Experimental study on the adverse effects of having excessive safety rules

NAKANOWATARI Hiroyuki[1], TAKAHASHI Makoto[2], and WAKABAYASHI Toshio[3]

1. Department of Quantum Science and Energy Engineering, Tohoku University, 6-6-01 Aza-Aoba, Aoba-ku, Sendai, 980-8579 Japan (hiroyuki.nakanowatari@luke.qse.tohoku.ac.jp)
2. Department of Quantum Science and Energy Engineering, Tohoku University, 6-6-01 Aza-Aoba, Aoba-ku, Sendai, 980-8579 Japan (makoto.takahashi@qse.tohoku.ac.jp)
3. Department of Quantum Science and Energy Engineering, Tohoku University, 6-6-01 Aza-Aoba, Aoba-ku, Sendai, 980-8579 Japan (toshio.wakabayashi@qse.tohoku.ac.jp)

Abstract: Operating and maintenance personnel are required to follow a large number of rules, with a higher level of safety maintained by appropriately following the rules. Although following the rules is quite important, incidents attributable to the erroneous interpretation of rules have not been eliminated. Furthermore, where the number of rules has increased rapidly, the task of following the rules becomes quite difficult. Although most working personnel in the field would agree that the number of rules seems excessive, no evidence has been provided to show that such an excessive number of rules for safety may have adverse effects. In the present study, cognitive experiments have been performed to show that excessive quantities of rules can result in degraded safety performance, and that task performance is affected by how rules are presented.

Keyword: rules; regulation; adverse effects

1 Introduction

Nuclear power plays an important role as a base electric power source in the world. On the other hand, higher levels of reliability and safety are required for nuclear power plants because of the potential risks associated with radioactivity. The safety of nuclear power plants is maintained by the three factors: the operators (human beings), the equipment (machines) and the interface which mediates between the operators and the equipment. To meet the need for a higher level of safety in nuclear power plants, much effort has been paid on the last two factors [1].

Although the interface system has been improved to meet the human factor requirements, organizational factors remain as the major cause of human errors in nuclear power plants [2]. In the present study, the focus has been set on the rules which have been imposed on personnel. The number of rules has increased monotonically in response to the troubles that have actually occurred at nuclear power plants. Consequently, the amount of rules has become too large, which can lead to the situation in which following the rules itself becomes quite difficult.

2 Method

In the present study, the following two hypotheses were validated based on cognitive experiments:

1. Excessive rules can result in degraded safety performance.
2. Safety performance can be influenced by how rules are presented.

The experiments were undertaken in two phases corresponding to the above hypotheses. In the phase one experiment, the influence of the quantity of rules on operator performance was evaluated. Task performance and mental workload were taken as the measures for evaluation. In this experiment, the relationship between the number of rules and the level of mental workload was evaluated using a simulated task in which subjects were instructed to make decisions based on the given information and rules. In the phase two experiment, the effects of presenting the relative priority of rules in a differentiated manner was evaluated. In this experiment, a smart grid system was used as an

Received date: May 31, 2012
(Revised date: June 21, 2012)

example task environment, in which subjects were instructed to control a smart grid system, which undergoes dynamic changes, whilst following the imposed rules. In both experiments, the operation logs were recorded and the behavior was analyzed based on these logs.

The NASA-TLX [3] (Task Load index) was used to evaluate the subject's mental workload. The NASA-TLX allows users to perform subjective workload assessments on operators working with various human-machine systems. The NASA-TLX is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales. These subscales include Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration.

3 Phase one experiment

3.1 Simulation environment

In the phase one experiment, a virtual incinerator plant (Fig.1) was taken as an example, in which subjects were instructed to determine the appropriate processing unit based on the situation shown on the screen in conjunction with the given rules. The purpose of this experiment was to verify the following hypothesis (restated from above) by performing cognitive experiments simulating actual working conditions.

Fig.1 The virtual incinerator plant.

