Defence Research and Development Canada | Recherche et développement pour la défense Canada

CAN UNCLASSIFIED

DRDC | RDDC
technologysciencetechnologie

An X-Ship Study on Crew Work and Rest Schedule and Fatigue aboard a Halifax-Class Patrol Frigate

Wenbi Wang, PhD
Matthew Lamb
Renee Chow, PhD
Henry Peng, PhD
Fethi Bouak, PhD
DRDC – Toronto Research Centre

Defence Research and Development Canada

Scientific Report
DRDC-RDDC-2018-R023
March 2018

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by Defence Research and Development Canada (DRDC) using the Schedule to the *Defence Production Act*.

The data collected as part of this study was approved either by Defence Research and Development Canada's Human Research Ethics Board or by the Director General Military Personnel Research & Analysis' Social Science Research Review Board.

Endorsement statement: This publication has been peer-reviewed and published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada. Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

Abstract

Sailors' work and rest schedules, and fatigue were studied onboard Her Majesty's Canadian Ship (HMCS) Montreal during a 10-day at-sea trial. We examined the crew's work hours and workload, sleep duration and quality, as well as daily fatigue and mood. The results indicated a significant increase of fatigue at the end of the trial, relative to the beginning of the trial. Significantly different patterns of results emerged when the data were analyzed by grouping participants according to their department, watch or rank. An elevated risk for fatigue was identified for watch keepers, particularly those on the bridge. Results from this study support the ongoing validation of a crewing analysis tool SCORE and provide a baseline for assessing crewing options in future X-Ship trials.

Significance to defence and security

Simulation for Crew Optimization and Risk Evaluation (SCORE) is a crew modelling software developed by the Defence Research and Development Canada (DRDC) for the Royal Canadian Navy (RCN).The results from this study will be used to validate two key algorithms in SCORE to predict crew sleep and performance which is critical for generating or evaluating robust crewing solutions for future RCN platforms. The fatigue data that were collected in this study will also serve as a baseline to assess the quality of alternative crewing options in future X-Ship crewing trials.

Résumé

Les horaires de travail et de repos ainsi que la fatigue des marins ont été étudiés à bord du Navire canadien de Sa Majesté (NCSM) *Montréal* durant une période d'analyse en mer de dix jours. Nous avons examiné les heures et la charge de travail des membres d'équipage, la durée et la qualité de leur sommeil, ainsi que leur fatigue et leur humeur au quotidien. Les résultats ont révélé une augmentation notable de la fatigue à la fin de la période d'analyse, par rapport au début de la période. Des modèles considérablement différents se sont dégagés de l'analyse des données par groupe de participants en fonction de leur service, de leur quart de travail ou de leur grade. Il existe un risque élevé de fatigue pour le personnel de quart, particulièrement sur la passerelle. Les résultats de cette étude vont dans le sens de la validation actuelle de l'outil d'analyse d'affectation des équipages SCORE et constituent une assise pour évaluer les options en matière d'affectation des équipages lors des futurs essais du navire X.

Importance pour la défense et la sécurité

L'outil Simulation pour l'optimisation de l'équipage et l'évaluation des risques (SCORE) est un logiciel de modélisation de l'équipage élaboré par Recherche et développement pour la défense Canada (RDDC) pour la Marine royale canadienne (MRC). Les résultats de cette étude serviront à valider deux principaux algorithmes de l'outil SCORE pour prédire le sommeil et le rendement des membres d'équipage, ce qui est essentiel pour créer ou évaluer des solutions robustes en matière d'affectation des équipages des futures plateformes de la MRC. Les données sur la fatigue qui ont été recueillies dans le cadre de cette étude serviront de base à l'évaluation de la qualité des options de rechange en ce qui concerne l'affectation des équipages lors des futurs essais dans ce domaine du navire X.

Table of contents

List of figures

List of tables

Acknowledgements

We are grateful to the commander and crew of HMCS *Montreal* for their participation in this study. This work would not have been possible without their outstanding support. We would also like to thank D Nav P&T, DNCI, Sea Trainers and DRDC colleagues for their assistance in this study: particularly Cdr Plaschka, LCdr Versteeg, CPO1 Bisal, CPO1 Labbé, Doug Saunders, John Townsend, Natalia Dubova, and Cpl Sheridan Torres.

1 Introduction

1.1 Background

The Royal Canadian Navy (RCN) is undertaking one of its most extensive fleet modernization and replacement programs in its history. Three ongoing acquisition projects, including the Joint Support Ship project (JSS), the Arctic/Offshore Patrol Ship project (AOPS), and the Canadian Surface Combatant project (CSC), will define the shape of the RCN for decades to come. One of the human factors issues considered in these acquisition projects is platform crewing which is concerned about how changes in crew size and job specifications on these future platforms affect their operational effectiveness.

To address this issue, a comprehensive approach has been adopted that involves a variety of investigation methods, ranging from Simulation and Modelling (M&S) to Human-In-The-Loop (HITL) experimentation. The Defence Research and Development Canada (DRDC) has been actively working with the RCN and providing assistance to ongoing crewing studies.

On M&S, a crewing analysis software, SCORE (Simulation for Crew Optimization and Risk Evaluation) has been developed for the RCN to conduct what-if assessment on crewing options [1]. This modelling tool allows analysts to articulate work requirements on an RCN vessel, its crewing options, alternative work assignment schemes in a computational model, and investigate the impact of crewing options using metrics such as utilization rate and work assignment conflict. With participation by RCN subject matter experts, particularly D Nav P&T, crewing models have been developed for JSS, AOPS and CSC [2–5].

Currently a research and development effort is taking place to expand SCORE by introducing a DRDC Fatigue Model (DFM) into this modelling software. The addition of DFM will enable users to assess crewing options by analyzing their impact on crew performance. Specifically, DFM uses two algorithms to generate performance prediction for each member of the crew based on one's assigned work schedules [6]. A fundamental pillar of DFM is a physiological model of human fatigue. Its development has leveraged extensively on past research conducted in the aviation domain. To use such a model for RCN applications, research is needed to verify and validate the model using data directly collected in maritime operations (e.g., [7]).

On HITL experiments, a series of crewing trials have been planned by the RCN with a general objective to study alternative work schedules and manning options, from specific departments to the entire crew. Typically in such an experiment, the impact of experimental manipulations (e.g., a change of crew size) will be examined using measures on sailors' physical or mental state, and work performance in a naturalistic setting during a naval exercise. One common challenge in designing such a study is that a performance redline (i.e., a criterion used to judge whether an option is acceptable or not) is hard to define in such a loosely controlled experimental setting. One solution is to collect a suite of benchmark data, based on operational experience onboard existing RCN platforms. Such data can be used in future crewing trials as a baseline to judge the quality of novel crewing concepts.

The RCN's Experimental-Ship (X-Ship) programme has provided an opportunity to collect empirical data for validating DFM and benchmarking future crewing trials. A human factors study was conducted on the modernized Halifax-class frigate HMCS Montreal between Oct 25th and Nov 3rd, 2016. During this study, the platform was operated based on a crew size of 216.[1] We sampled a large number of crew members and studied their work, rest patterns and fatigue levels during the 10-day trial. This Scientific Report is one of the publications produced from this trial. This work is supported by the DRDC project 01ab which is entitled “RCN crewing and human factors.”

