Jonathan Soper[1]
Faculty of Engineering and Applied Science,
Memorial University,
240 Prince Phillip Drive,
St. John's, NL, A1B 3X5, Canada
e-mail: jksoper@mun.ca

Erik Veitch
Department of Design,
Norwegian University of Science and Technology,
Kolbjørn Hejes Vei 2b,
Trondheim 7491, Norway
e-mail: erik.a.veitch@ntnu.no

Rebecca Thistle
Faculty of Engineering and Applied Science,
Memorial University,
240 Prince Phillip Drive,
St. John's, NL, A1B 3X5, Canada
e-mail: rebecca.m.thistle@mun.ca

Jennifer Smith
Fisheries and Marine Institute,
155 Ridge Road,
St. John's, NL A1C 5R3, Canada
e-mail: jennifer.smith@mi.mun.ca

Brian Veitch
Faculty of Engineering and Applied Science,
Memorial University,
240 Prince Phillip Drive,
St. John's, NL, A1B 3X5, Canada
e-mail: bveitch@mun.ca

An Analysis of Factors Influencing Ice Management Performance in an Experimental Marine Simulator and Their Application to Decision Support System Design

Ice management is essential for maintaining the safety of offshore operations in Arctic regions. We present the combined results of three experiments conducted in a full-mission bridge simulator specially designed for ice management. From a quantitative analysis of the results, we infer the effect of three variables on performance: (1) experience, (2) training, and (3) Decision Support System (DSS). The results confirm that experience and training improve performance for untrained and inexperienced simulator participants. The DSS also improves performance, but with a smaller effect. Qualitative observations using vessel position heat-map diagrams and exit interviews suggested that novice participants using the DSS adopted expert strategies but carried out their tasks more slowly and with less precision. This has important consequences for the design of a future DSS used in training simulators or onboard ships. Potential improvements to the DSS design might include real-time feedback to the user, a redesign of the human–machine interface (HMI), and increasing user input and customization with a human factors focus.
[DOI: 10.1115/1.4063617]

Keywords: marine, ice, navigation, decision support system, automation, simulator, training, human factors, ships in ice

1 Introduction

1.1 Purpose and Background. As the maritime domain incorporates higher levels of automation, it is important to appropriately adapt maritime training and to understand how human expertise can be retained. Complex maritime operations like ice management still depend to a large extent on human expertise, requiring automation efforts aimed at supporting such operations to shift their focus toward human factors.

Ice management refers to marine practices taken to ensure that safe marine operations can continue with minimal disruption in sea ice conditions. This can take many forms, from icebreaking and clearing to iceberg towing. The operations are highly specialized and variable, depending on the unique environments of a given region. As such, it is an activity that is very difficult to automate due to the limited quantity of data and its apparent reliance on human expertise.

In this research, the results of three experiments are analyzed. Experiments I, II, and III quantify and describe the effects of experience, training, and a Decision Support System (DSS) on ice management performance, respectively. An immersive marine simulator was used to test these effects using five cohorts of participants who were instructed to complete an ice management task. Two questions were investigated: (1) out of experience, training, and DSS, which factors are most effective at influencing seafarer performance? (2) How does performance when using a DSS compare to the factors of training or experience?

This paper will first outline the experimental equipment and process. Next, the qualitative and quantitative results will be shown. Finally, the implications of these results and next steps will be discussed.

1.2 Experiments. All three experiments used the same simulator, habituation scenarios, and design of experiment (DOE) to ensure consistency so that analysis between cohorts was possible. A full-factorial experiment with two factors and two levels was performed, and a split-plot approach was used to compare results across participant cohorts. Additionally, careful checks of analytical assumptions were made to ensure the results from the experimental design were valid.

---
[1]Corresponding author.

Experiment I examined the effects of the experience factor on seafarer performance. Seafarers with at least 10 years of experience were tested against novice participants from a local maritime navigation program. These corresponded to high and low levels of experience, as is required for a full-factorial analysis [1]. The novice group from Experiment I will be referred to herein as the *No DSS* group, since they were untrained, and were not provided with a DSS. The seafarers' group from Experiment I will be referred to here as the *Seafarers*.

Experiment II examined the effects of training on ice management performance. This study employed the same methods as Experiment I and recruited from the same pool of novices who did not participate in Experiment I. These participants were randomly assigned either one or two sessions of training, which corresponded to the low and high levels respectively. They were also assigned to a mild or severe ice concentration level, to allow for comparisons between experimental campaigns [2]. The novice participants from Experiment II will be referred to here as *Training I* and *Training II*.

Experiment III used a DSS to aid the performance of novices. This experiment recruited another cohort of novices from post-secondary institutions. The novice participants were provided with a Case-Based Reasoning (CBR) DSS that gave tactical guidance based on expert advice. The participants could request assistance, and the DSS would show them the most effective strategy for the ice management task, based on their current approach. Experiment III again used severe and mild levels of ice concentration to allow for comparisons between all experiments. The *DSS* group is referred to here as *DSS*. A summary of the participant cohorts can be seen in Table 1.

1.3 Summary of Results. This paper examines the results of all three experiments and qualitatively demonstrates the effects of each factor. It also discusses qualitatively the various tactics chosen by the different cohorts. A summary of the experiments is listed here:

- In Experiment I, it was hypothesized that more experience would lead to better performance and reduced variability [3].
- Experiment II hypothesized that increasing amounts of training would improve performance and reduce the variability in performance among the novice participants. Experiment II also hypothesized that this relationship could be used to provide a method of estimating the quantity of training required to achieve a required performance level [2].
- Experiment III hypothesized that the use of a DSS by novice participants would increase performance and decrease variability of performance when compared to novice participants who did not have access to the DSS. Experiment III also hypothesized that the tactics chosen by the *DSS* group would be similar to the *Seafarer* group from Experiment I, and the trained novices from Experiment II.

Experiments I and II showed that experience and training did indeed increase performance significantly in most metrics. In Experiment III, *DSS* group did not significantly improve their performance over the *No DSS* group at 95% confidence; however, they

did employ similar tactics to the *Seafarers* group, and the *Training I* and *II* groups. Some performance improvements were observed; however, they were often not statistically significant. These findings could be relevant to developers of navigational decision support systems and curriculum development at navigation colleges to help them improve their delivery of ice navigation knowledge. A summary of these results is presented in Table 2.

2 Materials and Methods

2.1 Simulator. Marine simulators are an effective, affordable, and low-risk way of training and evaluating mariners. This allows for studies of human performance in simulated high-risk situations at sea without endangering crew, passengers, or the environment. Simulation has been used to explore the effects of new technology, legislation, and organizational changes and has been found to be a valid method [4]. Prior to this experimental campaign, simulators have also been used to evaluate training on the ability of novices to operate lifeboats in ice [5].

All three experiments used for this analysis were performed using an ice management simulator operated by Memorial University of Newfoundland and Labrador. The simulator was built for marine safety research, specifically for scenarios involving sea ice. The simulator is immersive, featuring a 360-deg screen that displays images of 11 projectors. The diameter of the screen is approximately 8 m. In the center, a simplified bridge console provides basic controls for a simulated anchor handling tug supply vessel (AHTS). The simulated vessel is modeled after a vessel of 75 m in length typical of Newfoundland and Labrador's offshore oil and gas industry on the Grand Banks [6]. The simulated vessel has twin 5369-kW diesel engines coupled to fixed pitch propellers and is provided with 895-kW tunnel thrusters at the bow and stern. A schematic of the simulator can be seen in Fig. 1(a).