"Rules primarily aimed to enhance reliability may have adverse effects on the performance of operators when the total number of rules becomes excessive"

3.2 Experimental design and tasks

This experiment consisted of two sets; the first was a preliminary experiment to evaluate the validity of the simulated working environment. The number of rules required to perform the given task and the additional rules aimed to enhance task reliability were tuned. The following experiments were performed with modifications based on the results of the preliminary experiment. In both experiments, the rules were prepared so that they appeared to simulate real working conditions. In these experiments, two different kinds of rules were given to the subjects according to the experimental conditions. Firstly, the "operation rules" required to make appropriate decisions, and secondly, the "safety rules" for improving reliability and safety. These rules are briefly described below:

A. Operation rules

- Necessary to accomplish proper operation.
- It is not possible to make a decision when these rules are not available.

Example – *"If the recycling equipment is not working, it cannot be used."*

B. Safety rules

- For the improvement of reliability and safety of the operation.
- It is possible to make proper decisions even when these safety rules are not available.

Example –*"The operation check sheet (shown in Fig.2) must be completed, and a decision must be made only after referring to the check sheet"*.

The experimental conditions were divided into three levels as shown below:

Level 1: Only the operation rules are given to the subjects.

Level 2: Both the operation rules and the safety rules are given to subjects.

Level 3: Both the operation rules and the safety rules are given to subjects. However, the number of safety rules is larger than in level 2.

Table 1 shows the number of rules relative to the operation rules and the time allowed for each decision at each level. The time allowed for each decision was determined proportional to the total

		Answer Yes (Y) or No (N) as appropriate if the following conditions or procedures have been met:								
Garbage ID	Candidate of selection	State of driving	Process margin	Safety Margin	Process Margin measurement sensor	Combustion temperature	Combustion temperature measurement sensor	Safety alarm	Supervisor permission	Execu-tion of work
1	R	Y	Y	Y	Y	/	Y	Y	Y	OK
2	A	Y	Y	Y	N					NG
	B	Y	Y	Y	Y	Y	Y	Y	Y	OK
3										
:	:	:	:	:	:	:	:	:	:	:

Fig.2 Example of operation check sheet (Safety rules).

Table 1 The ratio of the number of rules and the time limit per question at each level

Group	Time limit [s]	Operation rules	Safety rules (Check sheet)	Safety rules (Process rules)	Amount of Total rules
A (level 1)	40	1	0	0	1
B (level 2)	60	1	0.25	0.25	1.5
C (level 3)	80	1	0.5	0.5	2

number of rules.

3.3 Subjects

Twenty-six graduate students participated in the experiments (Ages from 18 to 24 years old with normal vision). All the subjects have sufficient computer experience and mouse device manipulating skills. They were informed that the number of rule violations would be used as a human performance score.

3.4 Results and discussion

A. Objective evaluation results

Figure 3 shows the results for the incorrect answer rate for each level. Here, "error by safety rule", in which execution of a required procedure determined by the safety rules was omitted, can be categorized as an omission error. On the other hand, "Error by operation rule", in which an erroneous decision was made by incorrectly applying operation rules, can be categorized as a commission error. The operating performance in level 2 is higher than that of level 1, and the operating performance is degraded for level 3 despite the larger number of safety rules, which were conversely expected to contribute to enhancing total safety. The results of this experiment imply that too many "safety rules" could have a negative influence upon the operating performance. The hypothesis has been confirmed as far as our experimental results are concerned, although the number of subjects is rather limited and they are not real plant personnel. It is believed that the results obtained in this study suggest that there may be an appropriate number of safety rules beyond which adding excessive safety rules may not necessarily result in enhanced safety.

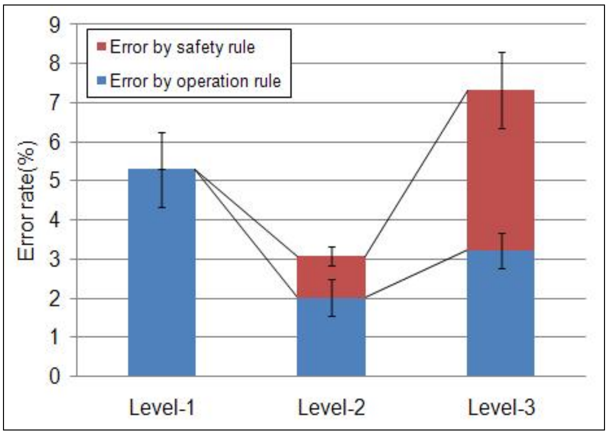

Fig.3 Error rate by experiment level.