1.2 Objectives

The purpose of this analysis was to examine sailors’ fatigue under the 217-crew model, analyze primary causes of fatigue, and establish a baseline dataset regarding the crew’s work and rest patterns under the specific operational scenario. Specifically, the analysis was focused on quantifying the following elements of this naval operation.

1. Work requirements for the crew, including both work duration and workload.

2. Rest quality, particularly daily sleep duration and disruption.

3. Changes in subjective fatigue and mood.

These topics of interest were investigated in this study based on a categorization of participants according to their department, watch and rank groupings.

1.3 A framework for naval crewing analysis

The study was designed based on a conceptual framework that we propose for guiding naval crewing analysis. As illustrated in Figure 1, fundamental issues of concern in a crewing study are the crew’s operational performance and their well-being. In the case of a full-ship crewing study, i.e., the type of analysis that is supported by SCORE, we suggest that a metric based on fatigue is most suitable as it can be applied to all members of the crew regardless of the types of work they perform which can differ significantly. According to International Maritime Organization (IMO), fatigue can be defined as “a state of feeling tired, weary, or sleepy that results from prolonged mental or physical work, extended periods of anxiety, exposure to harsh environments, or loss of sleep, and can impair a crew member’s alertness and ability to perform” [8]. The concept of fatigue allows us to use a single measure to compare crew effectiveness across ship departments, rank levels, and between existing and future platforms.

In a crewing study, the independent variables that are subject to design manipulation include both crew size and crew composition. The impact of such manipulations is reflected in work assignment, particularly two work-related variables, i.e., work duration and workload. Subsequently, work duration and its scheduling dictate the crew’s opportunity to rest. In the context of naval operation, the dominant factor to consider for sailors’ restorative rest is sleep. These three variables, i.e., work schedule, workload and sleep, are the primary factors to consider in the proposed framework. They are represented as bold arrows in the figure to highlight their significance in this study.

There exist other factors that influence fatigue and its subjective perception, such as environmental stressors (e.g., ship motion) and fatigue mitigation methods (e.g., alertness-promoting means like caffeine) [9]. While these factors themselves are commonly not a main subject of investigation in a crewing study, they should be considered in the experimental design to ensure their impact on fatigue is properly analyzed.

[1] There was one vacant position relative to the full complement of 217.

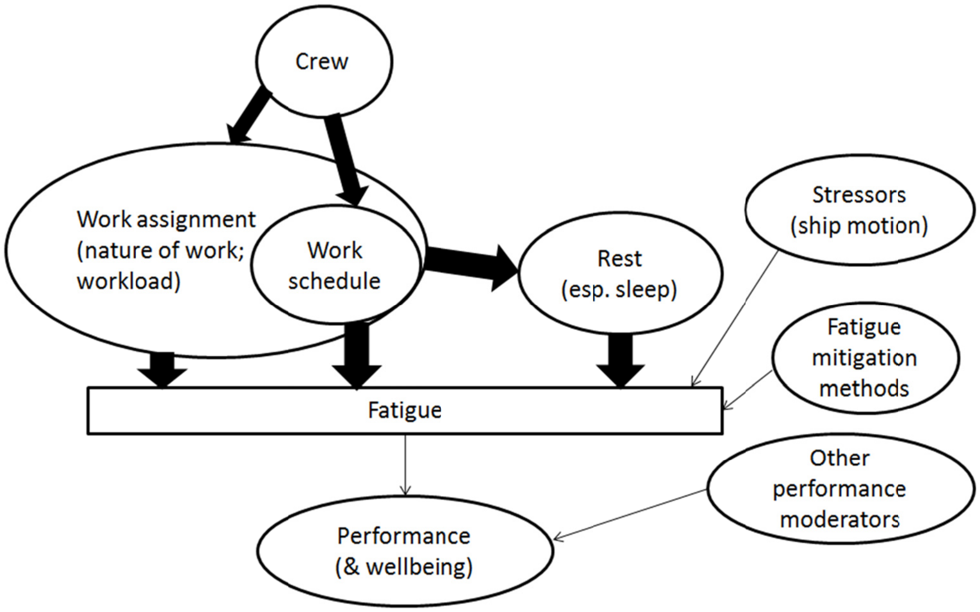


Figure 1: *A framework for guiding the design of naval crewing studies.*

It is useful to point out this framework does not consider performance moderators beyond fatigue. For example, individual characteristics such as their competencies, skills and abilities often have a significant impact on task performance. These factors are not included in this framework since they are currently not represented in the crewing model.

Surveys and questionnaires used in this study were selected or designed according to this framework. A detailed description will be provided in the next section.

2 Methodology

The section describes the trial methodology for this study. Its experimental protocol (#2016-036) was approved by the DRDC Human Research Ethics Committee.

2.1 Participants

A total of 160 volunteers from the HMCS Montreal participated in this study, which represented 74% of the entire ship's crew. They were sampled from six ship departments and across different watch syndicates. Table 1 is a summary of the demographic information for all participants, including their age, gender, rank, the ship department and to which the watch team they belonged.

Table 1: *Summary of participants' demographic information.*
Data are presented as Frequency (%) or Mean (M)± Standard Deviation (SD).

Demographic factor	Variable	Value
	Participant number	160
	Age	34.5 ±8.3
	Age (max, min)	(54, 19)
Gender	Male	139 (87%)
	Female	21 (13%)
Rank		
Jr Sailors	OS	13
	AB	13
	LS or CPL	50
	MS or MCPL	20
Sr Sailors	PO2 or Sgt	23
	PO1 or WO	17
	CPO2	5
	CPO1	1
Officers	SLt	2
	Lt(N)	14
	LCdr	1
	Cdr	1
Ship department	Combat (CBT)	59 (37%)

Demographic factor	Variable	Value
	Marine System Engineering (MSE)	36 (23%)
	Combat System Engineering (CSE)	27 (17%)
	Logistics (LOG)	23 (14%)
	Deck (DECK)	10 (6%)
	Executive (EXE)	5 (3%)
Watch syndicate	1-in-2 Starboard Watch Keeper (STBD)	50 (31%)
	1-in-2 Port Watch Keeper (PORT)	49 (30%)
	1-in-3 Bridge Watch Keeper (BWK)[2]	5 (3%)
	Dayworker or on-call (Dayworker)	56 (35%)

In this study, we also collected data regarding a participant's chronotype which is a personal characteristic that reflects an individual's preference in the timing of sleep and other behaviours. It is associated with one's circadian rhythm and is measured on a continuum from morningness to eveningness [10]. A Morningness-Eveningness Questionnaire (MEQ) was used to analyze a participant's chronotype and the result is categorized into five groups, that is, definite morning, moderate morning, intermediate, moderate evening, and definite evening, according to the time of day when one's daily peak alertness occurs.

A total of 156 responses were collected and the results are summarized in Table 2. Overall, 60.0% of participants were identified to be the intermediate type with the daily peak alertness occurring between morning and evening. Twenty-six point three percent (26.3%) were moderate morning type and 7.5% were moderate evening type. Only 6 participants were definite morning or definite evening type which together was accountable for 3.7% of all participants. The table also provides a detailed breakdown for three participant groupings investigated in this study, i.e., by department, watch syndicate, and rank level.

