Human Factors and Simulation, Vol. 30, 2022, 1–10
https://doi.org/10.54941/ahfe1001485

Novices as Models of Expert Operators: Evidence from the NRC Human Performance Test Facility

Jinchao Lin[1], Gerald Matthews[2], Niav Hughes[3], and Kelly Dickerson[3]

[1]University of Central Florida, Institute for Simulation and Training Orlando, FL 32826, USA
[2]George Mason University, Department of Psychology Fairfax, VA 22030, USA
[3]U.S. Nuclear Regulatory Commission Rockville, MD 20852, USA

ABSTRACT

One of the challenges for conducting human performance research in the nuclear domain is access to trained operators. Without sufficient sample size, it is difficult to perform analyses with adequate statistical power and draw substantial conclusions. This paper presents the analyses of data collected from three previous NRC Human Performance Test Facility (HPTF) studies to validate the "equal but different" principle. The analyses confirmed that the HTPF studies successfully induced same type and comparable level of cognition workload that would be experienced by expert operators in controlled lab experiments using student novices as participants. Overall, novices can be trained to be models of expert operators to help identify human factors issues in the nuclear domain as well as other domains where access to experts is limited.

Keywords: Nuclear power plant, Equal but different, Workload, Novices, Experts

INTRODUCTION

Humans are integral to the safe operation of a nuclear power plant (NPP). Following the Three Mile Island accident in 1979, the United States Nuclear Regulatory Commission (NRC) began focusing on incorporating good human factors engineering design principles in regulation and emphasizing the importance of adequate training of plant operations staff. As part of this focus, NRC amended its regulations to require facility licensees to have simulation facilities for use in administering NRC operating tests and licensed operator requalification training (52 FR 9460). Since then, the simulator has become an important tool for operator training and license examinations.

As technology develops, new designs and technology becomes available to the nuclear power community. The staff of NRC is responsible for reviewing and determining the acceptability of new designs to ensure they support safe plant operations. Since the human operator is vital to NPP safety, NRC must understand the potential impact of new designs on human performance to support sound regulatory decisions (Hughes, D'Agostino, & Reinerman-Jones, 2017). Despite the importance of human performance in plant safety, much of the basis for current NRC Human Factors Engineering guidance is from other domains (e.g., aviation, defense), qualitative data from operational experiences in NPPs, and limited empirical studies in a nuclear environment (Hughes & D'Agostino, 2016). To close this data gap, NRC launched the Human Performance Test Facility (HPFT) project to explore the impact of new designs, technologies, and concepts of operations on human performance using generic simulator platforms.

One of the challenges for conducting human performance research in the nuclear domain is access to trained operators. Without sufficient sample size, it is difficult to perform analyses with adequate statistical power and draw substantial conclusions. To overcome the participant access challenge NRC partnered with the University of Central Florida (UCF) and use college students as a proxy for expert operators to study the impact of traditional and new Main Control Room (MCR) designs, technologies, and concepts of operations on performance of common NPP tasks and physiological and subjective workload. This approach follows the principle of "equal but different".

Equal but Different Approach

Conducting laboratory-based research studies using novice college students as a sample to get meaningful findings that are generalizable to the real-world environments can be a challenge. Criticism of the methods used to perform laboratory experiments typically stems from a concern that the participant pool excludes operational experts and/or a misunderstanding of experimental strategies available to account for differences between novice and expert participants (Lackey, Reinerman-Jones, & Salcedom, 2014). In order to collect meaningful data from novices, the "equal but different" approach was adopted in the study designs of the HPTF project experiments. By following the principle of "equal but different" approach, although the simulated environment in the laboratory is different from the real-world working environment, the experimental task scenarios are controlled to induce the same type and comparable level of cognition workload that would be experienced by expert operators. Specifically, the complexity of the environment was reduced, and the operation procedure was modified so that the same type and comparable level of cognition workload can be experienced by novice participants without requiring them to have all the knowledge and skills of an expert operator.

METHODS

Data collected in three previously completed NRC HPTF studies were reviewed and reanalyzed to demonstrate the validity of using novices as models of expert operators in human factors studies in the nuclear operation domain. The studies investigated workload responses of both novices and experts using simulators with different types of interfaces in three common NPP operation task types. Designs of the three studies are summarized in Table 1.

Table 1. Summary of study designs.