The bridge console consists of a small platform with identical fore and aft-facing controls. This allows for the participants to choose whether they want to handle the vessel in a forward or aft-facing position, a practice that is common for operations with AHTS vessels. These controls are simplified and consist of two throttles for port and starboard main engine control, a bow and stern thruster control, and a ship's wheel for rudder control. A display is used to present information to the participant regarding vessel speed, engine speed, heading, and rudder angle. No radar or chart plotter is provided; however, the participant can use a very high frequency

Table 2 Summary of hypotheses and results for Experiments I, II, and III

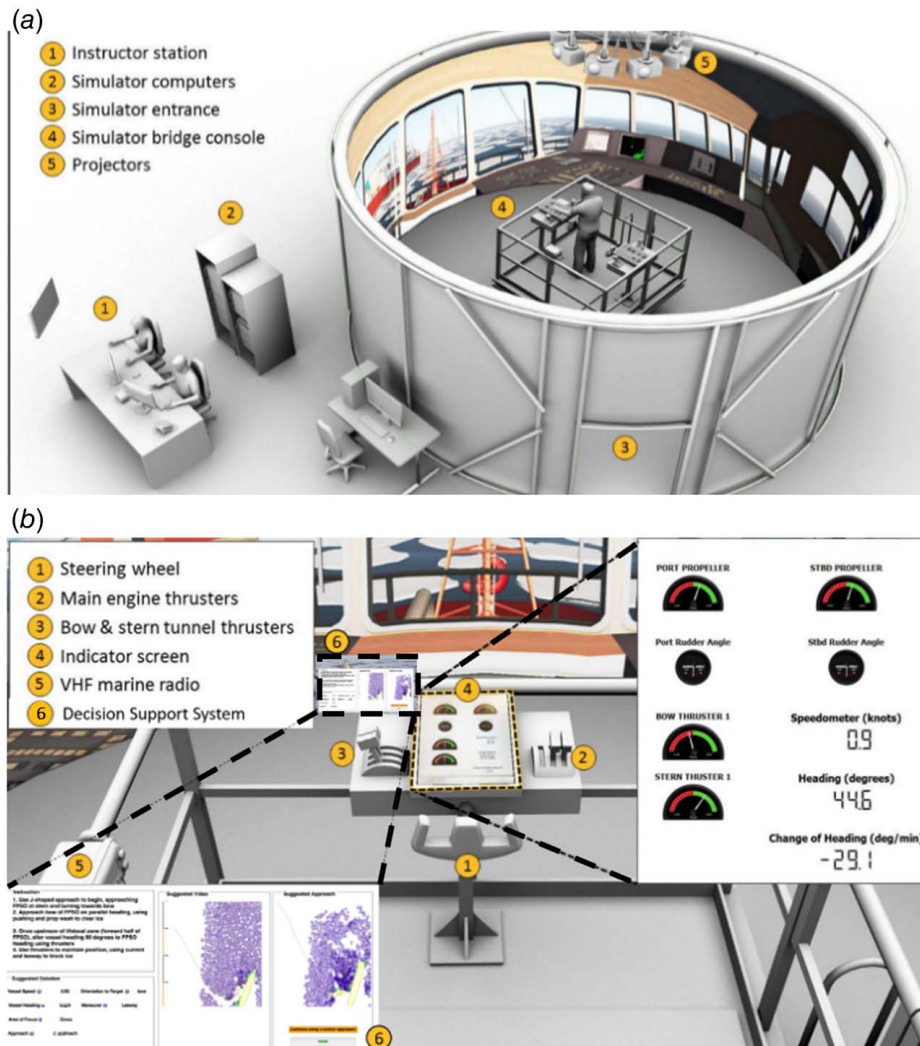| Study	Hypothesis	Result
Experiment I	Experience significantly improves ice management performance	Experienced participants performed significantly better than inexperienced participants
Experiment II	Increasing levels of training will significantly improve ice management performance	Performance improved with increasing amounts of training in severe ice conditions. In mild ice conditions, training brought inexperienced participants to the level of experts, but a second level of training did not further improve performance
Experiment III	The DSS will significantly improve performance	Improvements were not statistically significant, but observational improvements were found. Qualitative evidence suggests expert strategies were followed

Table 1 Description of participant cohorts

Cohort	Description
No DSS ($n = 18$)	Novice participants received no training or DSS assistance
DSS ($n = 18$)	Novice participants provided with a DSS
Training I ($n = 17$)	Novice participants provided with a single training session
Training II ($n = 18$)	Novice participants provided with two training sessions over different days
Seafarers ($n = 18$)	Experienced seafarers with a minimum of 10-year experience

Fig. 1 Simulator and controls: (*a*) complete simulator showing participant at controls and experimenters at instructor station (above), and (*b*) close view of controls from the position of participant (below) [3]. A closer view of the DSS (item 6) can be seen in Fig. 2.

(VHF) radio to request distances and bearings off objects from the experimenters. The experimenters sit outside the simulator but have access to this information on their display which they can relay to the participant as if they are a crew member on the bow or bridge wing. The DSS is positioned to the left of the participant when in the forward-facing position and is presented on a laptop. It can be activated by using the laptop's trackpad to press the on-screen "assist" button.

The scenario for the experiment consists of an ice management operation for a floating production, storage, and offloading (FPSO) vessel in an ice field with a 0.5–0.6 kt current. It is modeled after a plausible ice management assignment for a standby vessel working in the Newfoundland offshore region [6]. The ice is considered to be medium first-year ice with a thickness of 0.4 m and a concentration of either four-tenth or seven-tenth, depending on the randomized assignment of the participant. Concentration is defined as the ratio of ice cover to open water and is expressed in units of tenths (X/tenth). The task given to the participant is to use the supply vessel they are piloting, and clear ice from the lifeboat launch zone of the FPSO. Various metrics are used to quantify their performance, but ultimately, participants are instructed to clear as much ice as possible from the FPSO's lifeboat launch zone.

2.2 Decision Support System. The DSS used in this experiment is an example of a low-level form of automation and can be seen in Fig. 2 [7]. The DSS tested in Experiment III uses a CBR algorithm developed in a previous study using expert knowledge [8]. Specifically in the work of Yazdanpanah and Smith et al. [9], the DSS was created using a database of previous results from the first two experiments by Veitch [3] and Thistle [2], which is known as a case-base in a CBR model [8]. In the development of the DSS during a previous study [8,9], experienced mariners were interviewed and asked to critique the performance of participants in these cases and to identify the most critical factors for effective ice management. This work formed the basis of the DSS used in Experiment III. The factors identified included vessel particulars (heading, speed, and specific ice class), ice conditions (ice concentration, floe size, ice type, and thickness), and task objective (area of ice to be cleared) [6]. These factors were ranked by importance by the experienced seafarers for classification in the case-base. Following this, the experienced seafarers were also asked to participate in a simulator session so that their recommended approaches could be tested in the same setting.

During Experiment III, each new participant in the simulator had their approach matched by the DSS to previous cases in the case-base using factors such as position and heading in relation to the FPSO. This allowed for the closest match for an above-average case to be shown to the user as a case to emulate. In general, having more cases in the case-base improves the ability of the DSS to match specific cases and give good advice [8]. In practice, the case-base contains approximately 40 cases. Cases in which the performance was considered below average were not included to
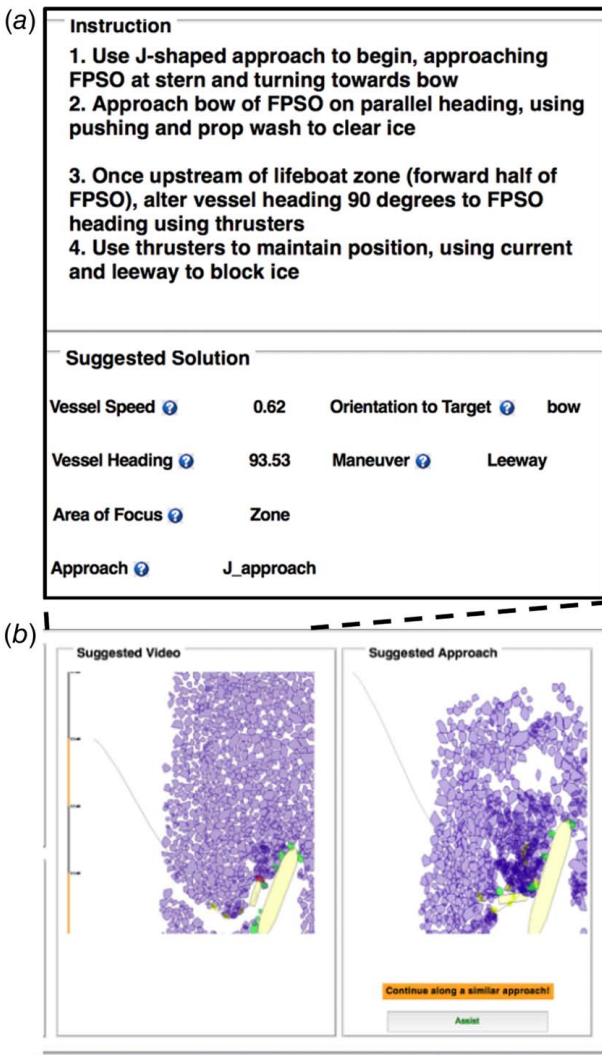
Fig. 2 The DSS GUI: (a) instruction portion is magnified for clarity (above), while (b) features the pictorial and video component of the GUI (below)

most important component based on observations of user preferences from tests made during development.

