A Study on Work Load Evaluation Method and Quantitative Evaluation Method for Engine-room Resource Management training

Tatsuro Ishida [a*], Takashi Miwa[a], Makoto Uchida [a]

[a] IAMU Member, Graduate School of Maritime Sciences, Kobe University, Kobe, 658-0022, Japan
[*]Corresponding author. E-mail: 152w951w@stu.kobe-u.ac.jp

Keywords: Engine-room Resource Management (ERM), Work Load Evaluation Method, Quantitative Evaluation, Non-technical skills

ABSTRACT

This study aims to propose work load evaluation method and quantitative evaluation method of trainee's non-technical skill regarding ERM training. The necessity of work load evaluation during simulator training exercise is considered to evaluate trainee's performance. The effect of the trainee's work load during training exercise is evaluated by objective evaluation using VACP (Visual, Auditory, Cognitive, Psychomotor) method and heart rate data comparing subjective evaluation NASA-TLX (Task Load Index). We studied the difference of trainee's work load and biological reaction during training exercise. VACP work load scale is modified for ERM training exercise. For quantitative evaluation, we used IMO model course 2.07 2017 Edition evaluation form: behavior markers for non-technical skills. Evaluation criteria is based on sample evaluation form and evaluation markers are modified. This paper report the result of both evaluation method from recorded video and audio data of ERM training evaluation experiment.

1. Introduction

In 2010 Manila amendments of Standard of Training Certification and Watch-Keeping(STCW) convention, the requirement concerning Engine-room Resource Management (ERM) have been introduced into mandatory requirement for engineer and full implementation is required on 1st January 2017. Amended requirement increased needs of simulator training to evaluate engineer's non-technical skills. To carry out ERM training and evaluate non-technical skills, IMO Model course 2.07 Engine-Room Simulator 2017 Edition program was introduced to provide guidance with a view to supporting training providers.

Even though model course program is designed to provide flexibility so as to allow training providers to adjust the course program to the needs of trainees. Evaluation forms are also suggested to develop by instructor in accordance with simulator facilities and functions.

According to these background, maritime education institute and universities are developing suitable ERM training program and effective evaluation method of the training exercise [1] [2].

This study aims to propose work load evaluation method and quantitative evaluation method of trainee's non-technical skill regarding ERM training. Evaluation of trainee's work load during ERM training is to understand validity of developed ERM training program and to know whether the training is effective or not when work load of trainee is too high or too low.

Also the necessity of work load evaluation during simulator training exercise is considered to evaluate trainee's performance. Relation of workload and task performance is considered in

several industry such as aviation industry and automotive industry [3].
The effect of the trainee's work load during training exercise is evaluated by objective evaluation using VACP (Visual, Auditory, Cognitive, Psychomotor) method [4] and heart rate data comparing subjective evaluation NASA-TLX (Task Load Index) [5]. We studied the difference of trainee's work load and biological reaction during evaluation experiment. For quantitative evaluation, we used IMO model course 2.07 2017 Edition evaluation form. Evaluation criteria is based on sample evaluation form and evaluation markers are modified.

2. Methodology

Work load evaluations are categories to subjective and objective evaluation. Subjective evaluation is carried out by collecting trainee's opinion using questionnaire. Subjective evaluation may pointed out that personal sense and feelings are influence to the result of questionnaire and certain results may not be reproduced. To avoid this concern, using rating scale such as NASA-TLX are developed. NASA-TLX is developed by human performance group at NASA and widely used for subjective work load evaluation.
Objective work load evaluation research is carried out by using biological data such as heart rate data to measure navigator's work load [6]. Using heart rate data for objective evaluation was effective in finding the difference of the work load. However, heart rate data may affected by physical activity and there is a concern in work load evaluation using only the heart rate data. Therefore, work load evaluation is performed by the proposed objective evaluation VACP (Visual, Auditory, Cognitive, Psychomotor) method. As a reference of the validity of the work load estimation, we also compared the workload estimation with the subjective evaluation NASA-TLX. Detail of VACP method is explained in the following.