B. Mental workload evaluation results

Figure 4 shows the mental workload evaluation result for each level of assessment. Subjects in the level 3 experiment felt much more time pressure than subjects in the other levels. The value of the subjective time pressure is large in level 3 although the allowed time for each decision has been balanced according the number of rules. This result also implies that the addition of the excessive safety rules may have adverse effects on perceived time pressure.

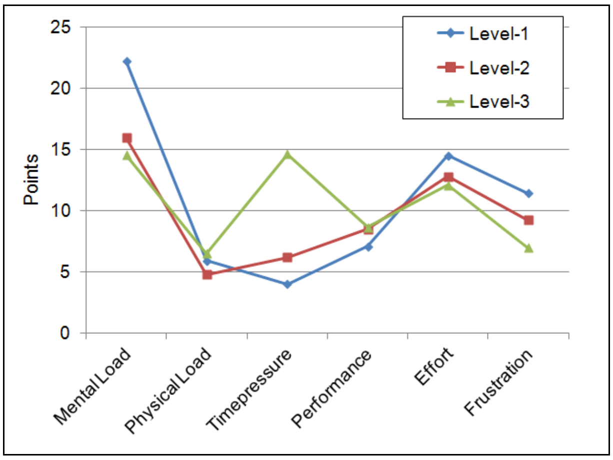

Fig.4 NASA-TLX for the first phase experiment.

4 Phase two experiment

4.1 Simulation environment

In the phase two experiment, the effects of highlighting the relative priority of the rules were evaluated using the smart grid (SG) simulation environment (Fig.5). In this experiment, the existence of difference levels of priority of rules was focused on and the hypothesis that "subjects perform better when the rules are given with different designated level of priority" was validated. The simulation system for smart grid operation consists of a fuel cell system (Fig.6), wind power generation system, photovoltaic generation system and a Sodium Sulfur (NAS) battery system as power sources. An industrial zone and a substation for the supply of electric power are simulated as demand–side loads.

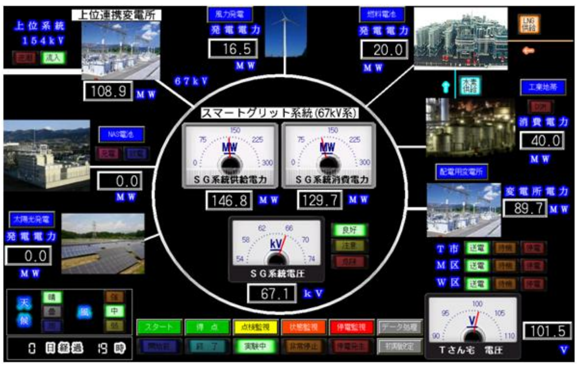

Fig.5 The smart grid simulation environment.

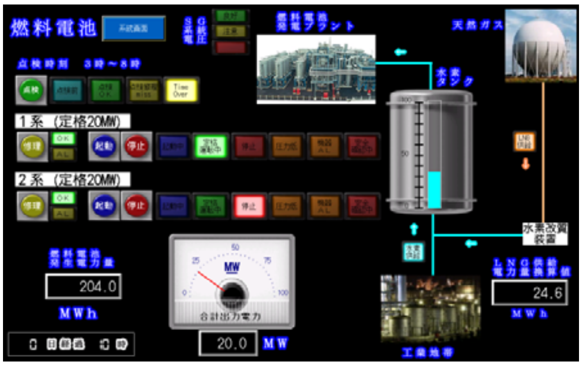

Fig.6 The fuel cell operation screen.

4.2 Experimental design and tasks

Subjects were instructed to follow the two rule sets (operation rules and safety rules) in the same way as in the phase one experiment. "Penalty factors" were applied to a variety of situations in order to designate undesirable performance conditions. Subjects in group A were instructed to follow the rules with less differentiation in the penalty factors. On the other hand, subjects in group B were instructed to follow the rules with large differentiation in their penalty factors. For example, the penalty factor for blackouts is four times larger for Group B than for group A. Subjects operated the SG system for five days of scenario time in the "micro world". Five days in the micro world corresponds to twenty minutes in the real world.