Table 2: *Summary of participants' chronotype.*

	Definite morning		Moderate morning		Intermediate		Moderate evening		Definite evening		Response missing	
	N	%	N	%	N	%	N	%	N	%	N	%
All participants	5	3.1%	42	26.3%	96	60.0%	12	7.5%	1	0.6%	4	2.5%
CBT	0	0.0%	14	23.7%	40	67.8%	5	8.5%	0	0.0%	0	0.0%
CSE	3	11.1%	8	29.6%	12	44.4%	4	14.8%	0	0.0%	0	0.0%
DECK	0	0.0%	4	40.0%	5	50.0%	1	10.0%	0	0.0%	0	0.0%
EXE	0	0.0%	3	60.0%	1	20.0%	1	20.0%	0	0.0%	0	0.0%
LOG	1	4.3%	7	30.4%	12	52.2%	0	0.0%	0	0.0%	3	13.0%
MSE	1	2.8%	6	16.7%	26	72.2%	1	2.8%	1	2.8%	1	2.8%
PORT	1	2.0%	10	20.4%	35	71.4%	2	4.1%	1	2.0%	0	0.0%

[2] Bridge watch keepers stood a 1-in-3 watch during this study which was comprised of three teams, each with a unique work/rest schedule. Due to the small size of the participant pool, they were grouped together and analyzed as a team.

	Definite morning		Moderate morning		Intermediate		Moderate evening		Definite evening		Response missing	
	N	%	N	%	N	%	N	%	N	%	N	%
STBD	2	4.0%	13	26.0%	28	56.0%	7	14.0%	0	0.0%	0	0.0%
BWK	0	0.0%	1	20.0%	3	60.0%	1	20.0%	0	0.0%	0	0.0%
Dayworker	2	3.6%	18	32.1%	30	53.6%	2	3.6%	0	0.0%	4	7.1%
Jr Sailors	5	5.2%	21	21.9%	58	60.4%	9	9.4%	1	1.0%	2	2.1%
Sr Sailors	0	0.0%	18	39.1%	27	58.7%	0	0.0%	0	0.0%	1	2.2%
Officers	0	0.0%	3	16.7%	11	61.1%	3	16.7%	0	0.0%	1	5.6%

2.2 Data collection instruments

Survey questionnaires were the primary data collection instrument used in this study. Three sets of surveys were administered, including a pre-trial survey, a post-trial survey, and a daily log for each trial date. The sleep patterns of a subset of participants were tracked by a wrist actigraph.

2.2.1 Pre-trial survey

The pre-trial survey was filled out once at the beginning of the trial. It is comprised of the following questionnaires.

Morningness-Eveningness Questionnaire (MEQ): a self-assessment questionnaire to measure a participant's chronotype, i.e., the circadian rhythm of an individual that influences the cycle of sleep and activity in a 24-hour period. Nineteen questions are included in MEQ to analyze whether a person's biological clock (circadian rhythm) produces peak alertness in the morning, in the evening, or in between [11].

Modified Fatigue Impact Scale (MFIS): a survey instrument commonly used in fatigue research to analyze a participant's functional limitations due to fatigue experienced within the previous two weeks [12]. It is comprised of 21 questions that examine the impact of fatigue in three domains, including physical, mental, and psychosocial.

Work Requirement Survey (WRS): a customized questionnaire developed in this study to analyze job requirements onboard a RCN platform. It is comprised of 52 questions that correspond to the set of elementary skills described by the Fleishman's taxonomy of human performance [13]. Participants are asked to rate the importance of each skill in relation to their work requirements on a five-point scale.

Baseline Sleep Survey (BSS): a survey developed in this study to record a participant's sleep pattern for three days prior to the start of the trial. It breaks down each day into 48 blocks, each with a duration of 30 minutes. One's sleep schedule is recorded by coding all time blocks that are used for sleep.

2.2.2 Daily log

A daily log was developed specifically for this study. It is two pages in length and participants were asked to fill out one log for each day during the study. The log is designed to record the following six elements in relation to a participant's work schedule, rest quality and perceived fatigue. All questionnaires and rating scales used in the daily log are obtained from past research (e.g., [7]), with minor customizations to support the current study.

1. Activity, sleepiness and workload tracking form:

This form breaks each day into forty-eight 30-min blocks. Participants were instructed to specify for each time block, the major activity that was conducted, a rating on sleepiness, and a rating on workload. Ten activities were examined in this study, including seven on-duty work activities and three off-duty activities. A unique one-letter code was created for each activity, as shown in Table 3.

Table 3: *Activity codes and their definition in the daily log.*

On-duty Activity	Code
Watch	**W**
Evolution (e.g., actions stations, emergency stations, force protection, RAS, NBP)	**E**
Departmental	**D**
Maintenance	**M**
Training (e.g., lectures, learning activities)	**T**
Secondary Duties (e.g., ships committee, CSD custodian, harassment advisor)	**X**
Other (e.g., personal administration, meetings, cleaning stations)	**O**
Off-duty Activity	
Sleeping	**S**
Food/Meals	**F**
Personal Time (PT, movies, reading, etc.)	**P**

The sleepiness was rated based on a 7-point Stanford sleepiness scale [14], as shown in Table 4. This scale has been previously used in similar studies for fatigue research [7, 15, 16].

Table 4: *A 7-point Stanford sleepiness scale.*

1—Feeling active and vital; alert; wide awake
2—Functioning at a high level, but not at peak, able to concentrate
3—Relaxed; awake; not at full alertness, responsive
4—A little foggy; not at peak; let down
5—Fogginess; beginning to lose interest in remaining awake; slowed down
6—Sleepiness; prefer to be lying down; fighting sleep; woozy
7—Almost in reverie; sleep onset soon; lost struggle to remain awake

Task demand was measured in this study using a 7-point workload scale, as shown in Table 5. Originally developed by Pearson and Byars [17], it was introduced in the daily log to track perceived workload from all on-duty activities.

Table 5: *A 7-point workload scale.*

1—Nothing to do; no system demands
2—Demands little to do; minimum system demands
3—Active involvement required, but easy to keep up
4—Challenging, but manageable
5—Extremely busy, barely able to keep up
6—Too much to do; overloaded: postponing some tasks
7—Unmanageable; Potentially dangerous; Unacceptable

2. Sleep quality rating:

The quality of participants' daily sleep was measured subjectively in the daily log based on responses to four aspects of sleep experience:

- Whether falling asleep is difficult
- Whether the sleep is deep
- Whether arising at wake-up is difficult
- Whether one felt rested

All questions were rated on a 4-point scale with a lower score indicating a higher sleep quality. Participants were instructed to answer these questions right after waking up from their daily long sleep (i.e., primary anchor sleep). This sleep quality measure has been used in the past DRDC studies (e.g., [16]).

3. Sleep disruption and reason:

Participants were also asked to report disruptions to their sleep, including the duration of each disruption, the time it occurs, and a brief description of its cause.

4. Fatigue and mood survey:

The survey uses the following eight questions to assess one's daily fatigue and mood.

- How alert do you feel?
- How sad do you feel?
- How tense do you feel?
- How much of an effort is it to do anything?
- How happy do you feel?
- How weary do you feel?
- How calm do you feel?
- How sleepy do you feel?

Participants were asked to respond to each question on a 5-point scale. They were instructed to fill out this survey prior to their daily primary sleep, in order to reduce potential bias introduced by survey completion time. This survey has been used in past DRDC studies (e.g., [16]).

5. Caffeine and sleep medication intake:

The daily log also tracks participants' daily intake of caffeine and sleep medication if applicable. The recorded information includes the type, amount and intake time.

6. Open comment:

Lastly, an open-ended question is provided that invites all participants to comment on concerns about daily work and rest schedule and fatigue levels.