Also the display, shown in Fig. 2(a), provides text and quantitative information to the operator. Recommendations are provided with step-by-step instructions. This information is presented as point-form instructions that were adapted using interviews with experienced seafarers. The instructions are customized for the specific approach being recommended. Below this, the "Suggested Solution" section provides specific operating parameters and strategy types. This includes a recommended heading, speed, orientation to the FPSO, and the name of the ice management approach suggested. The participant can hover their mouse over the suggested approach for a detailed description, in case they are not familiar with the term. This will also match the instructions given earlier.

The DSS will then match their current position, speed, and other parameters to the case-base and may suggest a new approach if these parameters differ from the last time assistance was requested. If there is no significant change, the advice will stay the same. For example, if the user requests assistance without changing their position, the advice will not be different.

Within the first 2 min of a scenario, there is not sufficient data for the algorithm to work accurately. However, given the importance of the first 2 min in deciding the initial strategy, the DSS was modified. For instance, in the first 2 min if assistance was requested, one of three ideal cases representing three different initial approaches would be presented depending on the heading of the participant's vessel. This allowed the participant to choose an approach and be shown the most appropriate strategy for said approach immediately. After 2 min, there was sufficient data to resume the use of the CBR algorithm.

2.3 Design of Experiment. The experiment design is a 2-k full factorial [1]. In Experiment I, the two factors studied were experience and ice concentration [3]. Experiment II examined the effects of two levels of training on novice participants, and ice concentration [2]. Experiment III studied the effects of the DSS guidance and ice concentration. In Experiment III, studying the DSS as a factor will allow for comparisons against the factors of the first two experiments.

To maintain consistency between experiments to allow for a combined analysis, steps were taken to minimize differences between the experiments. This included the use of scripts so that briefings to participants were consistent, the use of the same habituation scenarios, and the same experimental equipment including the simulator. Additionally, 18 participants were recruited for Experiment III to match the group size in Experiment I and Experiment II. This maintains a similar design power, assuming a similar effect using Cohen's d method [10,11].

2.4 Experimental Procedure. Experiment III was reviewed and approved by Memorial University's Interdisciplinary Committee on Ethics in Human Research (ICEHR). The experimental procedure was kept as close as possible to the procedures in Experiments I and II so that experimental results could be compared between all groups. The simulator, scenario, habituation script, and participant qualification criteria used in all three experiments were the same. A detailed description of the experimental procedures for Experiments I and II can be found in the literature by Thistle et al. [2] and Veitch et al. [3].

In Experiment III, the following steps were taken prior to beginning the experiment in the simulator:

- Prospective participants were recruited by open invitation.
- Participants who contacted the researchers were scheduled into a timeslot that was compatible with the schedules of both parties.
- Participants were then randomly assigned to an ice concentration group on the day of their simulator session, but were not informed which ice concentration group they would be placed in before arriving.

avoid giving bad advice to participants. A DSS can be improved by adding new cases to the case-base as more users participate. However, this function was not used for this controlled experiment to avoid altering the properties of the DSS between participants.

The human–machine interface (HMI) was altered from its original form developed by Yazdanpanah using knowledge gathered by Smith et al. [8,9]. Six individuals with various backgrounds were asked to complete an ice management task in the simulator using the DSS. They were then interviewed to see which components of the DSS were most useful, and which could be improved. Although a formal study on the effects and best practices of HMI design was not considered, this allowed for several incremental improvements to be made to the presentation of the assistance over the original design. It also allowed the CBR algorithm to be improved at the beginning of the scenario, when data from the participant was still lacking.

The HMI of the DSS was improved in several ways. A video replay of the bird's eye view of the suggested route was added. As shown in Fig. 2(b), the video provides a sped-up capture of the entire 30-min scenario in a six-second replay. The replay loops so that the participant can get a clear strategic view of the scenario. The video allows the participant to observe the dynamics of the ice flowing south in the current and the interactions of the supply vessel. The test instructions can also be referred to if a user desires further details explaining the steps required to emulate the video. The video was considered the

- Participants were provided with an informed consent form and were asked to complete a questionnaire to determine their level of experience. This captured information about their time spent in formal maritime education, time at sea, and time spent working on ice. To ensure safety, the participants were asked to complete a simulator sickness questionnaire to establish a baseline of their current physical state before entering the simulator. This questionnaire was based on the work by Kennedy [12]. Some common symptoms they were asked to be aware of are nausea, headaches, or dizziness. This was repeated periodically to ensure participant safety, and participants were informed that they could stop the experiment at any time should they no longer feel comfortable. No adverse effects were reported by any participants.

Upon completing the questionnaires, participants were shown the simulator controls and were given three different vessel habituation scenarios to complete. This was intended to familiarize them with the controls in the simulator and the virtual environment, but was not intended to provide them with ice management training. These habituation scenarios were identical to those used in Experiments I and II to ensure all participants across all cohorts had the same level of familiarity with the facility prior to beginning the experiment.

Following the habituation, participants were given the task required by the experimental scenario and, in Experiment III, were shown how to use the DSS. The DSS habituation gave the participants the opportunity to familiarize themselves with the assist function, gain knowledge about the information presented to them, and develop an understanding of the strengths and limitations of the technology. The latter point is important to ensure that the participants do not misunderstand the information being presented to them, which could result in incorrect use of the technology [13].

The experimental "Emergency Ice Management" scenario was used in Experiments I, II, and III. The purpose of the scenario was to use the AHTS standby vessel under the participant's control to perform an ice management operation where ice was to be cleared from two zones on the FPSO's port side. Figure 3 provides an example of the seven-tenth ice concentration of the emergency ice management scenario. The two zones to be cleared include the larger ice management zone shown as a semi-
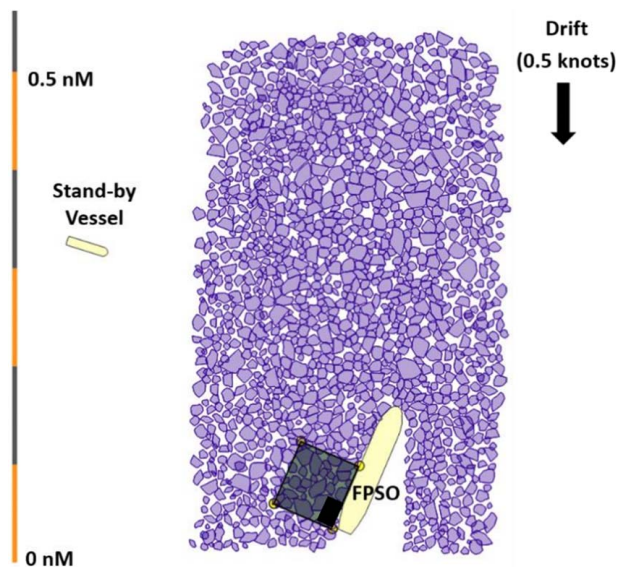
transparent square on the port side of the FPSO, and the lifeboat launch zone shown as a smaller black rectangle inside. The initial position of the participant's AHTS standby vessel is shown in the top left. There is a 0.5-kt current pushing the ice south.

The larger zone is a 120×120 m square area. It leads from the stern of the FPSO to approximately amidships. Figure 4(a) provides a closer view of this zone. The lifeboat launch zone is 16×8.2 m and is directly underneath the FPSO's lifeboat, as shown in Fig. 4(b). The instruction given to participants was to clear ice for the lifeboat launch over the course of the 30-min scenario. They were shown the lifeboat in the simulator which is visible on the projected image of the FPSO, but were not given specific instructions as to the dimensions of the location to clear. The scenario is derived from realistic ice management activities used in regions of the offshore oil and gas industry where the presence of sea ice is likely. The Newfoundland Grand Banks are an example of such a region [6].

The representation of this sea ice in the simulator was done through the generation of ice floes by randomly sampling a lognormal distribution of ice. The floes were given a uniform thickness of 40 cm [3]. The FPSO was mostly static throughout the scenario and did not significantly yaw, pitch, or roll. The ice drifted at 0.5–0.6 kt with the current from the port bow of the FPSO and flowed past the stern.

Participants in Experiment III were informed that they had the option to use the DSS as many or as few times as they wished. In the DSS habituation, participants were informed that the software evaluated their current position, heading, and orientation to the FPSO and used this to provide them with tactical recommendations for a strategy or approach to follow.