A. Evaluation criterion

Subject's performance for basic operation was measured by the following four kinds of evaluation index.

1. SG voltage stability point (SVP) [points]

A higher score was given when the SG voltage stabilized near the standard voltage (AC 67.0kV). On the other hand, the score decreased when the SG voltage became unstable or deviated from the standard voltage. The maximum points attainable for SVP was 60,000.

2. Blackout (BO) [times]

Blackout occurs in the substation for the supply of electric power when the SG voltage becomes unstable continuously for one hour of micro world time. Penalty points for Blackout differ depending on the experimental conditions. For the subjects in group A, the associated penalty points were -5000, while for the group B, penalty points were -20,000.

3. Emergency stop (ES) [times]

Each piece of equipment stops automatically when a violation of the operation rules occurs. For instance, the wind power generator stops automatically when the velocity of the wind exceeds 25**m/s**. For each ES, -2,000 penalty points were given.

4. Violation of inspection (VI) [times]

Subjects were instructed to perform a simulated inspection task at the predefined time. A violation of inspection occurs when the necessary check is not correctly performed. Penalty points for VI`s differ depending on the experiment scenario. For the subjects in group A, the penalty points were -1,000, while for the group B, the penalty points were -100.

On the other hand, the performance related to safety is evaluated based on a safety margin and a reaction time. The criteria used to evaluate the safety of the operation are shown as follows.

1. Electric power effective adjustment rate (EER) [%]

Ratio of the amount of power generation that contributes to stabilizing the SG voltage to the total amount of power generated.

2. Inspection average time (IAT) [hours]

This is the index of the reaction time, determined by the elapsed time after starting the inspection time window, which indicates how quickly the inspection was performed.

3. Abnormal state average time (ALAT) [hours]

Abnormal states are scheduled to occur at random frequency in each piece of equipment. Subjects were instructed to undertake repair actions, whenever abnormal states were found.

4. Recovery average speed time (RAST) [mins]

RAST is the total time spent on recovery operations after an emergency stop or blackout occurs.

B. Difference of priority of rules for each group

Table 2 shows the differences in the evaluation criterion for the group A and B. While the SVP was the same for both groups, the BO penalty was much larger for group B. This meant that the subjects in group B were strongly motivated to avoid blackout.

4.3 Subjects

Twenty-two graduate students participated in these experiments (Ages from 19 to 22 years old with normal vision). All the subjects had sufficient computer experience and mouse device manipulation skills. Subjects were instructed to try to obtain the highest points that they could by stabilizing the SG voltage and by avoiding blackouts, emergency stoppages and failing to undertake inspections.

4.4 Results and discussion

A. Objective evaluation results

Table 3 shows the evaluation results for task performance for basic operation. For these indices, no significant differences were found between the two groups. As for the blackout frequency of group B, almost half of the subjects of group B caused blackouts despite the subjects being strongly motivated to avoid their occurrence. This fact should be interpreted in relation to the level of mental workload, which is discussed in the following section. As for the frequency of failure of inspection, no difference between groups was found irrespective of the differences in the penalty points for inspection failure. This result implies that it is difficult to decrease violation by raising the penalty when the risk of occurrence is high

Table 2 Evaluation criteria for each group (SG voltage stability points and allocated penalty points per occurrence of violation)

Group (difference of priority)	SG voltage stability point (SVP)	Blackout (BO)	Emergency Stop (ES)	Violation of Inspection (VI)
A (small)	MAX +60,000, MIN -60,000	-5,000	-2,000	-1,000
B (large)	MAX +60,000, MIN -60,000	-20,000	-2,000	-100

Table 3 Task performance for basic operation

Group	SVP [points]	BO [times]	ES [times]	VI [times]
A	33,007	0.27	3.27	3.00
B	34,321	0.45	3.00	3.09

Table 4 shows the evaluation results of the task performance in relation to safety. In regards to the indices, a difference between the two groups was found in two indices. The ALAT of group B shows better performance although it is not statistically significant. This means that subjects in group B performed repair actions quicker than subjects in group A. As for the RAST, subjects in group B showed better performance ($p<0.05$). The result implies that the workload level for the subjects in group B may have been smaller than that of group A.