2.1 Work load evaluation VACP method

The basic idea VACP method is to divide trainee's information processing resource into four different channels: Visual, Auditory, Cognitive, Psychomotor (VACP) and express the amount of resource consumed as a scale. The total VACP scale is calculated as quantified work load by summing up each consumed scale.
The VACP method originates from the study of workload components in the operation of light weight helicopter [4]. They recognized four workload components and evaluation criteria description and scale are defined. There are research adopting VACP method to evaluate work load in different industry such as Anesthesiology field [7].
Developed VACP scale for ERM training is shown in Table 1. The standards of each scale and description is based on original research and action code is modified according to the characteristics of engine room operation and ERM training. In this research, work load of trainees are evaluated using this VACP scale from recorded video and audio data of ERM training evaluation experiment.
In order to quantify the work load of the trainee's tasks, the content of the tasks to be quantified by comparing with the description of the VACP scale to determine the amount of consumed resources. For example, when the subject verbal confirmed orders from C/E by transceiver, the auditory weight is A3. VACP scale should include a P1 since the trainee have to use on hand to hold the transceiver. Furthermore, notice that A4 is defined as the Auditory Interpretation of high attentional auditory signal, such as alarm signals, which happens in

blackout situation. Each scale is summed up in task execution timing. In this research, VACP scale were summed each for a four-second timing.
Main tasks occurred in evaluation experience and VACP scale of each task is shown in Table.2.
In addition to these main tasks, other people's conversations that can be heard indirectly is added as Auditory Disturbance A1 to the work load quantify. P1 was added as a psychomotor scale when push alarm stop button or pointing out of confirmation and so on.

Table 1 VACP Scale

VISUAL		
Scale.	Description	Action Code
V1	Visual Detection	Gaze<=2s
V2	Visual Discrimination	Gaze >2s, Static
V3	Visual Trucking	Gaze >2s, Dynamic, Monitor + Gauge
V4	Visual Read, Searching,	Dynamic, Visually high attentive. Find out target
AUDITORY		
Scale	Description	Action Code
A1	Auditory Detection	Digital Signal, Sound.
A2	Auditory Verification	Auditory Feedback. Answer back
A3	Auditory Decoding	Speech, Semantic Content, Order, Report
A4	Auditory Interpretation	Sound patterns, Alarm sound
A1	Auditory Disturbance	Non-directive
COGNITIVE		
Scale	Description	Action Code
C1	Automatic, Alternative Selection	e.g. Open valve as ordered, Answer back
C2	Sign/Signal Recognition	e.g. Alarm occur, Alarm information
C3	Evaluation/Judgement (Single Aspect)	e.g. Situation evaluation, Condition Judgement
C4	Evaluation/Judgement(Several aspect)	e.g. Alarm situation evaluation,
PSYCHOMOTOR		
Scale	Description	Action Code
P1	Discrete Actuation	e.g. Push Button, Talk, , monitoring, Point out
P2	Continuous Adjusting	e.g. Telegraph operation
P3	Symbolic Production	e.g. Writing, typing
P4	Convergent Multiple Operations	e.g. Walk, Check condition

Table 2 Main tasks VACP scale

Task Name	Visual	Auditory	Cognitive	Psychomotor
① Alarm Occur	2	4	4	0
② Order	0	3	0	2
③ Answer Back	0	2	2	0
④ D/G Auto Start	3	0	2	4
⑤ M/E Reset	3	1	3	2
⑥ Restart Boiler	3	0	3	1
⑦ Pump Auto Start	3	0	3	4
⑧ Steering Auto Start	3	0	3	4
⑨ E/R Round check	3	0	4	4
⑩ M/E Restart	3	1	3	2

