



## **Effectiveness of Self-Evaluation, Peer Evaluation and 2nd-Step Self-Evaluation- Covering Anchoring Training in Maritime Education and Training**

**Akihiro Nunome**

National Institute of Technology, Toyama College, Toyama, Japan

**Yoshiaki Kunieda**

National Institute of Technology, Toyama College, Toyama, Japan

**Yoshihiro Majima**

Japan Agency of Maritime Education and Training for Seafarers  
Yokohama, Japan

**Hideyuki Kashima**

Tokyo University of Marine Science and Technology, Tokyo, Japan

### **ABSTRACT**

**Evaluation is an essential part of education and training, and it can be used in maritime education and training to help learners organize their knowledge and improve their skills, as well as to improve education and training methods. In this study, self-evaluation and mutual evaluation were conducted during anchoring training on the training ship Shioji Maru belonging to the Tokyo University of Marine Science and Technology. In a survey of students trained in 2020 and 2021 regarding the acquisition of knowledge and skills, 94.5% of students rated themselves as “very effective” or “effective” in their self-evaluation and 92.0% of students rated themselves as “very effective” or “effective” in their mutual evaluation. Comparing the self-evaluation scores with the mutual evaluation scores, it was found that the mutual evaluation scores tended to rank higher than the self-evaluation scores. This is thought to be due to a lack of confidence in one’s own ship handling skills, which leads to harsh evaluations of oneself and more lenient evaluations of others. It was also found that the higher the instructor’s evaluation score, the smaller the difference between the self-evaluation score and the instructor’s evaluation score. Students with higher scores in the instructor’s evaluation were more confident in their ship handling skills, which is thought to indicate that they can evaluate themselves more accurately. On the other hand, self-evaluation was conducted at an early stage immediately after the training, and the bridge operation team and the entire team also conducted the self-evaluation again after the debriefing. In other words, a 2nd-step self-evaluation was conducted through two evaluations conducted at different times. We show the results of a qualitative analysis of the students’ impressions and opinions of these self-evaluations and peer evaluation using the steps for coding and theorization (SCAT) method.**

**Keywords:** Self-Evaluation; Mutual Evaluation; 2nd-step self-evaluation; SCAT (Steps for Coding and Theorization); Anchoring Training.

## INTRODUCTION

Trade and logistics have continued to develop in tandem with the world economy. Not only in island countries like Japan but also in many other countries around the world, logistics is mainly carried out by maritime transportation using ships that can transport large volumes of cargo at a low cost. However, a large number of maritime accidents, such as collisions and groundings, continue to occur. Taner Albayrak, et al. described many maritime accidents are caused by human factors, pointing to the importance of education and training of seafarers [1].

In the waters around Japan, according to Japan Coast Guard statistics, about 2,000 maritime accidents occur annually. In 2020, the number of vessel accidents recognized by the Japan Coast Guard was 1,954 in one year [2]. By type of marine accident, collision was the most common, accounting for 44% of all cargo ship marine accidents. The most common cause of collisions was "Improper ship handling" followed by "Insufficient lookout." These are all human factors, and human factors account for about 80% of all maritime accidents. Possible measures to prevent collisions caused by human factors include improving and reinforcing the education and training of seafarers, developing and promoting equipment and facilities to prevent collisions, and establishing and implementing appropriate management and systems to prevent accidents. This study focuses on education and training to eliminate the improper ship handling that can lead to collisions.

In maritime education and training, knowledge and skills are acquired through classroom instruction and onboard training. In classroom teaching, lectures are the norm, but to optimize effectiveness, a combination of exercises and experiments is used. On the other hand, training on board is conducted by the students themselves and is viewed as an instance of active learning, and considered to be more effective than lectures in which students simply listen to a one-way talk. While shipboard training is more effective, simply practicing what is taught is not enough. It is important to be able to do what has been taught or decided, but more effective training is needed to acquire independent thinking and problem-solving skills. Kashima et al. showed that anchoring training is effective for acquiring ship handling skills during training on actual vessels [3]. Kunieda et al. also showed that anchoring training is effective in developing students' ability to think actively and address problems on their own [4].