Table 4 Task performance in relation to safety

Group	EER [%]	IAT [hours]	ALAT [hours]	RAST [mins]
A	88.9	1.76	7.37	44.9
B	88.9	1.90	6.06	30.0

B. Mental workload evaluation results

The NASA-TLX was used to estimate the level of mental workload. Figure 7 shows the NASA-TLX evaluation results for both groups. In general, no large difference was found between the groups in each item contributing to the mental workload. Next, the subjects in group B were further divided into two groups: ones who caused blackouts and ones who did not cause blackouts. The difference in one of the items in the NASA-TLX, namely time pressure, was found as shown in Fig.8. In the phase one experiment, it was shown that time pressure plays an important role. Thus, the focus was narrowed to examine the relationship between time pressure and task performance.

C. Correlation of time pressure and performance

For the correlation between performance score and the subjective time pressure score to be made, the evaluation criteria for performance must be the same.

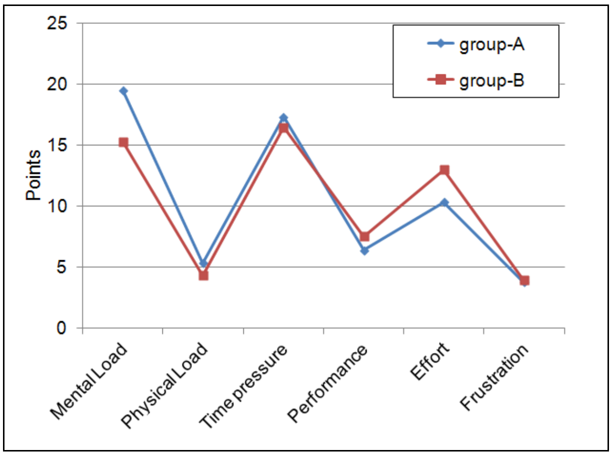

Fig.7 NASA-TLX of the second phase experiment.

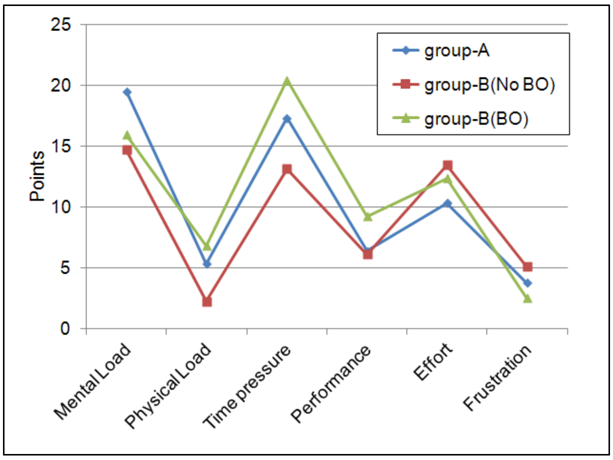

Fig.8 NASA-TLX of classification due to blackout in group-B.

Table 5 Example for evaluation of operation rules

Group	SVP [points]	BO [times]	ES [times]	VI [times]
Sample	30,000	1	2	3

Thus, the score of the subjects of each group must be re-calculated based on a common rule, combining the rules that were utilized for each of the groups and applying them to the resulting performance of each subject. The example data to demonstrate the common calculation rule is shown in Table 5. The score against the common rule is calculated by adding the resulting score for each rule. The calculation follows the three steps shown below:

Step1. Score calculation by criteria A

Criteria A score = 30,000+(-5,000)×1+(-2,000)×2+(-1,000)×3 = 18,000

Step2. Score calculation by criteria B

Criteria B score = 30,000+(-20,000)×1+(-2,000)×2+(-100)×3 = 5,700

Step3. Score calculations are added.

Score on common rule = Criteria A score + Criteria B score = 18,000 + 5,700 =23,700

All of the subjects' scores were re-calculated using these procedures, with the re-calculated scores of both groups shown in Table 6. The re-calculated scores for both groups are almost the same value. In addition, the standard deviation of both groups are also almost the same. However, the correlation pattern between the performance and time pressure shows a significant difference.