2.2 Quantitative evaluation method for Non-technical skill

In IMO model course 2.07 2017 Edition sample evaluation form: Behavior markers for non-technical skills consist four non-technical skill evaluation criteria which are shown in Table.2. Evaluation markers in sample form was described by four scale score A,B,C,D. In this report, evaluation criteria is followed this sample evaluation form and evaluation markers were modified to 90-0% and original four scale score was calculated by weight criteria after completed evaluation by A,B,C,D. Weight criteria is defined by score 4 to 1 meaning that the performance is Very good, good, fair and poor. Two type of evaluation markers were conduct to propose reliable and valid quantitative evaluation method for non-technical skill.

An evaluator team, consisted 14 evaluator who have different onboard experience, gave a two type of evaluation marker to evaluate based on their personal judgement.

Table 3 Evaluation criteria

Evaluation criteria
1. Understanding of roles and responsibilities in operating plant machinery as a part of an engineering watch
2. Instruction, report, answerback and other communication patterns
3. Leadership and assertiveness
4. Situational awareness and notification of any doubt

3. Evaluation experiment

3.1 Experiment device

The experiment was conducted using Marine Engine Plant Simulator (MEPS), installed at Kobe University Maritime Science Graduate School. MEPS shown in Figure 1 consist of control room system, Engine room system and instructor system. Two fixed point cameras for video recording are provided on the control room and engine room side. Recorded evaluation

experiment data is shown in Figure 2. A microphone at control room is used to record audio data. Heart rate data was collected using a wristband type optical heart rate meter (Polar A 360). This optical heart rate meter calculates heart rate by beat per minutes in every time step and this calculated heart rate is expressed as heart rate (bpm) in this report.

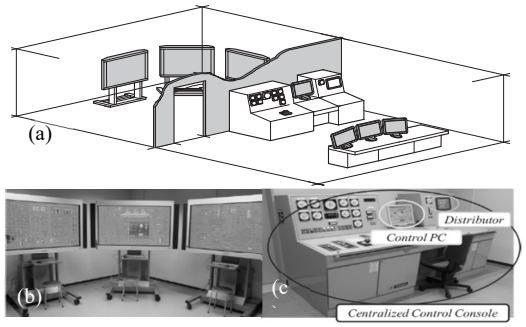

Figure 1 MEPS (a) outline of MEPS; (b) Engine room; (c) control console.


Figure 2 Recorded evaluation experiment data

3.2 Experiment contents

The evaluation experiment scenario is modified and proposed for MEPS environment, based on IMO Model Course 2.07 sample exercise ERM training blackout roll play.

The applicability of the proposed method for evaluating work load and non-technical skill quantitative evaluation regarding ERM training was demonstrated by proposed evaluation experiment. Four students (participant 1 to 4) who have onboard experience took the roles of engineer under the blackout scenario. Scenario was carried out twice with different role for each participant.

A scenario starts with occurrence of a blackout (power loss) due to an emergency stop of Turbo Generator and Main Engine emergency stop due to L.O. Low press. Then Diesel and emergency generator start automatically, power supply and recovery work at engine room will be carried out. After Confirmation of each machinery, restart operation of Maine Engine will be the end of scenario. Communication between control room and the engine room are carried out by transceiver during the experiment.

4. Result of Evaluation experiment

4.1 Result of Work load evaluation

Result of four participant work load evaluation are shown in Table 4.

By comparing the role of each participant into C/E, 1/E and 2/E, 3/E, the work load of control room side and engine room side are verified. Figure 3 to 10 shows result of VACP SUM and heart rate (HR) of each participants. Circled number describe main tasks shown in Table 2.