2.2.3 Post-trial survey

The post-trial survey is comprised of a single MFIS survey and it was filled out once on the last day of the study. Comparing MFIS responses on the first and the last day of the trial allowed us to assess the change of fatigue level during the period of this study.

2.2.4 Actigraph

A wrist actigraph is a watch-like accelerometer device that can discriminate a sleep state from a waking state. Such a device has been used in many studies (e.g., [16]). A subset of participants was identified to wear a wrist actigraph throughout the study. They were sampled from all ship departments and watch syndicates.

2.3 Study environment

The trial was conducted between Oct 25th and Nov 3rd, 2016. During this period of time, the crew of HMCS Montreal conducted a wide range of operational activities according to a mission plan modified based on a Phase Three work-ups scenario. The selection of specific exercise activities in this plan considered their potential impact on crewing. Those activities that may affect future design of crew size and composition were emphasized in this trial. In addition, the activities were scheduled on a ten-day timeline which was longer than the typical duration of a Phase Three Work-ups. For a detailed description of the scenario (or FLEX), readers are referred to [18].

During this study, the majority of the ship sailed in a 1-in-2 watch system (which breaks down watch keepers into two teams) that followed a 7-5-5-7 schedule. That is, the watch keepers were divided into two teams: starboard (STBD) and port. The 1-in-2 STBD team started work each day at 00:30 h for a 7-hour watch, then they were off watch for 5 hours before starting a second, 5-hour watch at 12:30 h. They had a 7-hour off-watch between 17:30 h and 00:30 h. The schedules for the 1-in-2 port team were complementary to those of the 1-in-2 STBD team. In total, each watch keeper was on watch for 12 hours every day, their detailed duty periods are shown in Table 6 and they remained the same throughout this trial.

Table 6: *1-in-2 watch rotation schedule adopted in this study.*

Watch	Duty periods
1-in-2 Stbd	00:30–07:30 and 12:30–17:30
1-in-2 Port	07:30–12:30 and 17:30–00:30

The only exception was the Bridge Watch Keepers (BWK) who stood a 1-in-3 watch rotation.[3] The duty periods of each watch team is described in Table 7. Consistent daily watch times were preserved for BWK during this study.

Table 7: *1-in-3 watch rotation schedule for bridge watch keepers.*

Watch	Duty periods
1-in-3 Red	03:30–07:30 and 15:30–19:30
1-in-3 Blue	07:30–11:30 and 19:30–23:30
1-in-3 White	11:30–15:30 and 23:30–03:30

The daily sea state during the trial was also recorded, as shown in Table 8. The highest sea state during this study was observed on the fourth and the fifth day. The information was used to interpret results obtained from other instruments, such as sleep quality and fatigue measures.

Table 8: *Daily sea state during the 10-day trial.*

Trial date	AM Swell	PM Swell
1	1–2 m	1–2 m
2	1–2 m	1–2 m
3	1 m	1–1.5 m
4	up to 2 m	3–4 m
5	3–4 m	2–3 m
6	2–3 m	2–2.5 m
7	2–2.5 m	2–2.5 m
8	1–1.5 m	1–1.5 m
9	1.0 m	1.0 m
10	< 1.0 m	< 1.0 m

2.4 Procedures

On the first day of the trial, all members of the crew were given a copy of the information sheet and a consent form for this study. Only those who provided written, informed consent were regarded as participants.

[3] This 1-in-3 watch rotation was based on the established crew count for the role. However in many operations, bridge watch keepers usually have additional personnel in their rotation and do not sail in this exact watch system.

All participants were asked to complete a pre-trial and a post-trial survey, and maintain a daily log throughout the study. Surveys were distributed to each participant on the first day and were collected by the experimenters on a daily basis.

Eighty-five participants were required to wear a wrist actigraph, which provided an alternative means to track their sleep patterns, in addition to self-report logs. The device was distributed on the first day and returned to the experimenters on the last day of the trial.

The data collection was performed by five DRDC and D Nav P&T experimenters, with the support of a team of domain SMEs from Naval Sea Training.

2.5 Statistical analysis methods

Both parametric and non-parametric statistical methods were used to analyze the experimental data. Unless otherwise specified, data are reported as mean (M) ± standard deviation (SD).

Parametric methods were used for ratio data collected in this study, primarily in the form of duration for various daily activities. In all cases, a mixed design Analysis of Variance (ANOVA) was applied with a within-subjects factor of trial date and a between-subjects factor of either department, watch or rank.

Most surveys in this study use an ordinal rating scale, including MEQ, MFIS, WRS, ratings of sleepiness, workload, sleep quality, fatigue and mood, non-parametric methods were adopted to examine the statistical significance. More specifically, when the analysis involved a between-subjects measure, such as comparison across ship departments, watch groupings or rank levels, the Kruskal-Wallis one-way ANOVA and the Mann-Whitney U test were adopted. Typically in this study, a Kruskal-Wallis test was performed first since the conditions for the factor of department, watch or rank were more than two levels. After a significant effect was revealed by the Kruskal-Wallis test, a series of Mann-Whitney tests was then carried out as a post-hoc analysis to pair-wise compare the factor of interest. On the other hand, when the analysis was focused on a within-subjects measure, such as comparison of a measure across multiple trial dates (i.e., a repeated measure), a Friedman's test was first conducted, followed by a series of Wilcoxon signed ranks test for post-hoc pair-wise comparison.

All statistical analyses were performed using Statistical Package for the Social Sciences Version 23 (IBM Corp., New York, USA) with a significant value of $p < 0.05$.

3 Results

The section presents key results obtained from this study. It is structured with three major subsections that explain analysis results in relation to sailors' work, rest, and fatigue in this ten-day trial. For each measure that was investigated, a global overview of results from all participants is first provided, followed by a comparison, where applicable, across three participant grouping factors: department, watch, and rank.

3.1 Work Requirement Survey (WRS)

Participants were asked in this survey to rate the importance of 52 basic human skills in relation to their work requirements. These skills are classified into four broad categories: sensory, psychomotor, physical and cognitive skills. A 5-point rating scale was used, ranging from 1 (not important) to 5 (extremely important). The response anchors have been suggested for similar surveys using Likert-type scales [19]. A total of 151 valid survey responses were collected.

Figure 2 shows the results of skill ratings averaged across all participants. Overall, a wide range of skills were required for work on a naval ship. A total of 21 skills received a median rating that was greater than 3 (i.e., the midpoint of the scale which indicates an important requirement), including 13 cognitive skills, 7 sensory skills and 1 psychomotor skill. Oral comprehension in particular was the only skill that had a median rating of 5. On the other end of the spectrum, only two skills had a median rating below 3, one was a physical skill (explosive strength) and the other a psychomotor skill (Arm hand steadiness).

Table 9 displays results from comparison of four skill categories. A Kruskal-Wallis test was performed and the result revealed a significant main effect of skill category, $\chi^2(3) = 445.13$, $p < .001$. The follow-up pair-wise comparison using the Mann-Whitney U test indicated a significant difference in rating between each pair of categories. The average rating for a cognitive skill was the highest, followed by sensory, psychomotor, and physical skills in that order.