Table 6 Score on common rule

Group	Score on common rule [points]	Standard deviation
A	42,715	21,536
B	41,878	21,827

The results of group A in Fig.9 shows no correlation between the performance and time pressure. On the other hand, in group B`s case (shown in Fig.10), a negative correlation is found. The fact that a negative correlation between the task performance and the subjective time pressure was found only for the subjects in group B implies that better performance can be achieved by controlling time pressure to make it lower, which is one of important factors determining mental workload. This means that the rules of group B , which emphasize the difference of priority between the rules, may be better than group A. Taking this into account, the task performance can potentially be explicitly controlled by monitoring the subjective time pressure and adjusting the weighting applied to different rules. Combined with the results showing better performance in the safety related criteria of ALAT and RAST for the subjects in group B as described in the earlier section, the working hypothesis "That safety performance can be influenced by how the rules are presented" has been confirmed to a reasonable extent.

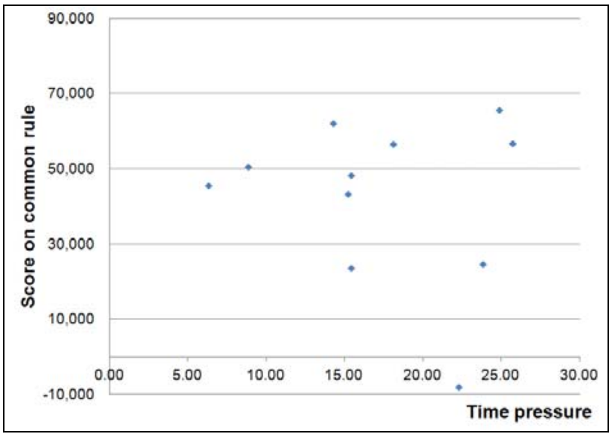

Fig.9 Correlation of group A.

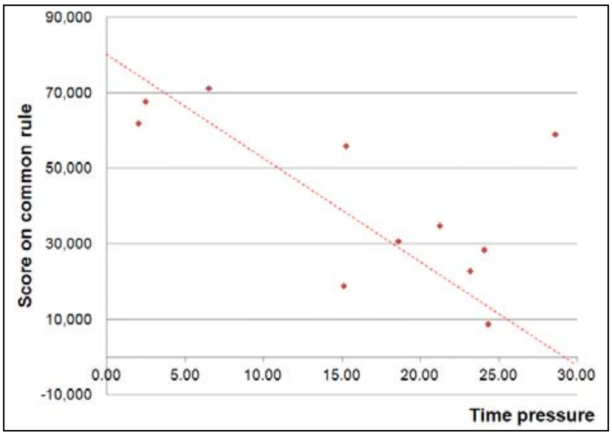

Fig.10 Correlation of group B.

5 Conclusion

Based on the results of the first and second phase study using micro-world simulation, the adverse effect of excessive rules and the effect of the way in which rules are presented on task performance have been demonstrated. In the phase one experiment, it was confirmed that when excessive rules were imposed the safety performance degraded. In the phase two experiment, it was shown that presenting rules with different emphasis on relative priority may have positive effects on safety performance and may provide potential to maintain a higher safety level by monitoring the subjective time pressure. In conclusion, the possibility of adverse effects on human performance has been shown when excessive numbers of rules are provided without considering the influence on time pressure.

Further validation of the proposed hypothesis in a more realistic working environment is underway.

References

[1] KIM, J.H., LEE, S., and SEONG, P.: Investigation on Applicability of Information Theory to Prediction of Operator Performance in Diagnosis Tasks at Nuclear Power Plants: *IEEE Trans. Nucl., Sci.* vol.50, 2003, 1238-1252

[2] MANNA, G.: Human and Organizational Factors Nuclear Installations: *JRC Scientific and Technical Reports* 2007

[3] HART, S.G. and STAVELAND, L.E.: Development of NASA-TLX (Task Load Index); Results of empirical and theoretical research. In P.A.Hancock and N. Meshkati (eds), Human Mental Workload, North-Holland, 1998, 139-1