Upon completion of the scenario, a final simulator sickness questionnaire was provided to the participants. They were given an exit interview for qualitative analysis of their impressions of the DSS efficacy. During the interview the participants were asked their reasoning for their chosen strategy, their perceived score on a scale of 1–5 (where one represented poor performance and five represented successful performance), and the usefulness of the DSS on a scale of 1–5 (where one is not useful, and five is very useful). They were

Fig. 3 Schematic of Emergency Ice Management scenario used in the experiment, showing the zones to be cleared. The figure also shows the participant's standby vessel, the FPSO, the direction and speed of the current, and the ice floes present. The larger square represents the ice management zone. The dark rectangle is the lifeboat launch zone.
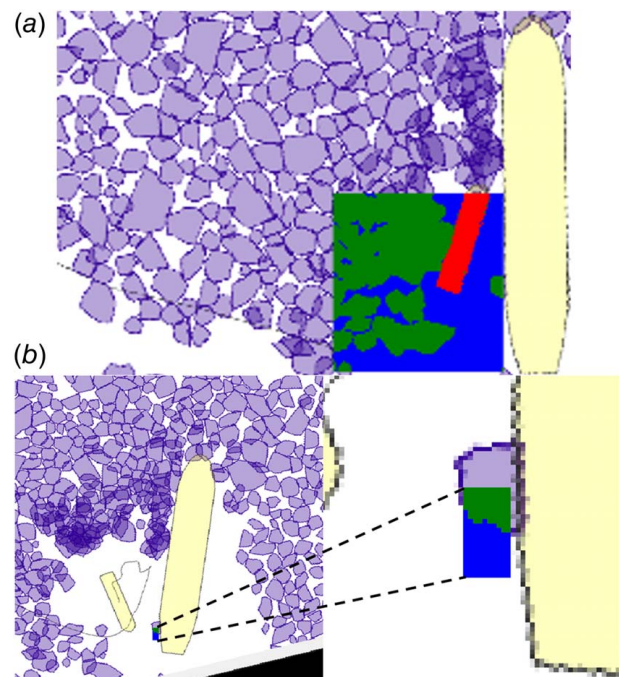

Fig. 4 Image processing of ice concentration by pixel count. (a), above, shows the larger ice management zone with the AHTS inside it, while (b), below, shows the lifeboat launch zone.

also asked which components of the DSS they found helpful, which they did not use, and which they thought could be improved.

2.5 Analysis. Experiments I, II, and III used the same methods to analyze the results quantitatively. To analyze the results, each participant's simulated case was replayed in real time, and screen captures of a top-down view from the instructor station were taken every second. After the scenario had been completed, 1800 images were captured. An example of the top-down view of captured images can be seen in Fig. 3. Each case also generated a text file comprised of speed and position data for the participant's vessel. Image processing scripts were then used to calculate the specific ice concentration in the zones at regular intervals. An image processing output frame for one of the participants can be seen in Fig. 4. Pixels were classified based on whether they are the vessel, ice, or open water. This was used to calculate the concentration.

All participants in each concentration sub-group were given the same initial condition to maintain consistency and allow for valid comparisons. A baseline was generated by measuring the change in ice concentration over the zones when no ice clearing was performed and was compared against each case to generate data for the following performance metrics: (1) mean change in ice concentration; (2) cumulative clearing time; and (3) clearing-to-distance ratio. These metrics will be described in detail below:

- The mean change in ice concentration metric was calculated by taking the difference in ice concentration in the larger zone shown in Fig. 4(a) from the baseline at each 30-s interval of the 30-minute scenario. Each interval was then averaged. Better performance was indicated by a larger value, which meant that more ice was cleared.
- The cumulative clearing time metric examined the smaller zone below the lifeboat, as shown in Fig. 4(b). When the zone was completely clear of ice with a concentration of zero-tenth during one of the 30-s time-steps, the lifeboat was considered able to launch. The total time in seconds that the lifeboat launch zone was completely clear of ice was summed for the duration of the 30-min case. A higher value was considered a better score since this meant that the lifeboat had more time to launch in open water during the case. An example of the output for concentration in the lifeboat launch zone over the course of a single case is shown in Fig. 5.
- The clearing-to-distance ratio measured the efficiency of a participant's case. This was done by dividing the quantity of ice cleared by the distance traveled by the vessel. A higher value indicated more ice cleared per distance traveled and was considered to be a more efficient performance.

3 Results

In this section, the assumptions required for analysis, the effect size, and the results of the three experiments will be presented.

3.1 Statistical Assumptions. To analyze the results as a full-factorial split plot, several assumptions were validated. This was completed through an analysis of the residuals for each metric [1]. The assumptions are listed here as follows:

- *Normal Distribution*: The first assumption is that the residuals are normally distributed. This was done by plotting residuals against probability. If the residuals form a straight line, the assumption of normality is assumed to be correct [1].
- *Randomness of Run Order*: Next, it was checked that no time-related variables affected the response. This was especially important to confirm in Experiment III since all three experiments were performed at different times. To validate this assumption, it was checked that the residuals were randomly scattered and not influenced by run order.
- *Equal Variance*: Another assumption that was checked is that the data have equal variance, a term known as
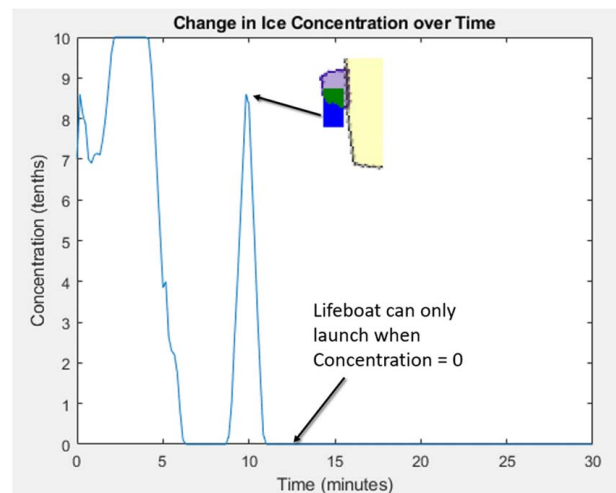

Fig. 5 Example measurement of ice concentration in the lifeboat launch zone for a single case. A capture of the ice floe present in the lifeboat launch zone can be seen at the 10-min mark to demonstrate how this affects the concentration metric. The lifeboat can launch when the concentration in this zone is zero.

heteroscedasticity [1]. To validate this assumption, it was checked that the residuals were evenly distributed between the upper and lower bounds of a plot.
- *Transformations*: Finally, it was checked that all recommended transformations were applied, if needed, by using a Box-Cox plot. A detailed analysis of these residuals can be seen for Experiment I and Experiment II in the studies by Veitch et al. [3] and Thistle and Veitch [2]. A detailed analysis of Experiment III can be seen in the study by Soper [14].

Experiment I demonstrated that sufficient design power and statistical significance had been attained in the experiment comparing the *No DSS* group to *Seafarers* [3]. Further, Experiment II demonstrated that sufficient design power was achieved to obtain significant results when comparing the Training groups to the *No DSS* group [2]. The trained groups were significantly better than the *No DSS* group who were not trained. A more detailed statistical analysis can be seen for all experiments in Veitch et al. [3], Thistle and Veitch [2], and Soper [14].

3.2 Effect Size. Effect size is a useful way of reporting the practical significance of results from empirical studies by demonstrating whether an effect exists from an experimental campaign, and if so, how big that effect is [15]. It also provides a standardized way to examine the magnitude of an effect across studies [15]. This can give a more nuanced view of the relationship between the different groups. Effect size can also be used to determine the expected design power for future experiments which is an important step in determining sample size.

In this analysis, the Cohen's *d* method of calculating effect size is used to determine the effect of each factor against each other. The effect size between factors was calculated the following way, with an example shown for the *DSS* group:
Equation (1): Calculation for Cohen's *d* [10]

$$d = \frac{\mu_{\text{DSS}} - \mu_{\text{Control}}}{\sqrt{\frac{\sigma_{\text{DSS}}^2 + \sigma_{\text{Control}}^2}{2}}}$$

where μ represents the mean value of each group, and σ represents the standard deviation. Cohen's *d* assumes that standard deviations are equal between groups. Since a perfect equal standard deviation is not likely to occur in the real world, a pooled standard deviation

can be used, which combines the standard deviation from both groups. This is done by adding the square of each standard deviation, dividing them by 2, and taking the square root [16]. It is a unitless value.