Table 9: *WRS rating comparison for four skill categories.*

Skill category	Number of skills in the category	Median	Mode	25th Percentile	75th Percentile
Cognitive	21	4	4	3	4
Sensory	12	3	4	3	4
Psychomotor	10	3	4	2	4
Physical	9	3	3	2	4

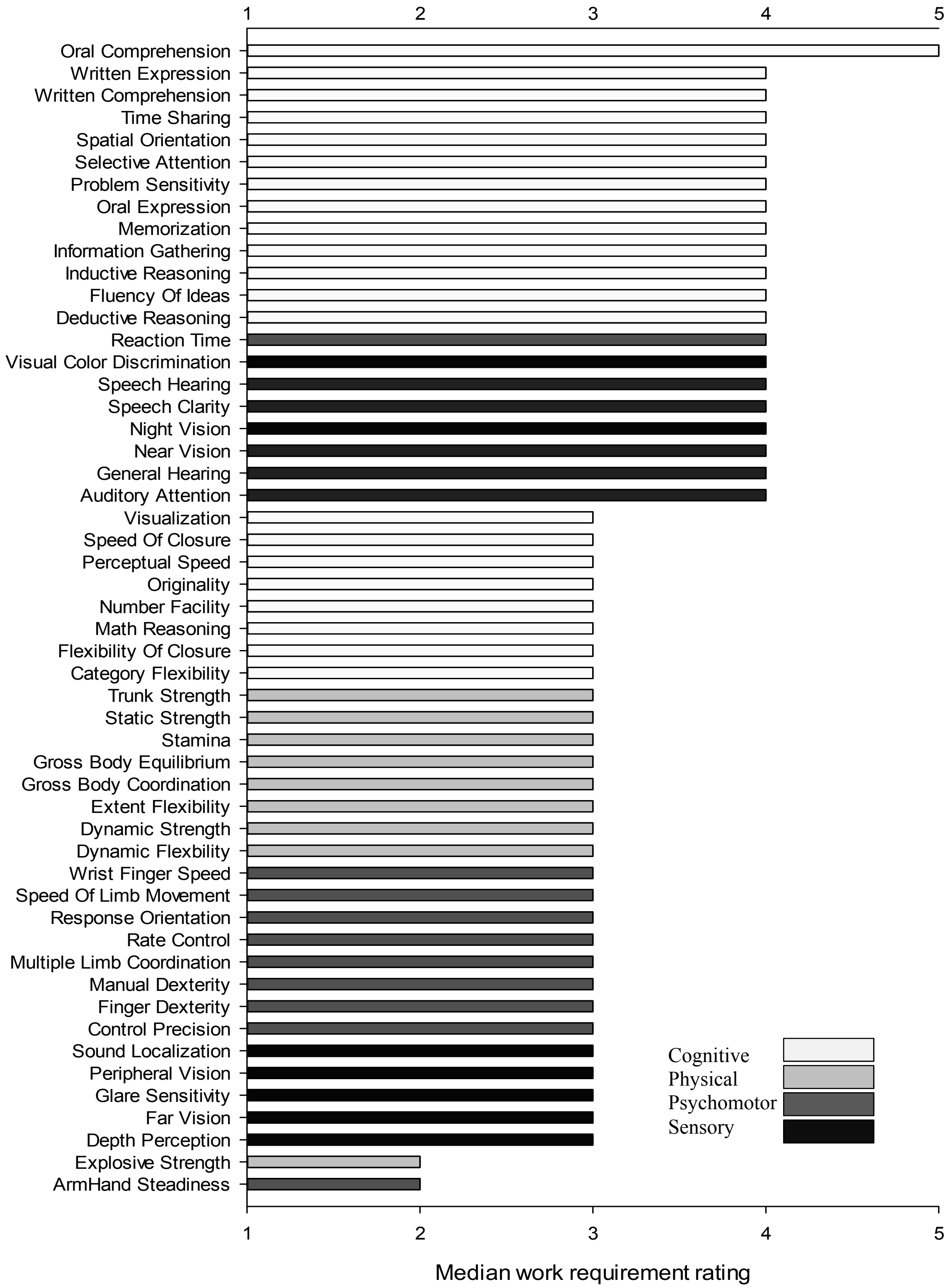


Figure 2: *Median ratings of importance for 52 skills in relation to sailors' work.*

Next, an in-depth analysis was conducted to compare skill ratings across ship departments. The results are summarized in Table 10. Notably, an asterisk in the last column indicates a statistically significant difference in the importance of the skill across the departments. Overall the ratings for a total of 33 skills were statistically different across departments, reflecting their differences in duties and responsibilities.

The skills related to communication (including both written and oral forms) and problem solving were highly rated by all departments. The following is a list of insights obtained from this analysis with an emphasis on additional critical skills required by each department.

1. For the CBT department, besides communication and problem solving, sensory and cognitive abilities such as visualization, memorization, selective attention, spatial orientation, timesharing were also highly rated.

2. The general skill requirements by two engineering departments (i.e., CSE and MSE) were similar to those of CBT, except that psychomotor skills such as control precision and manual dexterity were also rated as very important by the engineers, indicating an enhanced requirement for control movement and fine manipulative abilities.

3. Results from EXE included 14 skills with a median rating of 5 (i.e., extremely important), more than any other departments and generally reflecting a high level of skill requirements for this small group of crew members. Besides abilities associated with communication and problem solving, time sharing, memory, selective attention and quantitative ability were also highly rated by EXE.

4. The DECK and the LOG department emphasized more on sensory and physical skills. All visual abilities were rated by the DECK department as very important, including far vision, near vision, night vision, glare sensitivity, depth perception, peripheral vision, visual color discrimination. Six physical skills that are associated with body strength, stamina, and flexibilities were highly rated by both departments For LOG, quantitative ability was also highlighted as very important.

Table 10: *A breakdown of skill ratings by ship department.*

Category	Skill	CBT	CSE	DECK	EXE	LOG	MSE	Statistical significance
		Median	Median	Median	Median	Median	Median	
Cognitive	Category Flexibility	3	3	2	4	3	3	
	Deductive Reasoning	4	4	3	5	4	4	*
	Flexibility Of Closure	3	2	3	3	2	3	
	Fluency Of Ideas	3	4	4	5	3	4	*
	Inductive Reasoning	4	4	3	5	3	4	
	Information Gathering	4	4	3	5	4	4	
	Math Reasoning	3	3	2	3	3	3	
	Memorization	4	4	4	5	4	4	
	Number Facility	3	2	2	4	4	3	
	Oral Comprehension	5	4	4	5	4	4	
	Oral Expression	5	4	5	5	4	4	

Category	Skill	CBT	CSE	DECK	EXE	LOG	MSE	Statistical significance
		Median	Median	Median	Median	Median	Median	
Cognitive	Originality	3	3	3	5	3	4	*
	Perceptual Speed	3	3	3	3	3	3	*
	Problem Sensitivity	4	4	4	5	4	5	*
	Selective Attention	4	3	3	4	4	4	
	Spatial Orientation	4	3	4	3	3	4	
	Speed Of Closure	4	3	3	4	3	3	*
	Time Sharing	4	4	3	5	3	4	*
	Visualization	4	3	4	3	3	3	
	Written Comprehension	4	4	3	5	4	4	*
	Written Expression	4	4	4	5	4	4	*
Physical	Dynamic Flexbility	2	3	3	2	4	3	*
	Dynamic Strength	2	3	4	1	4	3	*
	Explosive Strength	2	3	3	2	3	3	*
	Extent Flexibility	2	3	3	2	4	3	*
	Gross Body Coordination	2	3	4	2	4	3	*
	Gross Body Equilibrium	3	3	4	3	4	4	*
	Stamina	2	3	4	3	4	3	*
	Static Strength	2	3	4	3	4	3	*
	Trunk Strength	2	3	4	2	4	3	*
Psychomotor	ArmHand Steadiness	2	2	3	1	3	3	*
	Control Precision	3	4	3	2	3	3	
	Finger Dexterity	3	3	3	3	4	4	*
	Manual Dexterity	2	4	4	3	4	4	*
	Multiple Limb Coordination	3	3	4	2	4	4	
	Rate Control	3	3	4	2	3	3	
	Reaction Time	4	3	4	3	4	3	
	Response Orientation	3	3	4	2	3	3	
	Speed Of Limb Movement	2	3	3	2	3	3	*
	Wrist Finger Speed	3	3	3	2	4	3	
Sensory	Auditory Attention	4	4	3	5	2	4	*
	Depth Perception	3	3	4	4	3	3	*
	Far Vision	3	3	4	4	3	3	*
	General Hearing	4	4	4	4	3	4	*
	Glare Sensitivity	3	3	4	3	2	3	*
	Near Vision	3	3	4	4	4	3	*