In this paper, to conduct more effective education and training in anchoring training, we conducted (1) a self-evaluation immediately after the training, (2) a mutual evaluation to evaluate each other, and (3) a second-step self-evaluation after debriefing, and examined the characteristics and effectiveness of each from questionnaires and comments, as well as improving the already proposed "Group work - Training by actual ship - Group work - Presentation (GTGP)" training model [5].

## ANCHORING TRAINING

Anchoring training, which is performed by student teams without instructor assistance, is a suitable exercise to improve ship handling skills through various maneuvers. In anchoring training, at first, the students heave up the anchor and sail a planned route. Then, after passing specified waypoints, the students anchor at the appointed anchorage. Anchoring training is suitable not only for learning maneuvering procedures based on actual performance but also for learning various elemental techniques for maneuvering, for information exchange and management skills and for improving the abilities related to those skills [3]. This paper

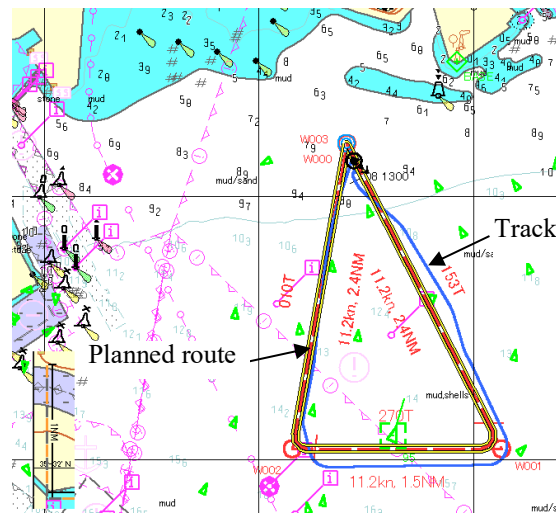
examines the effects of student self-evaluation, mutual evaluation, and then a second-step self-evaluation in anchoring training for third-year students of the Department of Maritime Systems Engineering at Tokyo University of Marine Science and Technology in 2020 and 2021. The number of students participating in anchoring training is shown in Table 1.

**Table 1. Number of students participating in anchoring training by time of year**

Time of year	Second half of 2020	First half of 2021	Second half of 2020	Total
Number of students	60	48	30	138