Cohen's d effect sizes can be classified as shown in Table 3. It should be noted however that these are guidelines.

The Cohen's d effect size was calculated for each metric and can be seen in Table 4. A comparison between each factor studied here is shown from all three experiments. Referring back to Table 3, the size of the effects can be seen. The largest effect is present in comparing the *No DSS* group to *Training I*, *Training II*, and *Seafarers* groups. This was expected as the *No DSS* group is a control group with no experience, training, or DSS assistance. The *DSS* group shows some effect over the *No DSS* group; however, this effect is limited when compared to the effects of experience and training. The effect size between the *DSS* group with the *Training I*, *Training II*, and *Seafarers* groups is smaller than that of the *No DSS* group to *Training I*, *Training II*, and *Seafarers*, further indicating that the DSS has some effect, but that it is not as effective as the effects of experience and training.

When examining the values in Table 4 and referring back to Table 3, it can be seen that most of the group pairs have a medium to large effect.

3.3 Quantitative Results.
A summary of the results can be seen in Tables 5 and 6. These tables show the mean value in each metric and the standard deviation. Each column contains one of the three metrics, with the different groups shown in each row. The mean value of all participants in a group is shown to indicate the average result and the standard deviation is shown to indicate the spread in the data for the specific group's metric.

In Table 5 for four-tenth ice concentration, it can be seen that for the Mean Concentration Change metric, the *No DSS* group has the lowest value, indicating the worst result. The standard deviation for this metric is considered medium relative to the other groups. This indicates that there was some variance in each participant. The *DSS* group had an improvement in performance for this metric in the four-tenth concentration, but this advantage is not present in the seven-tenth concentration, as shown in Table 6. In both concentration levels, the *Training I*, *Training II*, and *Seafarers* groups

Table 3 Description of Cohen's *d* magnitude and meanings [13]

Cohen's d	Effect description
$d = 0.2$	Small
$d = 0.5$	Medium
$d = 0.8$	Large

Table 4 Cohen's *d* effect sizes for all factors

Group comparison	Mean concentration change	Cumulative time clear	Mean clearing to distance
DSS to No DSS	0.7	0.25	0.84
DSS to seafarer	0.89	0.95	0.94
No DSS to T1	1.46	0.71	1.9
No DSS to T2	2.23	0.89	2.8
Seafarers to T1	0.69	0.44	0.38
Seafarers to T2	0.68	0.53	1.01
DSS to T1	0.97	0.86	0.89
DSS to T2	1.44	0.78	1.38

Note: Training I and Training II are shortened to T1 and T2, respectively.

Table 5 Summary of performance metrics for four-tenth concentration, showing mean ± standard deviation

	Mean con change (tenths)	Cumulative time clear (S)	Mean clearing-to-distance ratio
No DSS	1.1 ± 0.65	586 ± 406	0.25 ± 0.16
DSS	1.75 ± 0.84	530 ± 276	0.63 ± 0.45
Training I	2.23 ± 1.02	735 ± 224	0.61 ± 0.34
Training II	2.10 ± 0.48	547 ± 219	0.70 ± 0.29
Seafarers	1.76 ± 0.45	687 ± 210	0.52 ± 0.17

Table 6 Summary of performance metrics for seven-tenth concentration, showing mean ± standard deviation

	Mean con change (tenths)	Cumulative time clear (S)	Mean clearing-to-distance ratio
No DSS	1.87 ± 0.83	408 ± 304	0.48 ± 0.19
DSS	1.76 ± 0.85	461 ± 267	0.49 ± 0.30
Training I	2.48 ± 0.83	568 ± 273	0.95 ± 0.45
Training II	2.95 ± 0.71	743 ± 282	1.13 ± 0.42
Seafarers	2.78 ± 0.80	710 ± 418	0.86 ± 0.43

perform best. In the four-tenth concentration group, the standard deviation varies more between groups than in the seven-tenth group. The reason for this is unclear.

Looking at the next column in Table 5, it can be seen that the DSS offers little improvement over the *No DSS* group. In this metric, a higher value is considered better. The *Training I* group performs best, followed by the *Seafarers*. The *Training II* group is in between these. The primary change seen by the *DSS*, *Seafarers*, and *Training I* and *Training II* groups is in the standard deviation, which is half of what it is for the *No DSS* group.

In the last column of the tables, the Mean Clearing-to-Distance Ratio, it can be seen in Table 5 that the *No DSS* group performs the worst, with the other groups all over twice as high. For this metric, a higher value is considered better. The standard deviation is the lowest in the *No DSS* group. The reason for this is not clear. In Table 6, a similar trend is observed, with *the No DSS* group having the lowest score, and the other groups being higher. Again, the standard deviation is lowest in the *No DSS* group.

From these tables, it is clear that the DSS has a positive effect in some metrics over the *No DSS* group, but falls short of the effects of training and experience. Additionally, the standard deviation is largely unchanged over the *No DSS* group.

A closer look at the data will be provided through the use of box plots. Box plots for both the four-tenth and seven-tenth results are shown from Figs. 6–9. The median is represented by the horizontal line in the box plot, and the mean is represented by an X. The upper and lower whiskers denote the extreme values. The upper and lower bounds of the box correspond to the upper and lower quartile.

3.3.1 Mean Concentration Change. The results for the mean concentration change metric are compared in Figs. 6 and 7. In Fig. 6, there is a positive effect for the *DSS* group over the *No DSS* group with four-tenth concentration. In fact, the mean of the *DSS* group is similar to that of the *Seafarers* group, though a higher variance is seen. Of the participants using the DSS, the upper quartile performed better than the *Seafarers* group, although the lower quartile performed worse. The *Training I* and *Training II* groups are seen to have performed best of all.

In observing the results for the seven-tenth concentration groups, shown in Fig. 7, the difference is less pronounced between the *DSS* and *No DSS* groups. A larger spread is seen in the data with the *No DSS* group. The *Seafarers* group and the *Training II* group
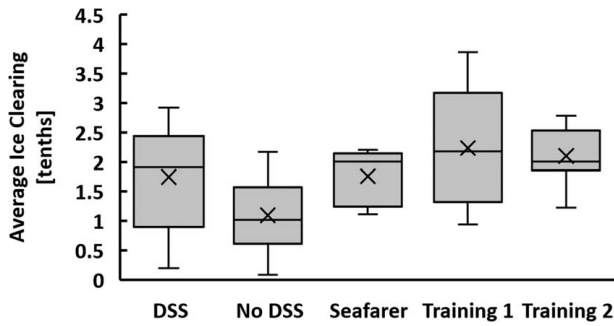
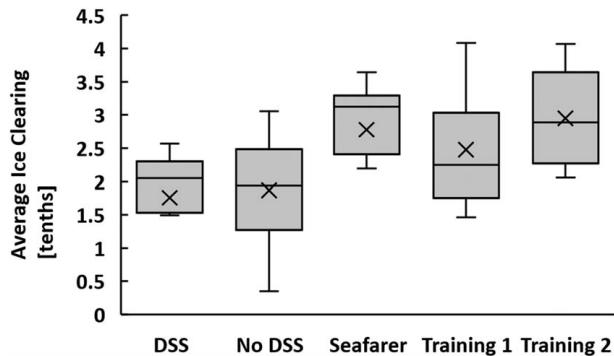**Fig. 6 The four-tenth mean concentration change**

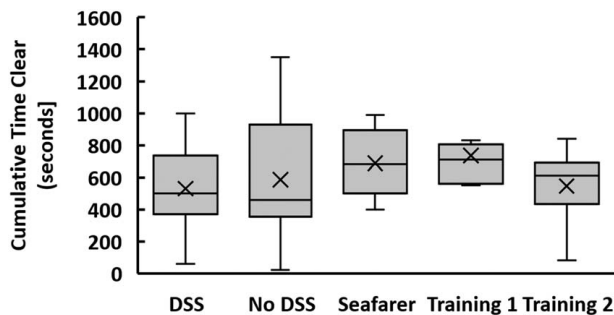**Fig. 7 The seven-tenth mean concentration change**

Fig. 8 The four-tenth cumulative time clear
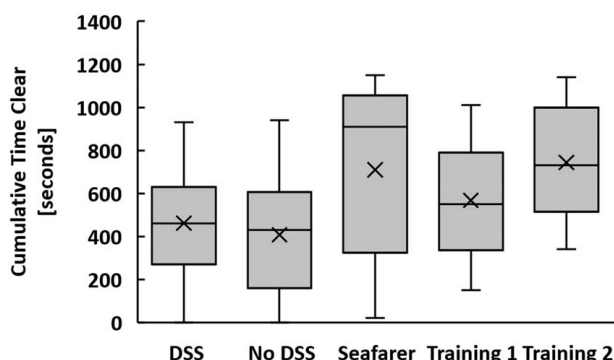