Category	Skill	CBT	CSE	DECK	EXE	LOG	MSE	Statistical significance
		Median	Median	Median	Median	Median	Median	
Sensory	Night Vision	4	3	4	3	3	3	*
	Peripheral Vision	3	2	4	3	3	3	*
	Sound Localization	3	3	3	3	2	3	
	Speech Clarity	4	4	4	5	4	4	*
	Speech Hearing	4	4	3	4	3	3	*
	Visual Color Discrimination	4	4	4	4	3	3	*

Summary: results from WRS revealed that a wide range of skills were required by naval work, and the majority of highly rated ones were cognitive in nature. In particular, oral and written communication skills were regarded critical by all departments, followed by problem solving (in CSE, MSE, CBT, EXE), and time sharing (in CBT and EXE). The Deck department deviated from this general pattern with an emphasis on sensory abilities particularly in the visual domain. In comparison, physical and psychomotor skills were rated lower. There are 19 skills in these two categories, none of them received an average rating that was greater than 4 (very important).

3.2 Daily work duration

Results presented in this subsection were analyzed based on 1258 daily activity logs that were collected from 159 participants over 8 days. Logs that were filled out on the first and the last day were removed from analysis due to the large amount of missing data they contained.

A participant's daily work duration was computed by summing up the time an individual spent on seven types of on-duty activities defined in the activity log, including watch, departmental, evolution, maintenance, training, secondary duty, and a catch-all category for all other work related activities.

On average, a participant worked 13.3 hours per day in this trial. The daily work time was affected by the mission scenario and fluctuated across trial date, as shown in Figure 3. ANOVA indicated a main effect of trial date, $F(7, 1250) = 3.858$, $p < .001$, $\eta_p^2 = .021$. Post-hoc analysis using Tukey's HSD test revealed that the average daily work time was longer on Day 9 than four other days (i.e., 3, 4, 6, 7) in this study.

Figure 4 shows a breakdown of daily work duration into seven types of on-duty activities, and how work composition varied for three participant grouping factors investigated in this study, that is, department, watch and rank.

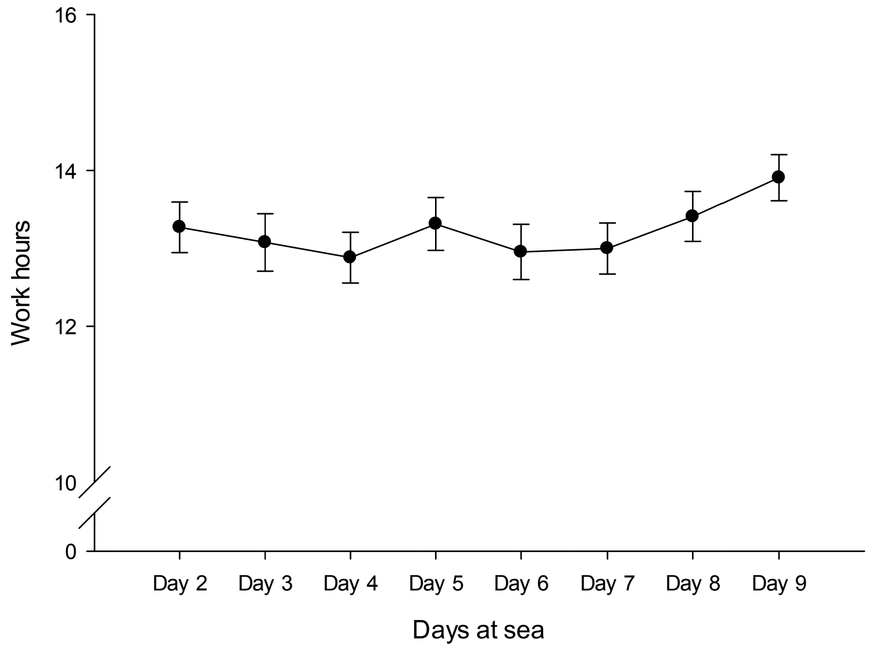


Figure 3: *Average daily work duration for eight days at sea.*

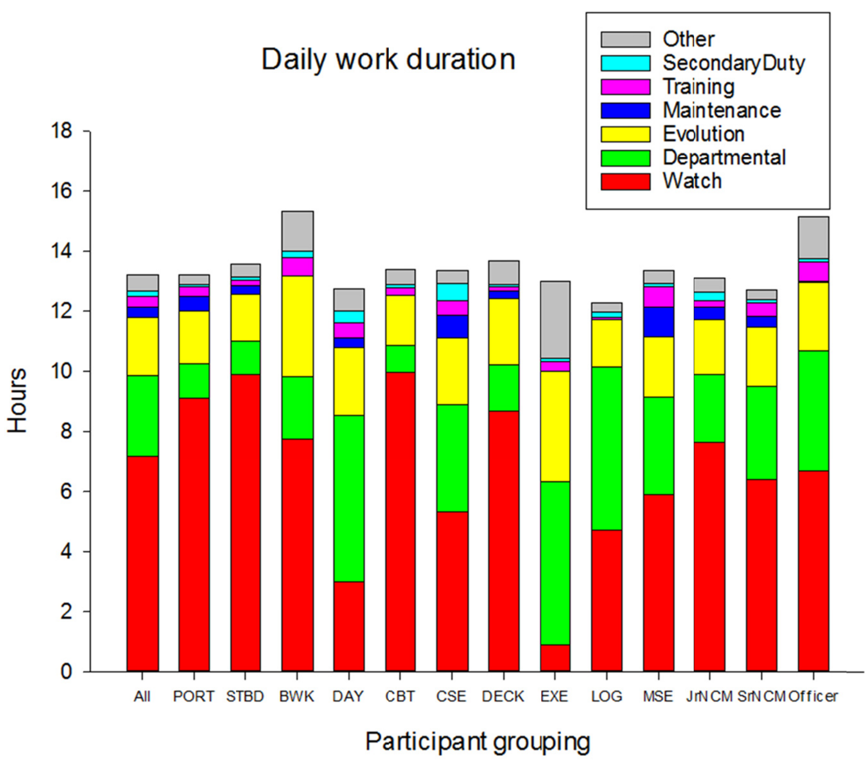


Figure 4: *Average daily work duration for participant groupings.*

The result summarized from all participants is shown by the first bar in Figure 4. It indicated a 13.3 h average daily work time, which was comprised of three major elements: watch keeping (7.2 h/day), departmental work (2.7 h/day) and evolution (2.0 h/day). A smaller portion of daily work was spent on maintenance (0.4 h/day), training (0.4 h/day) and secondary duties (0.2 h/day), with an additional 0.6 h/day logged as other on-duty activities.