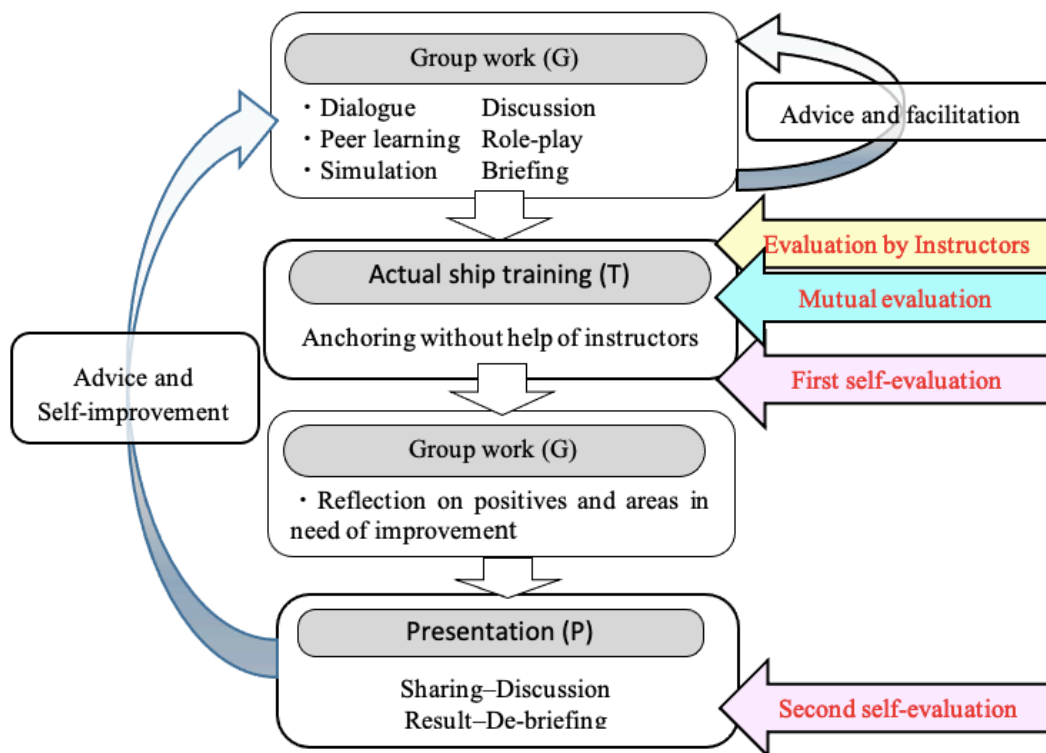
Anchoring training is generally performed by a team of four persons, each with a defined role: the captain (role of captain-ROC), the first officer (role of the first officer-RO10), the third officer (role of the third officer-RO30) and the quartermaster (role of the quartermaster-ROQ). Since the training is conducted in teams, the group work involves peer-to-peer learning as indicated by Mochizuki and Mizokami [6][7]. Thus, anchoring training is considered effective group work, with a flow as follows:

- (1) Taking the lead, the ROC develops a navigation plan for anchoring. This active student planning of ship handling is designed to develop the ROC's leadership skills and create opportunities for peer learning. The ROC explains the navigation plan to teammates and instructors, who then give advice and feedback to refine the plan. After making corrections to the navigation plan, the ROC briefs team members and instructors, who then check their notes and roles: (Group work) G.
- (2) Conduct actual ship training: (Training by actual ship) T.
  - a) The ROC positions the leaving anchorage station and, before heaving anchor chains, directs the RO30 to prepare the main engine and then directs the heaving up of the anchor.
  - b) When the anchor is aweigh, the ROC sets off on a predetermined course using the main engine and rudder.
  - c) The ROC corrects the course appropriately to ensure that the planned route is being navigated. The ship then passes two scheduled waypoints and navigates the predetermined route.
  - d) The ROC slows the main engine, adjusts the course and stops the ship by applying the main engine to the sternway to ensure it is positioned correctly at the appointed anchorage.
  - e) The ROC lets the anchor go at the appointed anchorage, lets out the cables to a predetermined length, stops the main engine and finally dismisses the anchoring station.
- (3) Shortly after the training ends, students conduct a self-evaluation based on a rubric.
- (4) Each team then discusses the anchoring training, that is, its positive aspects and those areas needing improvement: (Group work) G.
- (5) Each team member presents their ship handling notes, while the others engage in active listening and thinking. Instructors' comments based on the evaluation rubric: (Presentation) P.
- (6) Perform the second-step self-evaluation.
- (7) Students write their impressions and organize the content of the training.



**Figure 1. An example of a planned route and track**

An example of a planned route and track in the training area is shown in Figure 1. And figure 2 shows the GTGP training model for anchoring training proposed by the authors. In this training flow, the instructor and students from other teams playing the ROC evaluated the trainee students using a rubric evaluation list during the training. The trainee students conducted the first self-evaluation as soon as possible immediately after the actual ship training. In addition, after the practical training, a debriefing was conducted as a group exercise in which the four members of the ship’s bridge team discussed what went well and what could be improved upon, and the results of the group work were presented to the whole group, after which the self-evaluation was conducted again (the second-step self-evaluation).