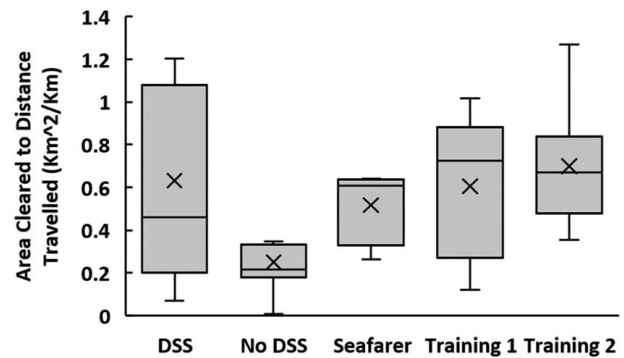**Fig. 9 The seven-tenth cumulative time clear**

Fig. 10 The four-tenth mean clearing-to-distance ratio

performed the best, although the *Seafarers* had a higher median value. The *Seafarers* group also had the lowest variance, indicating a more consistent result.

3.3.2 Cumulative Time Clear. The cumulative time that the smaller lifeboat launch zone is completely clear of ice is aggregated for each case and is shown for all groups. Figure 8 shows the cumulative time clear for the four-tenth ice concentration group. From this, participants in the *DSS* group performed the least well out of all groups. The effect of experience and training are shown to be the most effective. There is a decrease in performance with the *Training II* group over the *Training I* group. This is contrary to the trend seen in the other metrics and suggests that the *Training II* group performance for the four-tenth ice concentration of this metric is an outlier [2]. The best performance was observed in the *Seafarer* and *Training II* groups. The highest variance is observed in the *No DSS* group. In this, the whisker in the plot shows that one participant performed exceptionally well, and another participant performed exceptionally poorly, thus increasing the variance.

Figure 9 shows the cumulative time clear for the seven-tenth concentration. These results differ from the results of the mild ice concentration, showing a slight increase in the performance of the *DSS* group over the *No DSS* group. The *Seafarers* group performed best, followed by the *Training II* group. The variance is much higher in the seven-tenth concentration than in the four-tenth concentration.

3.3.3 Clearing-to-Distance Ratio. The clearing-to-distance ratio measures the reduction in ice concentration, in tenths, compared to the distance traveled by the participant's vessel. A larger value is considered best, since it represents a more efficient ice management performance. Figure 10 shows that the *DSS* group has a higher mean and median than the *No DSS* group. The DSS is not sufficient to bring the novice participants to the level of experienced seafarers or inexperienced seafarers who have been provided with specific training; however, some participants in the *DSS* group performed exceptionally well. A large variance can be seen in the results from the *DSS* group. The highest mean is observed in the *Training II* group; however, the median is higher in *Training I*. Both training groups performed better than the *Seafarers*. The *No DSS* group has the worst performance.

Figure 11 demonstrates the results for the clearing-to-distance ratio with the groups in severe ice conditions. The *Training I* and *II* groups and *Seafarers* are significantly more successful than both the *No DSS* group and the *DSS* group. It can be observed that the *DSS* group is slightly more effective than the *No DSS* group; however, the difference is slight. The *Training II* group performs the best of all, followed by *Training I* and *Seafarers*.

3.4 Qualitative Results. Qualitative results can provide great insight into the strategies adopted by the various groups and how these strategies influenced their quantitative results. Additionally, exit surveys can provide valuable insight as to the decision-
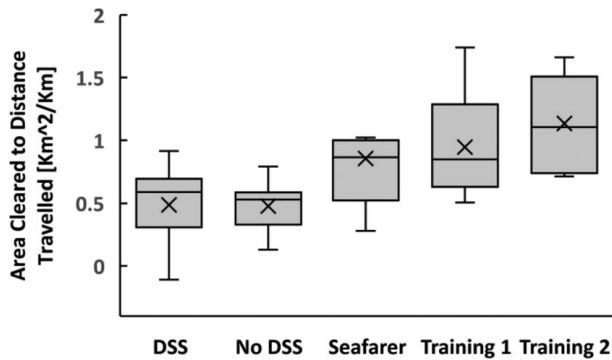
Fig. 11 The seven-tenth mean clearing-to-distance ratio

making of the participants and which factors of the experiment they found useful. This section describes the participants' strategies using heat maps and outlines the insights from the exit interviews.

3.4.1 Heat Maps. Heat maps are an effective way of qualitatively indicating the strategy or approach used by each group. To create the heat maps, each participant's trajectory in the simulator was overlaid onto the next. Pixels where more cumulative time is spent across all cases were given a lighter color. This provided some insight into the group's behavior and strategy.

Heat maps of each cohort for the seven-tenth scenario can be seen in Fig. 12. From this, it can be observed that the *No DSS* group has the least concentration when compared to the other cohorts and generally has a larger footprint in the figure. This lack of clustering near the FPSO, as seen by the light regions in the other cohorts, indicates that there was less focused group behavior and more variation between each case. The *DSS* group appears to be clustered just upstream of the lifeboat launch zone, a trend that was also found in both *Training I* and *II* groups, and in the experienced *Seafarers* group.

One of the most effective ice management techniques is to position one's vessel up-current from the zone that is to be cleared and position the vessel in a way that blocks the ice when holding position. This is a strategy that is not obvious to an inexperienced and untrained participant. For experienced seafarers, it is a well-known technique [6]. Due to the efficacy of this method, known as the leeway technique, it was incorporated into the curriculum developed by Thistle et al. [2] for use in the training provided to the *Training I* and *Training II* cohorts. Additionally, this method was recommended in almost all cases by the DSS [15].

3.4.2 Exit Interview Results. Exit interviews were a valuable way of recording expert tactics, the utility of training, and gauging the perceived efficacy of the DSS, expert techniques, and training efficacy. They were also used to deduce which components of the DSS could be improved and which training techniques were best.

In Experiment I, the exit interviews were able to determine the most effective expert tactics. The highest-performing results could be cross-referenced with the exit interview results to determine the strategies and techniques used. The replay video of the expert's simulated scenario was then extracted and used to inform both the training curriculum and the DSS development [10,17].

An exit interview was also conducted for the participants in Experiment II. Questions were asked as to the strategies of the participants and whether the training was found to be helpful or adequate. Questions were also asked as to what could be changed with the training to improve it in the future [17].

In Experiment III, all the participants were asked to gauge the utility of the DSS, and all participants reported that the DSS was generally helpful with an average rating of 4.2 out of 5. Of the
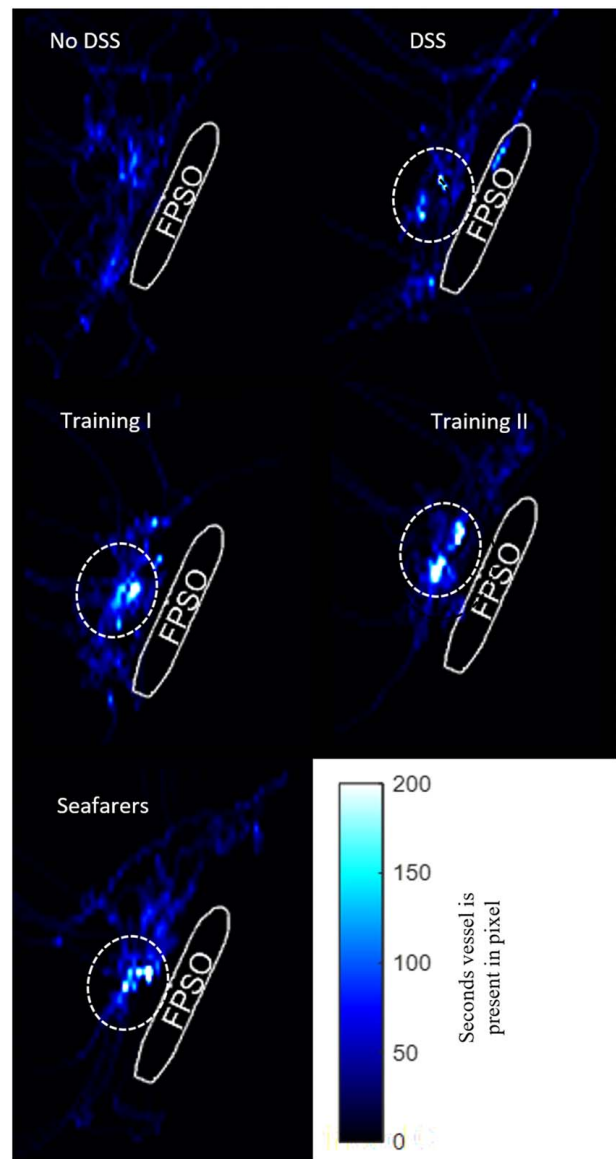

Fig. 12 Position heat maps of each cohort in the severe ice concentration metric. The shade corresponds to the number of seconds a vessel is in each pixel, as shown in the legend.

components of the DSS graphic user interface (GUI), 12 of the participants exclusively used the central looping video that showed the most effective strategy to follow given their current position. Three participants found that the text instructions were useful, and two participants used the suggested solution section of the DSS. The suggested solution section gave recommendations on the speed, heading, and approach to use. A common observation from all participants was that they found the DSS was too cluttered and distracting [18].