Result of VACP evaluation method indicates quantified work load of control room task are higher than engine room task. This result can be considered that control room role had higher consumption of each VACP resource by reacting for alarm occurrence and order for recovery tasks. Engine room tasks had higher psychomotor resource consumed by check condition and operating valves but other auditory and cognitive resource consumption were smaller compare to control room. Heart rate data of all participant are higher in the control room task than

engine room task. Also Subjective evaluation by NASA-TLX also resulted control room task required higher work load to all participant. The correlation between the heart rate data and the VACP evaluation quantified work load resulted positive correlation with a correlation coefficient of 0.4 or more for participant1,3 and 4 in the control room side results, and for participant 2 resulted weak positive correlation of 0.21.

These results may consider that higher work load of control room task explained higher performance were required to these task responses.

Proposed and developed VACP scale for ERM training provides a useful way to evaluate trainee's work load objectively from recorded video and audio data of evaluation experiment.

Table 4 Comparison of NASA-TLX, HR Average, VACP SUM and HR-VACP Correlation

		Task Location: Control Room					Task Location: Engine Room			
Participant	Role	NASA-TLX	HR Average(bpm) - HR Rest(bpm)	VACP SUM	HR-VACP Correlation	Role	NASA-TLX	HR Average(bpm) - HR Rest(bpm)	VACP SUM	HR-VACP Correlation
1	C/E	14.87	11.29	549	0.41	2/E	13.20	6.03	476	0.23
3	C/E	16.20	13.37	531	0.42	2/E	9.80	6.44	469	0.08
2	1/E	7.33	11.64	429	0.21	3/E	5.47	6.11	388	0.01
4	1/E	14.60	15.17	420	0.42	3/E	11.13	6.32	386	0.47

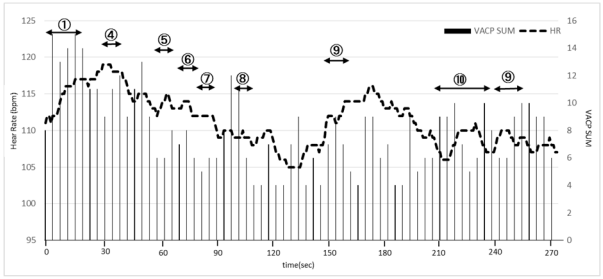

Figure 3 VACP SUM and HR of Participant1:C/E (1st Scenario)

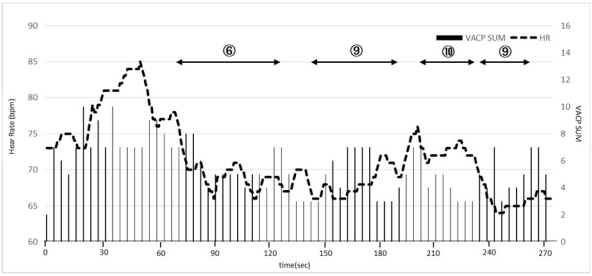

Figure 6 VACP SUM and HR of Participant4:3/E (1st Scenario)

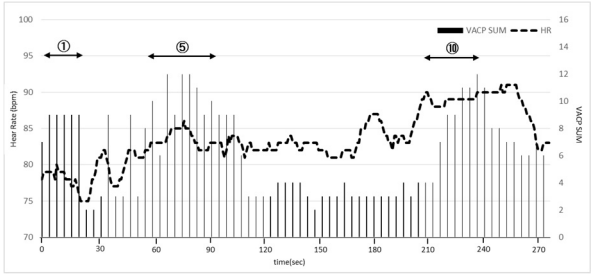

Figure 4 VACP SUM and HR of Participant2:1/E (1st Scenario)

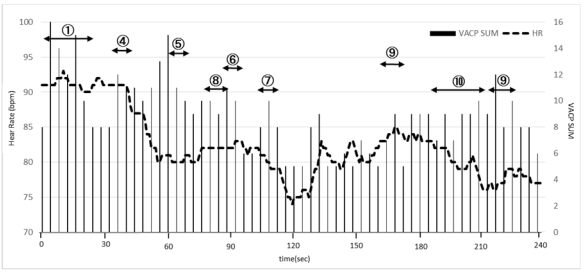

Figure7 VACP SUM and HR of Participant3:C/E(2nd Scenario)