**Figure 2. Improved GTGP training model**

## EVALUATION METHOD

In anchoring training, two instructors who are experienced captains of a large training ship conduct the training evaluation; the same two instructors performed consecutive evaluations in 2020 and 2021. For evaluation of anchoring training, 14 specific evaluation items were determined and the rubric evaluation, which sets evaluation criteria for each item, was adopted. Evaluation criteria were scored across four stages: 4 points (90 or more out of 100 points), 3 points (80–90 points), 2 points (60–80 points), and 1 point (fewer than 60 points). Self-evaluation by the students themselves and mutual evaluation by other students were also conducted using the same rubric evaluation list.

Rubrics are essential in the evaluation of learning, and the following are the reasons for this [8].

- 1) A rubric evaluates student performance using many perspectives.
- 2) A rubric demonstrates a shared viewpoint and standard of evaluation between students and instructors.
- 3) Different evaluation levels can serve as the students' desired values.
- 4) A rubric evaluates based on a clear standard rather than a teacher's intuition.
- 5) A rubric can standardize the evaluation items and criteria of judgment between two or more teachers. As a result, it can evaluate with high validity and reliability.

Moreover, a rubric place the educational objective before the student, thereby clarifying the aim of the lesson. Studies show that rubric evaluations are effective. In this study, the rubric evaluation list that the instructors and students used contained the following 14 items and was evaluated by a four-step evaluation indicator:

- a: Procedure for heaving up anchor
- b: Acceleration and course setting after heaving up anchor
- c: Lookout (visually finding)
- d: Lookout (visually continuous monitoring)
- e: Course setting
- f: Altering course
- g: Give-way/stand-on maneuvering
- h: Grasping the ship's position
- i: Anchor position
- j: Anchoring procedure
- k: Decrease in speed
- l: BRM/BTM (Information sharing)
- m: BRM/BTM (Intention sharing)
- n: Overall impression

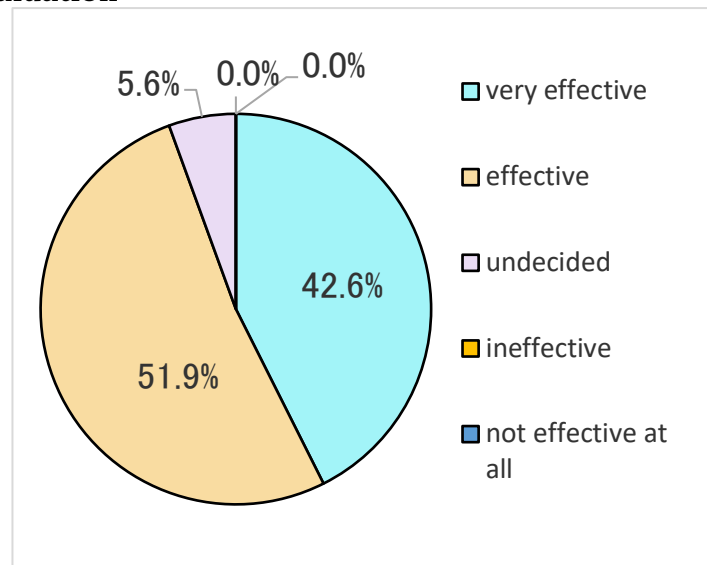
Table 2 shows an extract from the rubric evaluation list. The evaluation criteria are as specific as possible so that they can be judged by numerical values.