Study	Participant	Sample size	Simulator	Interface
Study A	Student novices	71 ($M = 20.2$, $SD = 2.7$)	GSE GPWR	Simulated analog/Touch
Study B	Expert operators	18 ($M = 45.9$, $SD = 10.6$)	GSE GPWR	Simulated analog/Touch
Study C	Expert operators	30 ($M = 55.5$, $SD = 7.8$)	TTC Westinghouse PWR	Analog/Manual

The experimental scenario developed for Studies A and B consisted of tasks reflecting common activities required when completing operating procedures: checking, detection and response implementation. Checking requires a one-time inspection of an I&C to verify that it is in the appropriate state. Detection requires continuous monitoring of a control parameter to identify a change in the state of the plant. Response implementation requires a fine motor response (mouse usage or finger touch) to change the state of the NPP by locating a control and subsequently manipulating the control in the required direction. There were twelve steps in the experimental scenario, grouped by task type (4 checking steps, 4 detection steps, and 4 response implementation steps). The order of task type block was partially counterbalanced across participants as a means of balancing the need for laboratory control (i.e., using tasks blocks) and realism (i.e., a checking task would never follow either of the other two tasks in the real environment) (Reinerman-Jones, Guznov, Mercado, & D' Agostino, 2013).

The experimental scenario for Study C was developed based on a generic version of an emergency operating procedure (EOP) for a "Loss of All Alternating Current Power (ECA-0.0)" scenario but modified for experimental use. The experimental procedure contained 69 steps supporting three different task types. In the experimental procedure, there were 30 steps (16 checking, 5 detection, and 9 response implementation) for RO 1, and 39 steps (27 checking, 1 detection, and 11 response implementation) for RO 2. The number of steps was not balanced nor was task type due to the nature of the original, realistic EOP, which requires steps to be taken in a prescribed sequence.

Multivariate Workload Assessment

There has been a longstanding debate in human factors over the optimal methodology for workload assessment. While the NASA Task Load Index (NASA-TLX) is the single most popular workload measure, it does not provide a comprehensive workload assessment. Studies conducted in the HPTF project utilized a multivariate workload assessment strategy and supplemented the NASA-TLX with additional subjective measures, such as the Multiple Resource Questionnaire (MRQ), psychophysiological measures derived from electrocardiogram (ECG), electroencephalogram (EEG), transcranial doppler (TCD), and functional near-infrared spectroscopy (fNIR), and performance-based measures including the index of effectiveness of communication and accuracy of task execution. This paper will focus on selected subjective and psychophysiological measures.

RESULTS

To validate the “equal but different” approach and demonstrate that the same type and comparable level of cognition workload would be experienced by expert operators can be induced in novices, we conducted Chi-square goodness-of-fit tests to compare the distribution of the NASA-TLX ratings between the novices and experts and ANOVAs to compare the workload metrics from both subjective and psychophysiological measures.

Chi-Square Goodness-of-Fit Test

To determine if workload level experienced by the novices were comparable to the workload experienced by the expert operators, a series of chi-square goodness-of-fit tests were conducted to compare the distribution of the NASA-TLX ratings between novices and experts as well between experts using different types of simulators. Based on a recent meta-analysis of NASA-TLX scores (Grier, 2015), the MCR operation tasks can be categorized as command-and-control tasks. Following Grier’s command-and-control task model, NASA-TLX ratings less 38 were categorized as low workload, ratings greater than 60 were categorized as high workload, and ratings between 38 and 60 were categorized as medium workload. The frequency of workload level ratings is shown in Table 2. The physical demand and temporal demand subscale ratings were not analyzed due to the nature of the frequency distributions. For the other subscales, the chi-square goodness-of-fit tests indicated that the workload experienced by the novice participants was similarly distributed to the workload experienced by the expert operators using the same simulator with touchscreen interface. By comparing the frequency distributions between study B and C, the chi-square goodness-of-fit tests revealed that only the experienced performance ratings are similarly distributed between expert operators using the digital simulator with touchscreen interface and the analog simulator.

Subjective Measures

Figure 1 illustrates a comparison of the overall NASA-TLX ratings collected after the entire experimental scenario in Study C and the averaged ratings from three task types in Study A and B. No significant difference in global workload and all six subscales between novices (Study A) and experts (Study B) using the same type of simulator was revealed by the analysis. Although there is no statistical difference between Study A and B, generally, the expert operators experienced lower workload than the novices and experts using the analog simulator. Compared to the experts using the analog simulator, the novices using the digital simulator experienced greater frustration, $F(2,90) = 3.35, p < .05, \eta_p^2 = .07$. In addition, expert operators using analog simulator reported greater temporal demand than experts using the digital simulator, $F(2,90) = 3.16, p < .05, \eta_p^2 = .07$. The findings are consistent

Table 2. NASA-TLX frequency table.