4 Discussion

The results of this analysis provide insights into the factors that influence ice management performance. The most effective expert strategy as identified in Experiment I will be discussed, as will its applicability to the development of training for inexperienced participants. The findings of Experiment III can be extrapolated to inform improvements for future DSS design.

Experience was found to be highly effective at influencing ice management performance as demonstrated by the superior

performance of experienced participants over novice participants in Experiment I. It should be noted that the upper quartile or upper outlier novices were often able to meet the mean seafarer performance, indicating that although the average expert performance was superior, the difference was not as great as expected in some cases [3].

Training was demonstrated to be an extremely effective way of increasing the performance of novices up to and often exceeding that of experts with only several hours of training over one or two sessions. The implication of this is that a well-designed training program can be extremely effective at improving the performance of inexperienced seafarers [2].

The DSS was found to positively influence participant strategies and performance. Further, the results provide evidence to answer the questions posed in this study. First, the goal was to determine which factors were the most effective at influencing ice management strategy. It is clear from both quantitative and qualitative metrics that training is a highly effective way to improve ice management performance, followed closely by experience for the cases studied here. The training regimes provided participants with the skills needed to effectively manage ice in this specific simulated scenario.

The second question pertained to the efficacy of the DSS. It was found that the DSS did positively influence performance, but these results were not as significant as those of the participants who were provided with training or who had more experience.

4.1 Participant Strategy Across All Groups.
The strategies of participants in Experiments I, II, and III will be discussed here. All participant groups, with the exception of the *No DSS* group, tended to position themselves in the area up-current from the lifeboat launch zone. This technique was first identified in Experiment I, as it was used by the most effective expert participants and yielded the best results. In this experiment, the *No DSS* group, being the least experienced and not being provided with any navigation aids or training, were expected to have the lowest performance, extreme values notwithstanding. Although the *DSS* group strategies did not translate into highly significant performance improvements, observational improvements are evident in the box plots. In the case of the *Training I* and *Training II* groups, performance met and often exceeded the performance of the *Seafarer* group in some metrics [2]. This indicates that training can be very effective at influencing performance for specific tasks, as it was capable of taking novices with little, if any, sea experience to the level of experienced captains in a simulated scenario. This strategy was communicated to the *DSS* group, and the *DSS* participants who made use of this strategy were more effective in their performance.

With this in mind, if the *DSS* group demonstrated tactics similar to those of the experienced seafarers and trained novices using these techniques, it raises the question of why the difference in their performance was not significantly better than the *No DSS* group. Examining replay videos of the cases can provide a partial explanation. The DSS explicitly recommends using the leeway technique as employed by many of the other groups and, in fact, demonstrates the best of these cases from Experiments I and II in the form of the replay video. While the *DSS* group tried to follow this advice and used similar expert strategies employed by the seafarer groups and trained group, they were slower in their responses and seemed to be uncomfortable getting as close to the FPSO as the other groups.

This resulted in many of the *DSS* group positioning their vessel too far away from the FPSO to prevent ice floes from drifting in between their hull and the hull of the FPSO. These ice floes would often drift into the lifeboat launch zone, resulting in a lower score. An example of this can be seen in Fig. 4, where an ice floe is positioned just north of the lifeboat launch zone and is poised to drift southwards into it. It is interesting to note that in Experiment II, this issue was initially seen with both training groups early into their training; however, by the time they

reached the testing scenario used for data collection, they were more comfortable operating close to the FPSO [2]. This further indicates that training was more effective than the DSS for the inexperienced participants.

Employing the leeway method resulted in less distance traveled since the vessel maintains the same position for much of the case. When performed effectively, this can result in a high clearing-to-distance ratio, as a large amount of ice can be cleared from the zone without requiring too much travel [3]. Before entering the pack ice, the *Seafarers* were much more likely to have a plan or strategy in mind, and they were able to adapt to a secondary strategy if they found their initial approach to be ineffective [9]. This strategy was not observed in the *No DSS* group. The *Training I* and *Training II* groups were taught the most effective ice management strategies and were allowed to practice them, while being provided with performance feedback. This was very effective, as can be seen by the heat maps and improved performance [2]. The *DSS* group who requested assistance from the DSS before entering the pack ice were also able to formulate a plan; however, many stated that they had a high degree of uncertainty about what this plan would entail in practice, as it was their first time using the simulator in ice. That said, despite no significant quantitative performance improvements, it can be seen from the heat maps in Fig. 12 that the *DSS* group were concentrated near the lifeboat launch zone in a manner which more closely resembled the experts or trained participants.

This uncertainty contrasts with the results from Experiment I of the *Seafarer* group (and the experienced seafarers interviewed in Ref. [10]), who stated in exit interviews that they often had a primary strategy as well as a secondary backup plan in place if required. Through their experience, they were able to ascertain whether their primary strategy was not likely to succeed and could switch to their secondary strategy with sufficient time to finish the scenario effectively. These results further demonstrate the effect of experience on ice management performance [3].

4.2 Training.
The training curriculum for Experiment II was designed using the results of Experiment I and recommendations of experienced seafarers. The development of the curriculum was completed using a formal development approach, including a needs assessment, curriculum design, and evaluation of curriculum effectiveness [2]. Observations from Experiment I demonstrated that participants with poor performance results had no discernable technique, while those who performed well used one or a combination of three techniques. These were the pushing, propwash, and leeway techniques, and the basis of the training curriculum taught these three methods [2]. These methods were also included in the DSS advice which aimed to communicate the expert strategies featured in the training in real time to participants.

4.3 Decision Support System Equipment Design.
Exit interview observations indicated that the participants in the *DSS* group preferred the pictorial information in the graphic user interface (GUI) over the text instructions. They generally found that the text-based instructions were too distracting and that the GUI was cluttered and distracting, so most gravitated toward the video in the center of the display. The few who did read the text-based instructions in conjunction with the video did find the information helpful, as each bullet point instruction is specific to the video. The implication of this is that future work on the DSS should take more care to incorporate best practices of HMI design and should avoid overloading the HMI with information [18]. Information overload can be caused by computer-based systems with a high number of available features displayed all at once. This can lead to a cluttered display which can cause difficulty for a user attempting to prioritize information correctly [19]. The DSS provided to users in Experiment III presented the same data in multiple forms, and several users reported that it could be distracting at times.

A more comprehensive habituation of the DSS could also be beneficial so that participants have a better understanding of the functionality and limitations of the DSS. This is a point that is supported by Bainbridge's Ironies of Automation (1983), which discusses how a lack of understanding of automated systems in all industries may negatively affect the skills of the operators of these systems [20]. Further, Nilsson et al. (2009) studied the effects of technology on the safety of marine operations and observed findings that should be tested with future DSS participants [13]. It was found that when performing simulated fairway navigation scenarios, experienced navigation officers performed better with conventional ship bridges over technically advanced bridges. The opposite was found to be true for inexperienced officers [14]. If the bridge simulator without the DSS is considered to be a conventional bridge, while the DSS outfitted bridge can be considered to be more technically advanced, a similar study of DSS technology could be undertaken to examine the effects of a DSS on varying levels of seafarer experience to test whether this conclusion holds.