**Table 2. The rubric evaluation list (extract)**

No	Evaluation Item	Excellent (90%-full marks)	Good (70%-90%)	Passing mark (60%-70%)	Failure (Less than 60%)
1	Procedure for heaving up anchor	The procedure for heaving up anchor was well understood and clearly ordered without error.	The procedure for heaving up anchor was understood, and although there was one mistake, it was clearly ordered.	The procedure for heaving up anchor was not fully understood, and there were about 2-3 mistakes, which could have been clearly ordered, but were not.	The procedure for heaving up anchor was not fully understood, and most of the time, the ship was ordered relying on memos.
2	Acceleration and course setting after heaving up anchor	The speed increase procedure and the setting of the course after heaving up anchor were completed accurately and without error.	The speed increase procedure and the setting of the course after heaving up anchor were completed with only one mistake.	There were 2-3 instances of mistakes in the speed increase procedure and setting the course after heaving up anchor, but he managed to complete the task.	The procedure for setting the speed increase and the course after heaving up anchor was not done well enough and no fewer than four mistakes were made.
3	Lookout (Visually, detection)	Visual lookouts were properly conducted without error.	A visual lookout was conducted with about one error.	There were two or three errors in the visual lookout, but he managed to implement them.	Continuous lookout, such as direction change, was not carried out three or more times. It cannot be said that a sufficient lookout was carried out.

## RESULTS AND CONSIDERATIONS

### Self and Mutual Evaluation

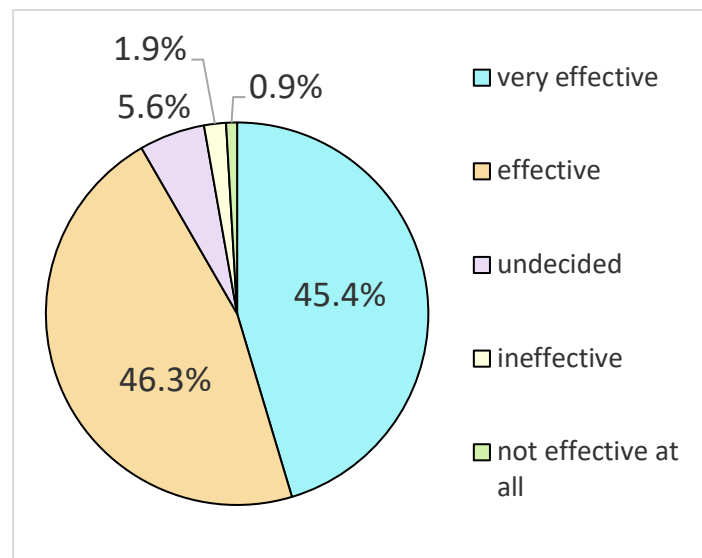


**Figure 3. Effectiveness of self-evaluation**

A questionnaire survey was conducted on self-evaluation and mutual evaluation using a five-point method which students were asked to rate as follows: (1) very effective, (2) effective, (3) undecided, (4) ineffective, and (5) not effective at all.

Figure 3 shows the results of a questionnaire that asked about the effectiveness of the self-evaluation. Of the students, 94.5% responded “effective” or very effective. In addition, the main impressions (comments) of the students on the self-evaluation are as follows.

- (1) By self-evaluating, I was able to identify what I did wrong.
- (2) I was able to clarify what was good and what was not.
- (3) It was good because it recorded my reflections and positive aspects so I would not forget them.
- (4) I was able to review the content of my training, which improved my own retention of the training.
- (5) I was able to recognize the problems that need to be addressed to make the next step.



**Figure 4. Effectiveness of mutual evaluation**

Figure 4 shows the results of a questionnaire that asked about the effectiveness of mutual evaluation. The majority of students (91.7%) responded with “effective” or “very effective,” while 0.9% (1 person) responded, “not effective at all” and 1.9% (2 persons) responded, “not effective.” All of these negative respondents were in charge of recording ship operations and were therefore “too busy recording to evaluate others” or “could not evaluate others effectively.”

In addition, the main impressions (comments) of the students on the mutual evaluation are as follows:

- (1) I was able to observe from a different perspective from my own (self-evaluation), and furthermore, I understood what I need to improve on.
- (2) Someone was able to evaluate my training content objectively and I appreciate the shortcomings better
- (3) I was able to view the practical training objectively and from a bird’s eye view, which enabled me to work calmly during the actual training.
- (4) It was good that I was able to get a better picture and apply it to my group’s practical training.
- (5) Evaluating others helped me to recognize mistakes and to use them as a reference point for my own ship handling.