Frequency	Study B	Study A	Study C	χ^2 B vs A	χ^2 B vs C
Global workload				2.28	67.85*
Low	11	43	12		
Medium	6	25	6		
High	1	6	12		
Mental demand				5.17	15.40*
Low	11	33	12		
Medium	4	20	5		
High	3	16	13		
Performance				1.98	5.60
Low	12	41	22		
Medium	4	20	2		
High	2	8	6		
Effort				.85	7.35*
Low	12	45	21		
Medium	4	18	2		
High	2	6	7		
Frustration				.59	12.36*
Low	10	41	26		
Medium	6	22	1		
High	2	6	3		
Physical demand				NA	NA
Low	17	55	26		
Medium	2	12	3		
High	1	2	1		
Temporal demand				NA	NA
Low	14	39	51		
Medium	4	25	4		
High	0	5	5		

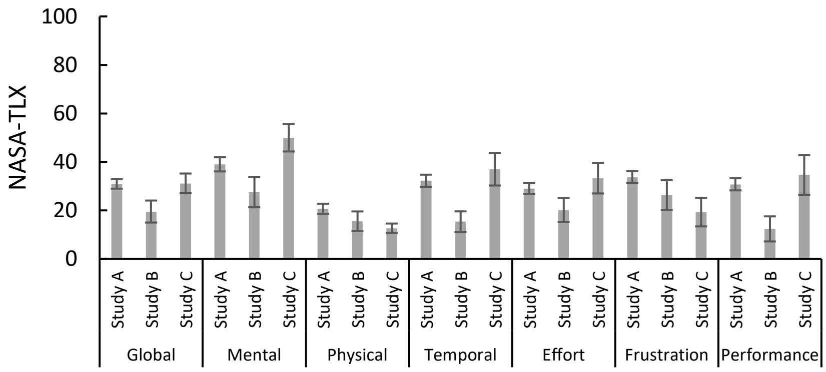

Figure 1: NASA-TLX ratings. Error bars represent standard errors.

with Grier's classification of command and control tasks, workload experienced by novices and experts was low to medium across the NASA-TLX subscales.

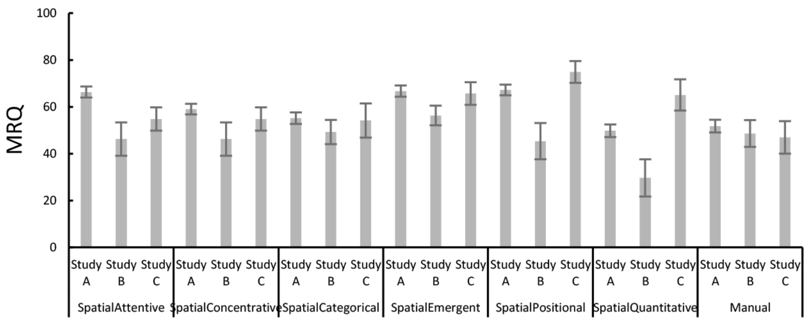

Figure 2a: MRQ ratings. Error bars represent standard errors.

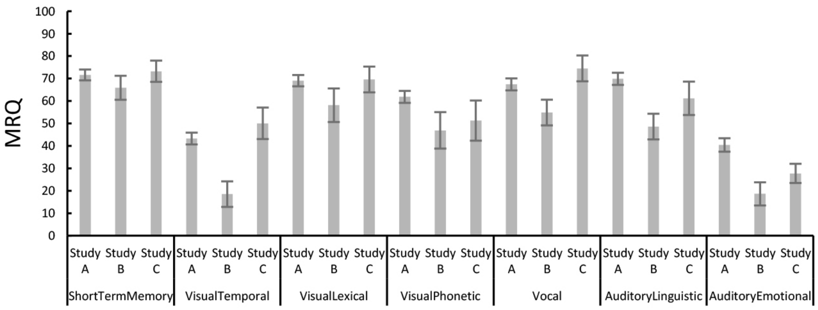

Figure 2b: MRQ ratings (Continued). Error bars represent standard errors.