Notably, 10.6% of all observations (i.e., with a total number count of 133) revealed a daily work time that was longer than 16 h. A breakdown of these long work days by participant groupings is shown in Table 11. Notably is the result for BWK, close to half of the time in this study (i.e., 3.6 out of 8 days), each member of this small group of participants worked more than 16 hours per day.

Table 11: *Distribution of 16+ hour work day by participant groupings.*

Participant grouping	N	Long work day (>16 h) count	Long work day per participant
ALL	159	133	0.8
PORT (1-in-2 Port Watchkeepers)	49	22	0.4
STBD (1-in-2 Starboard Watchkeepers)	49	34	0.7
BWK (1-in-3 Bridge Watchkeepers)	5	18	3.6
DAY (Day workers or on-call)	56	59	1.1
CBT (Combat department)	59	42	0.7
CSE (Combat Systems Engineering department)	27	15	0.6
DECK (Deck department)	10	11	1.1
EXE (Executive department)	5	7	1.4
LOG (Logistics department)	23	14	0.6
MSE (Marine Systems Engineering department)	35	44	1.3
Jr Sailor (OS to MS)	95	52	0.5
Officer (SLt to Cdr)	18	20	1.1
Sr Sailor (PO2 to CPO1)	46	28	0.6

Detailed composition of daily work is of interest to SCORE modelling and will be analyzed and reported elsewhere. For the purpose of this study, we focused the investigation on whether there existed a significant difference in total work duration across department, watch team, or rank level.

Departmental comparison

A mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of department was performed. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 76.165$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .868$). The results revealed main effects of date, $F(6.076, 868.885) = 5.260$, $p < .001$, $\eta_p^2 = .035$, and department, $F(5, 143) = 2.313$, $p = .047$, $\eta_p^2 = .075$, that were qualified by an interaction between date and department, $F(30.38, 868.885) = 13.196$, $p = .040$, $\eta_p^2 = .005$.

To study the significant interaction, a simple effects analysis was conducted. The results showed a main effect of department for Day 8, F(5, 151) = 3.060, p = .012, $\eta_p^2 = .092$. A post-hoc Tukey's test revealed that the daily work duration for LOG (M=12.1, SD=2.0) was significant shorter than CBT (M=13.8, SD=1.6) and DECK (M=14.5, SD=2.6). For all other trial dates, there was not a significant difference in work duration across departments.

Watch comparison

Based on a similar analysis method, a mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of watch was performed. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 75.646$, $p < .001$), and degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .876$). Main effects of date, F(6.134, 889.390) = 3.234, p = .004, $\eta_p^2 = .022$, and watch, F(5, 145) = 5.616, p = .001, $\eta_p^2 = .104$, were qualified by an interaction between date and watch, F(18.40, 889.390) = 4.630, p < .001,$\eta_p^2 = .087$.

A simple effects analysis was carried out to examine the significant interaction and the results are summarized as following:

- BWK (M=15.3, SD=1.7) consistently reported the longest daily work time among four watch groupings. For two days in the trial (Days 4 and 9), BWK's work duration was significantly longer than the three other groupings. On Day 8, BWK's work duration was longer than STBD and Dayworkers.
- For three days in this study (Days 2, 5, 6), the work duration of STBD was significantly longer than that of PORT and Dayworkers.

Rank comparison

A mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of rank was performed. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 67.054$, $p < .001$), consequently degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .885$). The results showed a main effect of date, F(6.196, 904.574) = 3.980, p = .001, $\eta_p^2 = .027$, and rank, F(2, 146) = 17.170, p < .001, $\eta_p^2 = .190$. A post-hoc Tukey's HSD test revealed that Officers (M=15.2, SD=1.8) worked longer than either Jr Sailors (M=13.1, SD=2.0) or Sr Sailors (M=12.7, SD=2.3).

3.3 Workload rating

Workload ratings from the modified Pearson and Byars 7-point workload scale were obtained from daily activity logs. However, only the data from days 2 to 9 were used, since too many participants failed to complete the logs on days 1 and 10.

Overall, the average workload rating from all participants was 2.7, which was between 2 ("Demands little to do; minimum system demands") and 3 ("Active involvement required, but easy to keep up"). As shown in Figure 5, daily mean workload was consistently rated between 2 and 3, a Kruskal-Wallis test indicated a main effect of date, $\chi^2(7) = 16.351$, p = .022. Post-hoc pair-wise comparison using Mann-Whitney U test shows a significantly elevated workload on Day 2, relative to days 4–8. Figure 6 shows average workload ratings for all participant groupings.

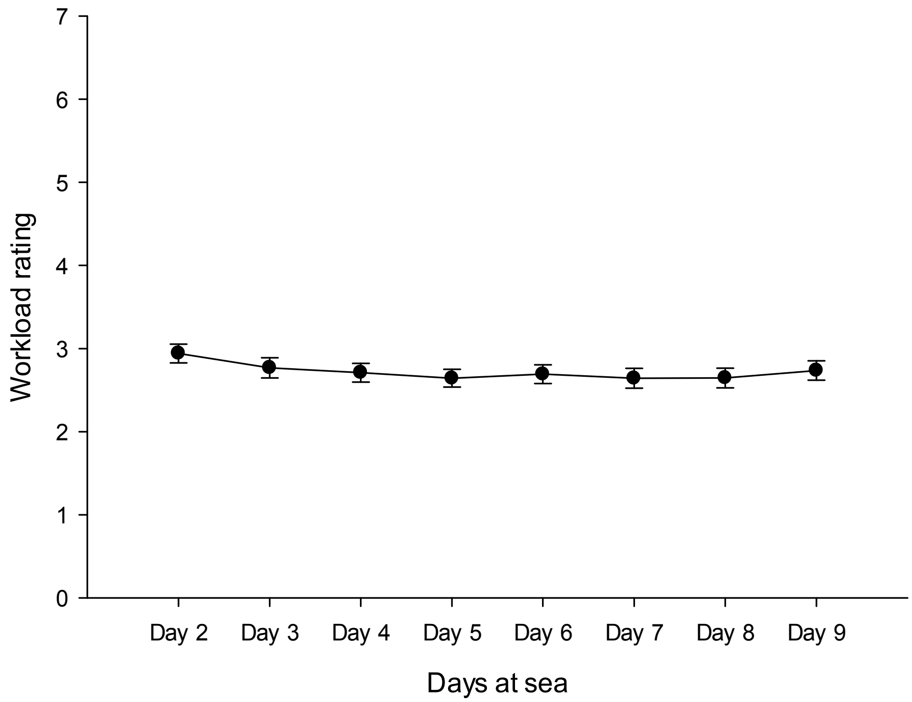


Figure 5: *Average daily workload ratings for eight days at sea.*

Departmental comparison

A Kruskal-Wallis test was conducted to investigate departmental difference in workload rating. Each trial date was analyzed separately and the results indicated that a main effect of department was identified only for two days in this trial, Day 6 (χ^2 (5) = 12.501, p = .029) and Day 7 (χ^2(5) = 15.770, p = .008).