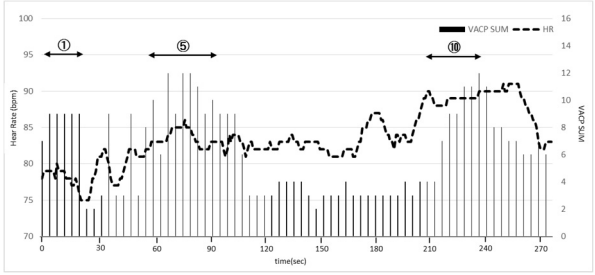

Figure 5 VACP SUM and HR of Participant3:2/E (1st Scenario)

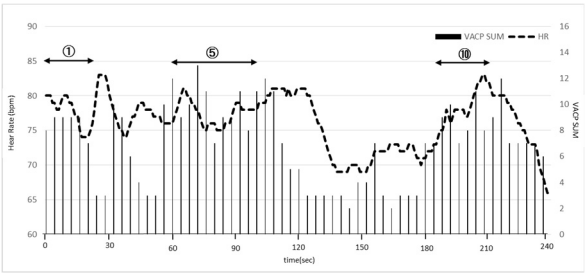

Figure 8 VACP SUM and HR of Participant4:1/E(2nd Scenario)

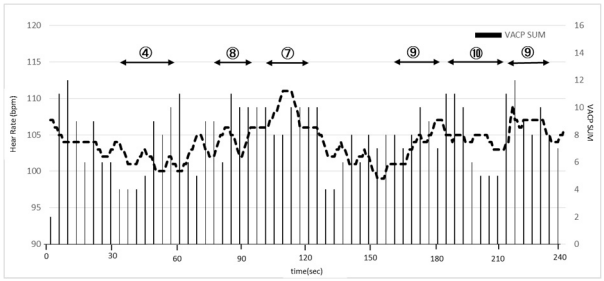

Figure 9 VACP SUM and HR of Participant1:2/E(2nd Scenario)

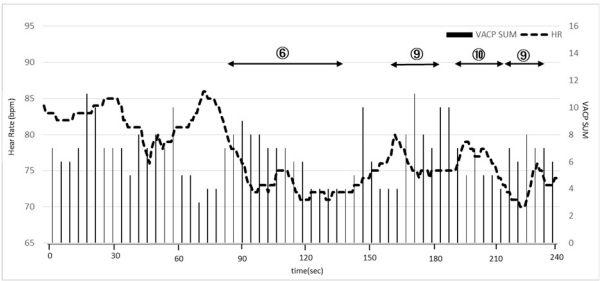

Figure 10 VACP SUM and HR of Participant2:3/E (2nd Scenario)

4.2 Result of Quantitative evaluation for Non-technical skill

Quantitative evaluation of non-technical skill was carried out by evaluator team and result of averaged evaluation marker score is shown in Table 5. Also Fig 11 shows each participants average score of control room side and engine room side are verified

From result of participant 1, role difference can be described that control room side task had higher non-technical markers specially by comparing evaluation criteria 3, 2/E task was lower score compare to C/E in both four-scale score and percentage score. Similar result can be seen for participant 3 who took the same role of 2/E and C/E. Difference between 2/E and C/E can be described by qualitative evaluation.

In other hand, Participant 2 and 4 who took a role of 1/E and 3/E had different tendency of result. Participant 4 resulted evaluation markers scored for 1/E engine room task was slightly higher compare to 3/E engine room task, but participant 2 shows 3/E evaluation markers was higher than 1/E in all evaluation criteria. Result of this quantitative evaluation shows C/E and 2/E had more clear difference of non-technical skills compare to the role of 1/E and 3/E. Propose quantitative evaluation clarified the difference of evaluation markers score for each participant role and each evaluation criteria.