Figure 2a and 2b illustrate a comparison of the overall MRQ ratings collected after the entire experimental scenario in Study C and the averaged ratings from three task types in Study A and B. The MRQ revealed similar trends between novices (Study A) and experts (Study B), especially in the subscales that would play a critical role in influencing the cognitive demands during NPP MCR operations, such as short-term memory, spatial attentive demand, spatial concentrative demand, and spatial categorical demand. Consistent with the findings revealed by NASA-TLX, experts in Study B using digital simulator reported the lowest workload across the MRQ subscales. Notably, some of the group differences were significant. Specifically, expert operators using analog simulator reported lower auditory emotional demand than novices using the same type of simulator, $F(2,90) = 4.77$, $p < .05$, $\eta_p^2 = .10$. Expert operators in Study B also reported significantly lower spatial positional demand ($F(2,90) = 6.91$, $p < .01$, $\eta_p^2 = .13$), spatial quantitative demand ($F(2,90) = 6.73$, $p < .01$, $\eta_p^2 = .13$), and visual temporal demand ($F(2,90) = 6.05$, $p < .01$, $\eta_p^2 = .12$) than participants in Study A and C.

Psychophysiological Measures

One-way ANOVAs were conducted to determine if the psychophysiological indices of workload were significantly different between novices and expert operators on a task-type basis. Selected metrics from four psychophysiological sensors, including ECG, EEG, fNIR, and TCD are reported in the following sections.

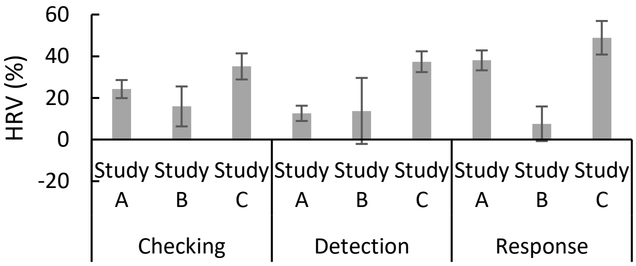

Figure 3: Heart rate variability percentage change from baseline by task type. Error bars represent standard errors.

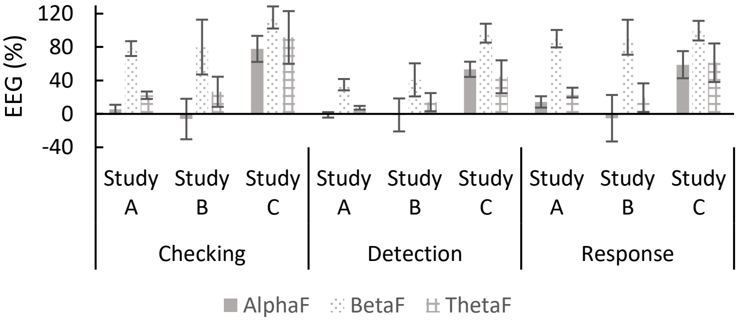

Figure 4: EEG alpha/beta/theta percentage change from baseline by task type. Error bars represent standard errors.

No significant difference between novices and expert operators in HRV change after exposure to checking, detection, and response implementation tasks was revealed (Figure 3).

For EEG analysis, three bandwidths, alpha (8-13 Hz), beta (13-30 Hz), and theta (4-8 Hz), were used as the workload indices in the spectral frequency analysis. No significant difference between novices (Study A) and expert operators using the same digital simulator (Study B) in EEG alpha, beta, or theta change in the frontal lobe after exposure to checking, detection, and response implementation tasks was revealed (Figure 4). Although the difference between novices and experts using the same type of simulator was not significant, compared to experts in Study C using the analog simulator, the EEG changes were in a much smaller magnitude in all three bands among novices, especially in detection tasks.

Figure 5 demonstrates the change in oxygen saturation in left and right hemispheres measured by fNIR. Similar to ECG and EEG results, no significant difference between novices and expert operators in oxygen saturation change during the checking, detection, and response implementation tasks was revealed.

Figure 6 demonstrates the change in cerebral flood flow velocity measured by TCD. Similar to other psychophysiological metrics, no significant difference between novices and expert operators in oxygen saturation change

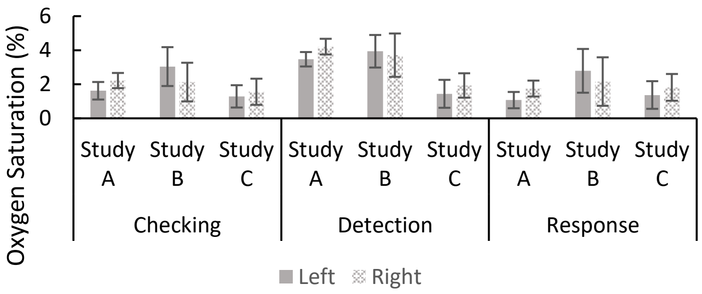

Figure 5: Oxygen saturation in left/right hemisphere percentage change from baseline by task type. Error bars represent standard errors.