4.4 Future Improvements to the Decision Support System.
Future work to improve the DSS could incorporate a more user-friendly method for requesting assistance, such as a physical button instead of a trackpad. This would reduce the need for the user to shift their attention from the task to search for a virtual button on the monitor. The use of physical buttons over virtual ones is encouraged and supported by the findings from the marine accident of the U.S.S. John S McCain, which found that an overreliance on digital throttle controls on a touch screen contributed to a loss of situational awareness by the crew, contributing to the collision [21,22]. The HMI should be designed in such a way that it matches the style of the simulator display and controls, to avoid the inconsistencies often found in multivendor bridge systems [22]. Future work would also seek to reduce text on the DSS or replace it with pictograms. It may also be beneficial to participants to provide the option of controlling the speed at which the replay video is played. The current format is that the video plays at 30 times the original speed, in a six-second loop. It was reported by several users in their exit interviews that this was too fast and that they would have liked the opportunity to pause the playback. This could be enabled through a simple physical knob control so that easy playback speed alterations could be done. A play/pause button could also be included. This would provide participants with the ability to customize the format of the feedback they are given.

Further improvements could be made to the CBR algorithm in the DSS. In some situations where a participant's approach is far removed from an approach in the case-base, the algorithm may suggest an approach that is not easy to achieve from the user's current position. This is an issue of automation reliability, where the automated system does not operate or perform well in certain situations, which can lead to distrust by the user [23]. For example, one participant was given a suggestion by the DSS which was not effective given their position at the time. They decided to no longer use the DSS and performed the rest of the task without it.

The current version of the DSS uses a simplified CBR algorithm which does not revise or retain the solutions based on new data. This approach was chosen to avoid altering the advice of the DSS between participants; however, it limits the ability of the DSS to improve when provided with new data. Retaining the results of new users and integrating them into the case-base would likely result in improved performance of the CBR. Additionally, integrating a revision step into the CBR algorithm (as it is designed to include) where the advice is updated, could result in improvements.

Future work to improve the DSS could involve a more formalized human-centered design through an extensive literature review and a formalized incremental design approach. For example, steps could also be taken to validate the results of the

DSS strategies with subject matter experts (SME) in the form of interviews or simulator experiments, similar to the experiments performed here. As these experiments with the DSS were performed using novices, a gap remains as to whether the DSS performance would be acceptable to SMEs. SME engagement in future work would be highly beneficial to optimize and improve the HMI design.

Improving on these findings could result in a tool that could be valuable to an onboard navigation team or a maritime college using simulators for training. An effective DSS could complement traditional navigation practices on board and complement the teachings of an instructor. That said, it is important to recognize that a single prototype DSS was tested in Experiment III, and as such, these results should not be generalized or used to validate the use of other forms of DSS.

5 Conclusion

Overall, this research concludes that while experience is important to seafarer performance, a well-planned and targeted training program can significantly improve the performance of less experienced operators for specific tasks. When the DSS was tested, it was found to be effective at influencing ice management strategies to be similar to those of experienced seafarers. Nevertheless, performance improvements were limited and participants did not appear to be proficient in the higher order decision-making skills that come with experience. All three experiments in this analysis tested factors that can lead to improved ice management performance and identified ways in which ice management strategies can be communicated to inexperienced operators. The results from Experiment III provide insights as to which components of the DSS were perceived as useful, and how future iterations can be improved to provide tactical advice clearly and concisely. Visual displays such as video-based replays and pictures were preferred by most of the participants and these should be emphasized going forward over text-based instruction.

The authors recognize that there are other factors which should be considered for future DSS development and testing, including tactical feedback and control. In the DSS tested, providing a visual animation of an optimal solution was found to be very helpful to participants and should be improved upon, while text and unnecessary numerical data should be reduced. Additionally, future iterations of the DSS should expand the case-base by retaining new cases. Revising cases to better fit an optimal solution would improve the accuracy of the DSS advice.

Ultimately, despite promising results from the DSS, the factors of experience and training were demonstrated to be the most effective at influencing ice management performance. Training, when provided to inexperienced participants was able to significantly improve performance such that it was often on par or better than that of experienced seafarers. The DSS was able to positively influence strategies, but had a marginal effect on performance when compared to these other factors of experience and training.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Montgomery, D. C., 2013, *Design and Analysis of Experiments*, John Wiley & Sons Inc., Hoboken, NJ.

[2] Thistle, R., and Veitch, B., 2019, "An Evidence Based Method of Training to Targeted Levels of Performance," Proceedings of the SNAME Maritime Convention, Tacoma, WA, Oct. 30–Nov. 1.

[3] Veitch, E., Molyneux, D., Smith, J., and Veitch, B., 2019, "Investigating the Influence of Bridge Officer Experience on Ice Management Effectiveness Using a Marine Simulator Experiment," ASME J. Offshore Mech. Arct. Eng., **141**(4), p. 041501.

[4] Lützhöft1, M., Porathe, T., Jenvald, J., and Dahman, J., 2010, "System Simulations for Safety," Proceedings of the International Conference on Human Performance at Sea, Glasgow, June 16–18, p. 3.

[5] Power, S., Power, J., MacKinnon, S., and Simões Ré, A., 2010, "Effect of Simulator Training on Novice Operators' Abilities to Navigate in Ice," Proceedings of the SNAME 9th International Conference and Exhibition on Performance of Ships and Structures in Ice, Anchorage, AK, Sept. 20–23.

[6] Dunderdale, P., and Wright, B., 2005, *Pack Ice Management on the Southern Grand Banks Offshore Newfoundland, Canada*, Noble Denton Canada Ltd., St. John's, Canada.

[7] Kolodner, J. L., 1992, "An Introduction to Case-Based Reasoning," Artif. Intell. Rev., **6**(1), pp. 3–34.

[8] Smith, J., Yazdanpanah, F., Thistle, R., Musharraf, M., and Veitch, B., 2020, "Capturing Expert Knowledge to Inform Decision Support Technology for Marine Operations," J. Mar. Sci. Eng., **8**(9), p. 689.

[9] Yazdanpanah, F., 2021, "Designing a Cased Based Reasoning Decision Support System for Ice Management Operations Using Expert Knowledge," Master's Thesis, Memorial University of Newfoundland, St. John's, Canada.

[10] Cohen, J., 1969, *Statistical Power Analysis for the Behavioural Sciences*, Academic Press, New York.

[11] Cohen, J., 1992, "Statistical Power Analysis," J. Clin. Psychiatry, **51**(3), pp. 98–101.

[12] Kennedy, R., Lane, N., Berbaum, K., and Lilienthal, M., 1993, "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness," Int. J. Aviat. Psychol., **3**(3), pp. 203–220.

[13] Nilsson, R., Gärling, T., and Lützhöft, M., 2009, "An Experimental Simulation Study of Advanced Decision Support System for Ship Navigation," Transp. Res. Part F: Traffic Psychol. Behav., **12**(3), pp. 188–197.

[14] Soper, J., 2022, "An Investigation of the Influence of a Decision Support System on Simulated Ice Management Performance," Master's Thesis, Memorial University of Newfoundland, St. John's, Canada.

[15] Lakens, D., 2013, "Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs," Front. Psychol., **4**.

[16] Soper, J., Smith, J., Power, J., and Veitch, B., 2022, "An Evaluation of Decision Support Technology in Simulated Offshore Ice Management," Conference on Ocean, Offshore, and Arctic Engineering, Hamburg, Germany, June 5–10.

[17] Thistle, R., 2019, *An Evaluation of the Effects of Simulator Training on Ice Management Performance*, Memorial University of Newfoundland, St. John's.

[18] Lee, D. J., Wickens, C. D., Liu, Y., and Boyle, N. L., 2017, *Designing for People: An Introduction to Human Factors Engineering*, CreateSpace, Charleston, South Carolina.

[19] Dung Vu, V., and Lützhöft, M., 2019, "Frequency of Use—The First Step Toward Human-Centred Interfaces for Marine Navigation Systems," J. Navig., **75**(5), pp. 1–19.

[20] Bainbridge, L., 1983, "Ironies of Automation," Automatica, **19**(6), pp. 775–779.

[21] National Transportation Safety Board, 2017, *Collision Between US Navy Destroyer John S McCain and Tanker Alnic MC*, National Transportation Safety Board, Washington.

[22] Nordby, K., Mallam, S. C., and Lützhöft, M., 2019, "Open User Interface Architecture for Digital Multivendor Ship Bridge Systems," WMU J. Marit. Aff., **18**(2), pp. 297–318.

[23] Mallam, S. C., Nazir, S., and Sharma, A., 2020, "The Human Element in Future Maritime Operations—Perceived Impact of Autonomous Shipping," Ergonomics, **63**(3), pp. 334–345.