Table 5 Average of evaluation marker score

Participant	Evaluation criteria	C/E		2/E	
		ABCD	%	ABCD	%
1	1	88.39	85.00	83.57	80.71
	2	88.39	85.00	85.18	85.00
	3	88.39	84.29	75.54	81.43
	4	80.36	80.71	80.36	80.00
Participant	Evaluation criteria	1/E		3/E	
		ABCD	%	ABCD	%
2	1	69.11	72.14	81.96	75.71
	2	77.14	73.57	81.96	80.00
	3	64.29	62.86	67.50	73.57
	4	67.50	70.71	73.93	76.43
Participant	Evaluation criteria	2/E		C/E	
		ABCD	%	ABCD	%
3	1	73.93	70.00	73.93	72.86
	2	70.71	74.29	80.36	80.71
	3	65.89	65.71	80.36	77.14
	4	67.50	67.86	72.32	75.00

Participant	Evaluation criteria	3/E		1/E	
		ABCD	%	ABCD	%
4	1	69.11	66.43	70.71	70.00
	2	70.71	71.43	75.54	77.14
	3	62.68	61.43	67.50	67.86
	4	61.07	63.57	69.11	72.14

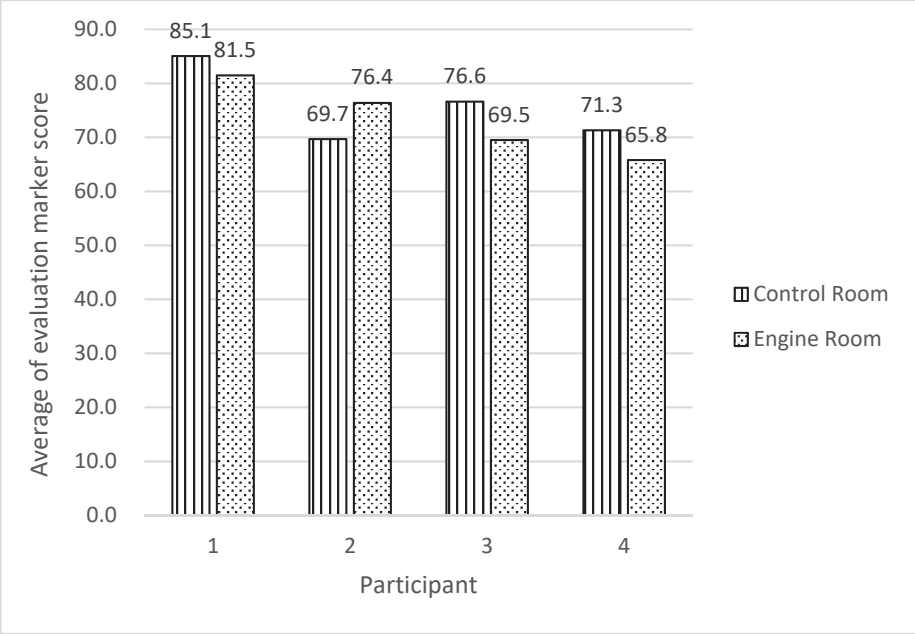

Figure 11 Average of evaluation marker score

4.3 Overall Result

Each result of work load evaluation and quantitative evaluation of non-technical skill are explained and discussed. From result of work load evaluation, control room task showed higher work load compare to engine room task for all participants. When difference of work load was occurred by task response as work performance, result of quantitative evaluation of non-technical skill may explain more precise for each participant role.

For participant 1 and 3 who took role on C/E and 2/E had clear difference of performance required and as mentioned in result of quantitative evaluation, evaluation criteria 3 had biggest difference for both participants.

For participant 2 and 4 who took role on 1/E and 3/E work performance also resulted control room task performance are higher compare to engine room task. But result of quantitative evaluation for non-technical skill for participant 2 indicated higher evaluation criteria score on 3/E role. When we compare result of evaluation criteria 3 score difference for C/E, 2/E and 1/E, 3/E, result of 1/E, 3/E score difference was lower score. Also result of work load VACP SUM indicate 1/E, 3/E difference were lower compare to C/E, 2/E.