The average self-evaluation score was 2.94, while the average mutual evaluation score was 3.44, indicating that the mutual evaluation scores tend to be higher than the self-evaluation scores. This is thought to be due to a lack of confidence in students’ own ship handling skills, which causes them to be hard on themselves and more lenient in their evaluation of others.

Moreover, it was found that the higher the instructor’s evaluation score, the smaller the difference between the self-evaluation score and the instructor’s evaluation score. Table 3 shows the average evaluation scores and average self-evaluation scores of the top five and bottom five evaluation scores by the instructors. For the top five students, the difference between the average instructor’s evaluation score and the average self-evaluation score was



0.19. On the other hand, for the bottom five, the difference between the average instructor’s evaluation score and the average self-evaluation score was 0.44—a larger point difference. In all of the top five, the average self-evaluation score was lower than the average evaluation score by the instructor. In contrast, for the bottom five, there were times when the average self-evaluation score was both higher (three cases) and lower (two cases) than the average evaluation score by the instructor, and the absolute values were higher. These results suggest that students with higher average evaluation scores awarded by instructors are more confident in their ship handling skills, and thus, can evaluate themselves more accurately.

**Table 3. The difference between self-evaluation scores and the instructor’s evaluation scores for the top five and the bottom five**

Top 5			Bottom 5		
Average instructor evaluation score	Average self-evaluation score	Average score difference	Average instructor evaluation score	Average self-evaluation Score	Average score difference
3.79	3.64	0.14	2.61	2.86	0.25(-)
3.50	3.07	0.43	2.68	3.29	0.61(-)
3.50	3.29	0.21	2.71	2.50	0.21
3.43	3.36	0.07	2.82	2.29	0.54
3.36	3.29	0.07	2.86	3.43	0.57(-)
Average of Score Difference		0.19	Average of Score Difference		0.44

### 2<sup>nd</sup>-step Self-evaluation

For the anchoring training in FY2021, in addition to the self-evaluation immediately after the actual shipboard training, another self-evaluation was conducted after the group work of the reflection and the presentation by each group (second-step self-evaluation). Students’ impressions of the second-step self-evaluation were analyzed by the steps for coding and theorization (SCAT), a method of qualitative analysis. SCAT is a method of analyzing qualitative survey data, etc. Data obtained through interviews, etc. are coded (coding), and the codes are used as the basis for theorizing. The SCAT describes segmented data in a matrix, which collates:

- (1) Words and phrases to focus on in the data
- (2) Words outside the data to paraphrase it
- (3) Words and phrases to describe it
- (4) Themes and constitutive concepts that emerge from this.

This is an analytical method consisting of a four-step coding process in which codes are considered and attached in this order, and a procedure for deriving (4) themes and constitutive concepts, thereby describing a storyline, from which a theory is described. This method is also useful for analyzing relatively small amounts of qualitative data, such as data from a single case or the free-text sections of a questionnaire. In addition, it has explicit and formulaic procedures, making it easy for ingénue researchers to embark on the analysis. Research using SCAT has been published in a truly diverse range of fields, including educational technology, the sociology of education, early childhood education, childcare education, media literacy education, Japanese language education, legal education, clinical psychology, medical education, clinical research, pharmacology, nursing, sports business research, and human services research. It is

used in many fields because of the clarity of its analytical procedures and the fact that it may apply to small data sets [9].

The results of the SCAT analysis of the students' impressions of the second-step evaluation in anchoring training are shown in Table 4.