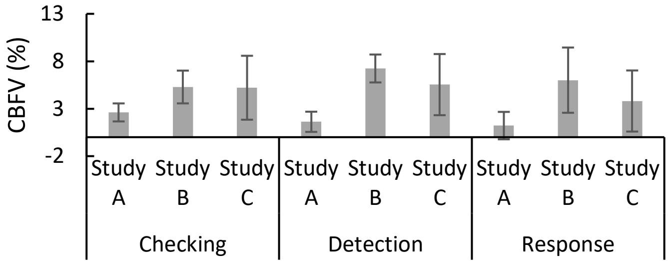

Figure 6: Cerebral flood flow velocity percentage change from baseline by task type. Error bars represent standard errors.

after exposure to checking, detection, and response implementation tasks was revealed.

DISCUSSION

The analyses compared data collected from three experiments and revealed evidence of using novices as models of expert operators in the nuclear domain. The chi-square goodness-of-fit tests confirmed that the workload (measured by NASA-TLX) experienced by the novice participants recruited in the HPTF study were similar distributed as the workload experienced by the expert operators. The non-significant ANOVA results from the subjective measures, such as NASA-TLX and MRQ, suggested that the novices not only fit the distribution of the expert operators, but experienced same type and comparable level of cognition workload that would be experienced by expert operators in the controlled experiments. Although not all resources covered by the subscales of MRQ showed non-significant results in novice-expert comparisons, the experimental manipulations successfully induced the most important resources in NPP operations, such as short-term memory and the processes related to various spatial demand, in comparable levels. The non-significant ANOVA results from the psychophysiological measures indicated that the novices not only experienced and reported the same type and comparable level of cognition workload, but the workload manipulation indeed induced similar changes in physiological response. Taken together the studies revealed that the "equal but different" approach induced same type and comparable cognitive level of cognition workload in novices and expert operators.

Student novices can not only stand in for expert operators who work in the current NPPs to help identify workload-related safety concerns but also serve as good representatives of future workforce working in the next-generation NPP MCR in the future. Most of current expert operators were trained to use the traditional analog control system. Currently, there are no experts for the emerging technology. As technology advances, the industry calls for plant modernization. Future workforce may not have the training and expertise that current expert operators have. In addition, compared to current expert operators, future operators are likely to have different attitude toward technology which may affect their interaction with the control system and introduce novel human factors issues. Future research could use novices as models of expert operators beyond workload research and extend the "equal but different" approach to human factors issues related to other MCR technologies, such as automation and novel control room configurations. Further, the method developed in the HPTF can be applied in other domains where access to experts is limited.

In conclusion, the analyses confirmed that the HTPF studies successfully induced same type and comparable level of cognition workload that would be experienced by expert operators in controlled lab experiments using student novices as participants and validated the methodology of the "equal but different" principle. Overall, novices can be trained to be models of expert operators to help identify human factors issues in the nuclear domain as well as other domains where access to experts is limited.

ACKNOWLEDGMENT

This paper was prepared as an account of work sponsored in part 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 any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the U.S. Nuclear Regulatory Commission.

REFERENCES

Grier, R.A., 2015, September. How high is high? A meta-analysis of NASA-TLX global workload scores. In Proceedings of the HFES Annual Meeting (Vol. 59, No. 1, pp. 1727–1731). CA: SAGE.

Hughes, N. and D'Agostino, A., 2016, September. Gathering meaningful data from novices and or in simplified operating environments to inform us about highly complex operational environments. In Proceedings of the HFES Annual Meeting (Vol. 60, No. 1, pp. 47–50). CA: SAGE.

Hughes, N., D'Agostino, A. and Reinerman-Jones, L., 2017. The NRC Human Performance Test Facility: An approach to data collection using novices and a simplified environment. In Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries (pp. 183–192). Springer, Cham.

Lackey, S., Reinerman-Jones, L.E. and Salcedo, J., 2014. Equal but different: 5 research strategies for improving conclusions drawn from novice populations. In Proceedings of the 2014 MODSIM World Conference, Hampton, VA (April, 2014).

Reinerman-Jones, L., Guznov, S., Mercado, J. and D'Agostino, A., 2013, July. Developing methodology for experimentation using a nuclear power plant simulator. In International Conference on Augmented Cognition (pp. 181–188). Springer, Berlin, Heidelberg.