp. 53-85. 10.1016/B978-0-08-101869-9.00003-0.

Gugerty, L. J., & Tirre, W. C. (1996). “Situation awareness: a validation study and investigation of individual differences,” In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 40, No. 12, pp. 564-568). Sage CA: Los Angeles, CA: SAGE Publications.

Hallbert, B., Boring, R., Gertman, D., Dudenhoeffer, D., Whaley, A., Marble, J., Joe, J., and Lois, E. 2006. "Human Event Repository and Analysis (HERA) System," NUREG/CR-6903, Overview, Vol. 1.

Hart, S. G., Staveland, L. E. 1988. “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” *In Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Hogg, D. N., Folles∅, K., Strand-Volden, F., & Torralba, B. (1995). “Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms,” *Ergonomics*, 38(11), 2394-2413.

Jang, I., Kim, A. R., Al Harbi, M. A. S., Lee, S. J., Kang, H. G., & Seong, P. H. (2013). An empirical study on the basic human error probabilities for NPP advanced main control room operation using soft control. *Nuclear Engineering and Design*, 257, 79-87.

Jang, I., Kim, A. R., Jung, W., & Seong, P. H. (2016). Study on a new framework of human reliability analysis to evaluate soft control execution error in advanced MCRs of NPPs. *Annals of nuclear energy*, 91, 92-104.

Joksimovich, V., Spurgin, A., Orvis, D., Moieni, P., and Worledge, D. 1990. *EPRI (Electric Power Research Institute) operator reliability experiments program-Training implications*. Accessed on: INSERT DATE.

Jung, W., Park, J., Kim, Y., Choi, S. Y., and Kim, S. 2020. "HuREX—A framework of HRA data collection from simulators in nuclear power plants," *Reliability Engineering & System Safety*, 194:106235. 10.1016/j.ress.2018.07.036.

Kim, Y., Jung, W., & Kim, S. (2014). "Empirical investigation of workloads of operators in advanced control rooms," *Journal of Nuclear Science and Technology*, 51(6), 744-751.

Kim, Y., Park, J., Jung, W., Jang, I., & Seong, P. H. (2015). A statistical approach to estimating effects of performance shaping factors on human error probabilities of soft controls. *Reliability Engineering & System Safety*, 142, 378-387.

Kim, Y. (2020). Considerations for generating meaningful HRA data: Lessons learned from HuREX data collection. *Nuclear Engineering and Technology*, 52(8), 1697-1705.

Kirwan, B., Basra, G., and Taylor-Adams, S. 1997. *CORE-DATA: a computerised human error database for human reliability support.* Paper presented at the Proceedings of the 1997 IEEE Sixth Conference on Human Factors and Power Plants, 1997. Global Perspectives of Human Factors in Power Generation.

Kwon, K.-C., Park, J.-C., Jung, C.-H., Lee, J.-S., and Kim, J.-Y. 1997. Compact nuclear simulator and its upgrade plan. https://inis.iaea.org/collection/NCLCollectionStore/_Public/29/043/29043412.pdf.

Liu, Y., Jin, X., Luo, Z., Dai, L., Liu, Z., & Li, P. (2021). "Methodology for dynamic reliability assessment of team situation awareness of digital nuclear power plants," *Progress in Nuclear Energy*, 104086.

Lois, E. (2009). *International HRA Empirical Study--phase 1 Report: Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Similar Performance Data*: Office of Nuclear Regulatory Research, US Nuclear Regulatory Commission.

Martins, R., & Carvalho, J. M. (2015). Eye blinking as an indicator of fatigue and mental load—a systematic review. *Occupational safety and hygiene III*, 10, 231-235.

Marusich et al., (2016). "Effects of information availability on command-and-control decision making: performance, trust, and situation awareness," *Human factors*, 58(2), 301-321.

Niehorster, D. C., Hessels, R. S., and Benjamins, J. S. (2020). GlassesViewer: Open-source software for viewing and analyzing data from the Tobii Pro Glasses 2 eye tracker. *Behavior Research Methods*, 1-10.