**Table 4. Students' Impressions of Second-Step Self-Evaluation (Extracts)**

No	Text	<1>Notable words and phrases in the text	<2>Rephrasing words in the text	<3> Concepts that are not in the text that would explain <2> of the left	<4>Themes/constitutive concepts
1	<ul style="list-style-type: none"> <li>• Receiving evaluations from other people's perspectives enabled us to grasp more specifically the areas for improvement, and I was able to conduct a very meaningful self-evaluation again.</li> <li>• I was able to review the anchor training of all four groups, including our own, so I was able to gain a better understanding of this practice.</li> </ul>	evaluation from others' points of view, able to understand areas for improvement, meaningful self-evaluation	objective evaluation and advice, understanding of what went wrong, and effective, useful reflections	incorporation of others' perspectives (causes), evaluation that is not self-indulgent (results), the effects of reflection (impacts)	effects of new insights from other people's and professionals' perspectives, fair evaluation to promote growth, effect of other teams' reflection, deeper understanding
2	<ul style="list-style-type: none"> <li>• I was often made aware of the good and bad points by others pointing them out.</li> <li>• In the first step, I was a little depressed because I did not think it would work at all, but when I reviewed the evaluation again in the second step, I was encouraged by what was good and motivated to pay attention the next time to what was not good enough.</li> </ul>	pointing out to others, good points bad points, being reminded, not doing well at all, evaluating again, reviewing, what was good is encouraging, what was missing	advice from others, what I did well, what I did not do well, what I noticed.	effects of noticing (result), effects from the second evaluation (result), difference between first and second impressions	perspectives of others in debriefing, etc., the effect of noticing much and details by pointing out of others, and differences between the first step, second-step evaluations
3	<ul style="list-style-type: none"> <li>• In the first step of self-evaluation, I could not organize my actions and did not find many accurate reflections, but in the second step of self-evaluation I was able to do so.</li> </ul>	first-step self-evaluation, organizing my actions, accurate reflection, second-step self-evaluation	problems with the first-step evaluation, excellent points of the second-step evaluation, identification of appropriate areas for improvement	availability of behavior organization (comparison), improvement in the ship handling (result)	reflecting on the big picture of my actions, improving training content, creating many specific areas for improvement

4	<ul style="list-style-type: none"> <li>• The briefing, self-evaluation, and debriefing allowed me to understand the content of the training and to think more deeply about it.</li> <li>• It was effective because it allowed me to know what I need to improve for the next time. I would like to incorporate the second-step self-evaluation not only in this training but also in others in the future.</li> </ul>	<p>to understand the content of the training, to be able to think deeply about the contents of the training, improvements for next time, self-evaluation of the second step, to incorporate the second-step self-evaluation</p>	<p>effectiveness through a series of training content, improvement of understanding through the second-step evaluation, improvement of understanding through reflection</p>	<p>depth of training (result), development of two-step evaluation (dimensions, development), effective improvement (result)</p>	<p>a greater depth of understanding, deepening of thinking, identification of many areas for improvement, application of the two-step evaluation to other</p>
Storyline (what can be said at this point)	<ul style="list-style-type: none"> <li>• The second step of the evaluation is conducted after receiving many new insights from the opinions and suggestions of the others, and advice and an indication of the professional perspectives, and thus, has the effect of deepening the understanding of anchoring training as well as identifying many specific and accurate areas for improvement and positive aspects.</li> <li>• The second-step evaluation has the effect of deepening the understanding of the anchoring training and identifying many specific areas for improvement and good points, since time has passed and the evaluation can be reviewed calmly, broadly, and in detail.</li> <li>• Students realized that the second step of the evaluation could be implemented in situations other than anchoring training because of its perceived effectiveness.</li> </ul>				
Theoretical description	<ul style="list-style-type: none"> <li>• Many new insights can be gained through the opinions and suggestions of other students and instructors.</li> <li>• Many specific and accurate improvements and good points can be found.</li> <li>• It has the effect of increasing understanding of anchoring training, ship handling, BRM, etc.</li> <li>• With the passage of time, the students can reflect calmly on the bigger picture and details, for example, by listening to the opinions of others.</li> <li>• To share the results and reflections of other teams so that new insights and areas for improvement can be identified.</li> <li>• Since the effectiveness of the second-step evaluation can be felt, it can be considered for implementation in other situations.</li> </ul>				

The purpose of collecting student impressions was threefold.