5. Conclusion

This study focused on work load evaluation method and quantitative evaluation method of trainee's non-technical skill regarding ERM training. The proposed method of workload

evaluation provides a useful way to measure trainee's work load objectively and required performance difference of control room task and engine room task regarding to ERM training was expressed. Result of correlation between the heart rate data and the VACP evaluation quantified work load pointed out higher accuracy of work load evaluation in control room task. In other hand, Engine room task resulted variation for each participants regarding evaluation timing of VACP and evaluation scale. Evaluation scale consider possible extension to develop scale action code for each control room and engine room task. Also further verification for evaluation timing should be considered. In this study, Work load of each participants are evaluated by comparing subjective NASA-TLX and objective VACP scale and heart late. In the further studies additional objective assessment such as EEG measurement would be considered to interpreting the cognitive resource of the participants.

Quantitative evaluation method is also indicated each trainee's non-technical skill during ERM training. Evaluation markers proposed in this method was applied to compare evaluation criteria qualitatively and difference in each role and each criteria were discussed.

Proposed evaluation experiment scenario was developed with assuming that control room tasks require higher non-technical skill compared to engine room task.

Difference of C/E role and 2/E role resulted that proposed evaluation scenario and quantitative evaluation method can demonstrate role behavior of ERM training without personal bias. Result of 1/E and 3/E role was discussed from both performance and non-technical skill point of view. Both result may consider that role difference effecting performance and non-technical skill evaluation between 1/E and 3/E was not as clear as C/E and 2/E in this ERM training evaluation experiment. Difference in both results considered possible extension of proposed evaluation experiment ERM training scenario. Also further studies to the proposed method would be to consider evaluation criteria.

Main idea of the proposed method is to evaluate work load objectively and also evaluate non-technical skill qualitatively to discuss the performance of trainee and conduct sufficient ERM training. Both evaluation method and developed evaluation experiment scenario described certain effective result for evaluating ERM training by objective and quantitative point of view.

REFERENCES

[1] Yanbin Wu, Takashi Miwa, Kaoru Shimamoto, Makoto Uchida, Development of quantitative team performance evaluation method for ERM, *WMU Journal of Maritime Affairs*, October 2015, Volume 14, Issue 2,PP 333-347.

[2] Sachiyo Horiki, Mitsuo Kojima, Tatsuro Tsukamoto,Engine-Room Simulator for ETM/ERM Training at Faculty of Marine Technology, Tokyo University of Marine Science and Technology, *Journal of JIME*, 2014, Volume 49,Issue 3, PP 370-372

[3] Fátima Pereira da Silvaa, Mental Workload, Task Demand and Driving Performance: What Relation?, *Procedia - Social and Behavioral Sciences*, 2014, Volume 162, PP 310-319

[4] Bierbaum, Carl R., Sandra M. Szabo, and Theodore B. Aldrich. Task analysis of the UH-60 mission and decision rules for developing a UH-60 workload prediction model: Summary report. Vol. 1. US Army Research Institute for the Behavioral and Social Sciences, 1989.

[5] Hart, S. G., & Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, pp.139-183.1988.

[6] Koji Murai, Yuji Hayashi, Noriko Nagata, Seiji Inokuchi, Mental Workload of Ship's

Navigator, A Few Comments on Heart Rate Variability during Navigational Watch Keeping, Knowledge-Based Intelligent Information and Engineering Systems, 7th International Conference, 2003, Proceedings, pp 897-903
[7] Pfeffer, Stefan, Patrick Decker, Thomas Maier, and Eric Stricker, Estimation of operator input and output workload in complex human-machine-systems for usability issues with iflow, In International Conference on Engineering Psychology and Cognitive Ergonomics, 2013, pp. 167-176.