Park, J., Boring, R. L., Ulrich, T. A., Lew, R., Lee, S., Park, B., and Kim, J. (2022a). "A Framework to Collect Human Reliability Analysis Data for Nuclear Power Plants Using a Simplified Simulator and Student Operators," *Reliability Engineering & System Safety*, 108326. 10.1016/j.ress.2022.108326.

Park, J. and Jung, W. 2006. "A study on the validity of a task complexity measure for emergency operating procedures of nuclear power plants—comparing with a subjective workload." *IEEE Transactions on Nuclear Science* 53.5: 2962-2970.

Park, J. and Jung, W. 2007. "OPERA—a human performance database under simulated emergencies of nuclear power plants," *Reliability Engineering & System Safety*, 92(4), 503-519. 10.1016/j.ress.2006.01.007.

Park, J., Jung, W., and Kim, J. 2020. "Inter-relationships between performance shaping factors for human reliability analysis of nuclear power plants," *Nuclear Engineering and Technology*, 52(1), 87-100. 10.1016/j.net.2019.07.004.

Park, J., Lee, D., Jung, W., and Kim, J. 2017. "An experimental investigation on relationship between PSFs and operator performances in the digital main control room," *Annals of Nuclear Energy*, 101:58-68. 10.1016/j.anucene.2016.10.020.

Park, J., Jang, I., Park, J., Boring, R. L., & Ulrich, T. A. (2022b). "A Framework to Integrate Human Reliability Data Obtained from Different Sources Based on the Complexity Scores of Proceduralized Tasks," Probabilistic Safety Assessment and Management (PSAM) 16.

Parry, G., Beare, A., Spurgin, A., Moieni, P., 1992. An approach to the analysis of operator actions in probabilistic risk assessment. EPRI Report TR-100259.

Preischl, W. and Hellmich, M. 2013. "Human error probabilities from operational experience of German nuclear power plants," *Reliability Engineering & System Safety*, 109:150-159. 10.1016/j.ress.2012.08.004.

Preischl, W. and Hellmich, M. 2016. "Human error probabilities from operational experience of German nuclear power plants, Part II," *Reliability Engineering & System Safety*, 148:44-56. 10.1016/j.ress.2015.11.011.

Sträter, O. 2000. "Evaluation of human reliability on the basis of operational experience," *Gesellschaft für Anlagenund Reaktorsicherheit (GRS) mbh.* GRS-170.

Swain, A. D. and Guttmann, H. E. 1983. *Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report*.

Takacs, Z. K., & Bus, A. G. (2018). How pictures in picture storybooks support young children's story comprehension: An eye-tracking experiment. *Journal of Experimental Child Psychology*, 174, 1-12.

Taylor, R. M. 2017. "Situational awareness rating technique (SART): The development of a tool for aircrew systems design," Paper presented at the meeting of the Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478, 1989.

Tian, P., Xu, G., Han, C., Zheng, X., Zhang, K., Du, C., ... & Zhang, S. (2022). Effects of Paradigm Color and Screen Brightness on Visual Fatigue in Light Environment of Night Based on Eye Tracker and EEG Acquisition Equipment. *Sensors*, 22(11), 4082.

Ulrich, T. A. (2017). *The Development and Evaluation of Attention and Situation Awareness Measures in Nuclear Process Control Using the Rancor Microworld Environment.* PhD Dissertation, University of Idaho.

Ulrich, T. A., Lew, R., Werner, S., & Boring, R. L. (2017). Rancor: A Gamified Microworld Nuclear Power Plant Simulation for Engineering Psychology Research and Process Control Applications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61,* 398-402.

Wei, H., Zhuang, D., Wanyan, X., & Wang, Q. (2013). "An experimental analysis of situation awareness for cockpit display interface evaluation based on flight simulation," *Chinese Journal of Aeronautics*, 26(4), 884-889.

Yan, S., Yao, K., & Tran, C. C. (2021). "Using artificial neural network for predicting and evaluating situation awareness of operator," IEEE Access, 9, 20143-20155.

Yang, T., Park, J., Boring, R. L., Kim, J. 2022. "Human Performance Analysis Depending on Expertise and Simulator Complexity," 2022 American Nuclear Society Annual Meeting, Anaheim, Jun 12-16.