- (1) How good or bad the second step of the evaluation was and how the students felt about it
- (2) Confirmation of the appropriateness of the evaluation of the second-step evaluation
- (3) Educational effects of the second-step evaluation

The results of the SCAT analysis of the students' impressions, in which the second-step evaluation took place after the "debriefing" and "presentation" had been conducted, yielded the following six different theoretical descriptions.

- a) Many new insights can be gained through the opinions and suggestions of other students and instructors.
- b) Many specific and accurate improvements and good points can be discovered.
- c) It has the effect of increasing understanding of anchoring training, ship handling, BRM, etc.
- d) With time, the students can reflect calmly on the bigger picture and details, for example, by listening to the opinions of others.
- e) By sharing the results and reflections of other teams, new insights and areas for improvement can be identified.
- f) Since the effectiveness of the second-step evaluation can be felt, it can be considered for implementation in other situations.

There were no negative comments about the second-step evaluation, all of them were positive, and the content of the second-step evaluation was a good effect of the evaluation.

### CONCLUSION

Self- and mutual evaluations were conducted, and students' impressions of anchoring training were surveyed as effective maritime education and training. The results showed that a high percentage of both self- and mutual evaluations were rated highly as useful. It was also found that when they evaluated others in a mutual evaluation, students gave better evaluations compared to their self-evaluations. Furthermore, it was found that the self-evaluation scores of students with higher instructor-rated scores were closer to the instructor's scores and slightly lower. On the other hand, students with lower scores evaluated by the instructor differed significantly from the instructor's scores, in some cases higher than the instructor's scores and in some cases lower.

A first self-evaluation was conducted immediately after the training had been conducted, and a second-step self-evaluation was conducted after the debriefing and presentation. The results of the analysis of students' impressions of the second-step self-evaluation using SCAT, a method of qualitative analysis, allowed us to describe the effects of the evaluation, such as "gaining new insights," "finding new areas for improvement and good points," "deepening understanding of training content," and "calm and big-picture reflections." In addition, developmental feedback was obtained, such as "I would like to use the multi-step evaluation in other situations." In the future, we would like to examine methods to quantitatively seek the effects of the second step self-evaluation and verify its effectiveness on education and training. In addition, we would like to introduce an evaluation method in which students can think for themselves to obtain further educational effects, and verify the effects of this method.

### References

Taner Albayrak., et al., (2012): Encouraging Research in Maritime Education & Training, Journal of Maritime Transport and Engineering, Volume 1 (1), 4-9.

Japan Coast Guard (2021): Current Status and Countermeasures for Marine Accidents in 2020.

Kashima, H., Kunieda, Y. and Takemoto, T. (2001): About the Training Effect of Ship Handling Training, Journal of the National Institute for Sea Training, 2001, 1, Vol. 1, 17-38.

Kunieda, Y., Murai, K. and Kashima, H. (2018): Kunieda Yoshiaki et al., Study on Education of Seamanship in the Anchoring Training, the 16th World Congress of the International Association of Institutes of Navigation Proceedings.

Kunieda, Y., Kashima, H., Murai, K. and Niwa, S. (2020): The Influential Features of an Effective Model of Anchoring Training in Maritime Education, *Psychology Research*, March 2020, Vol. 10(3), 106-117.

Mochizuki, M. (2013): On Peer Learning and Learning Process in JSL Learners, Kansai University, *Bulletin of Faculty of Foreign Language Studies*, 8, 87-97.

Mizogami, S. (2014): *Conversion of Active Learning and a Lesson Study Paradigm*, Toshindo.

Tanaka Hiroyuki (2017): *Evaluation of Learning Achievements for Active Learning*, Gakuyo Shobo, 74-83.

Otani T. (2011): SCAT: Steps for Coding and Theorization - Qualitative Data Analysis Methods that are Easy to Initiate with Explicit Procedures and Applicable to Small Data Sets-, *Kansei Engineering International Journal*, Vol.10(3), 155-160.