Post-hoc pair-wise comparison using Mann-Whitney test revealed the following significant difference in workload ratings:

- For both days (i.e., Days 6 and 7), the rating by MSE was higher than those of CBT, CSE and DECK.
- For Day 7, the ratings by EXE and LOG were higher than those of CBT, CSE and DECK.

Watch comparison

A similar analysis method was followed to compare watch teams. The results indicated a main effect of watch for Day 7 only (χ^2(3) = 9.603, p = .022). A follow-up pair-wise comparison using Mann-Whitney test showed a higher level of workload was reported by Dayworkers than 1-in-2 watch keepers on Day 7.

Rank comparison

A Kruskal-Wallis test indicated a main effect of rank for four different trial dates: Day 3 ($\chi^2(2) = 7.793$, $p = .020$), Day 7 (χ^2 (2) = 6.642, $p = .036$), Day 8 (χ^2 (2) = 7.973, $p = .019$) and Day 9 ($\chi^2(2) = 9.149$, $p = .010$). The post-hoc pair-wise comparison confirmed that Jr Sailors reported a lower level of workload than either Officers (on all four days) or Sr Sailors (on three days).

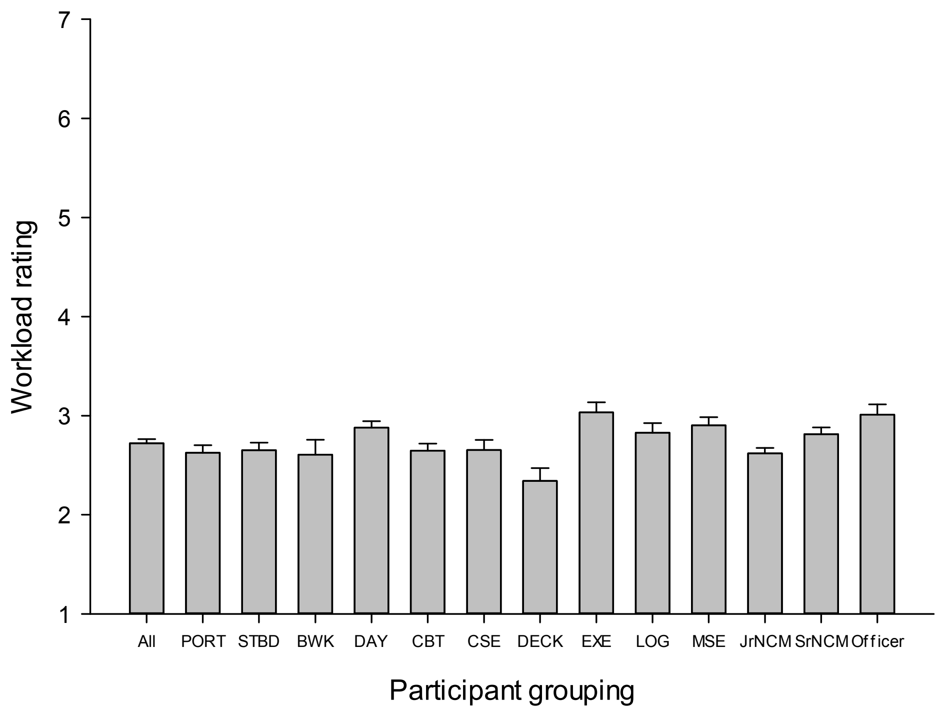


Figure 6: *Average workload ratings for participant groupings.*

3.4 Daily rest duration

A participant's daily rest duration was computed by summing up the time that an individual spent on three off-duty activities, that is, sleep, meal, and personal time. Together with daily work duration, these two measures added up to the 24 h day. Figure 7 shows the results from this measure for different participant groupings.

Overall, a participant's average daily rest time was 10.7 h in this study, which was distributed between sleep (7.4 h), personal time (2.0 h) and meals (1.4 h). Detailed analyses of this measure were not reported as the results were closely related to the daily work duration measure. In this report, we focus on sleep data that is the dominant factor in crew rest.

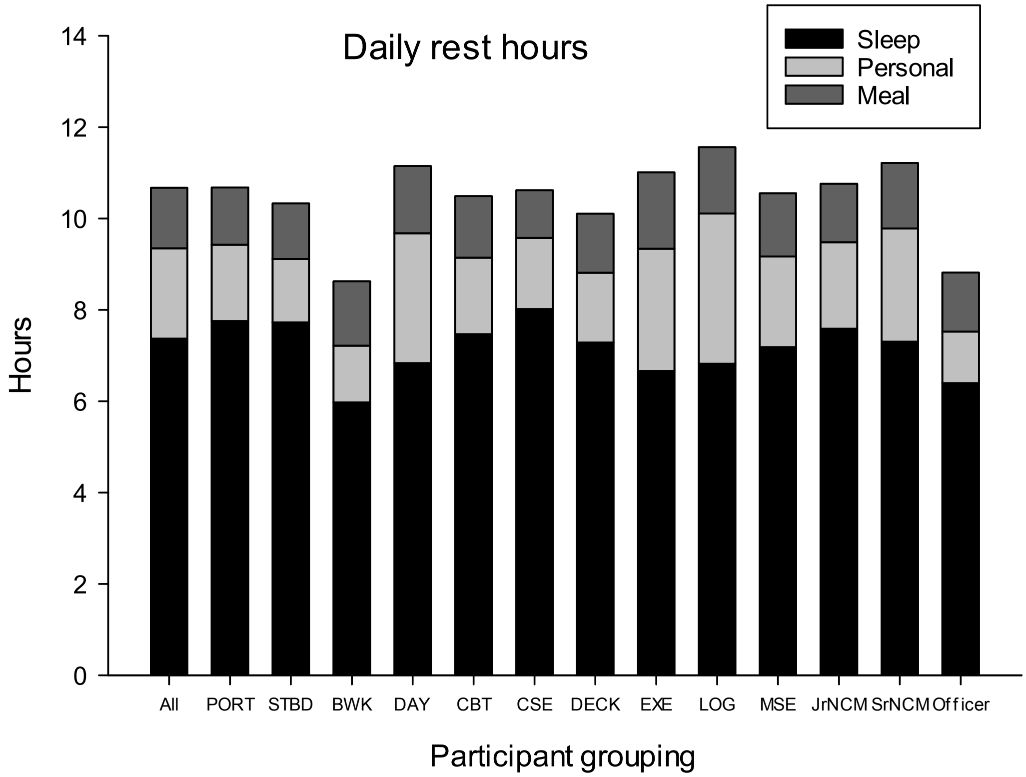


Figure 7: *Daily rest hours for participant groupings.*

3.5 Baseline sleep

Participants' sleep data for three days prior to the start of the trial was collected using a baseline sleep survey. A total of 145 valid responses were collected. Since the measure was based on self-report survey, it should be interpreted as Time-In-Bed (TIB). Mean TIB durations and their standard deviations are shown for various participant groupings in Figure 8.

On average, each participant had 8.4 h daily TIB during the three days before the start of the trial. ANOVA did not show a main effect in either department, $F(5, 139) = 1.828$, $p = .111$, or watch, $F(3, 141) = 0.660$, $p = .578$. But a significant main effect of rank was identified, $F(2, 142) = 3.105$, $p = .048$, and a post-hoc Tukey's test revealed that the Jr Sailors (M=8.6, SD=1.6) had a longer TIB than Sr Sailors (M=7.9, SD=1.1).

It is useful to point out that Day 1 of this trial coincided with the first day of the sail, while the baseline sleep reflected participants' sleep on shore, before they boarded the ship.

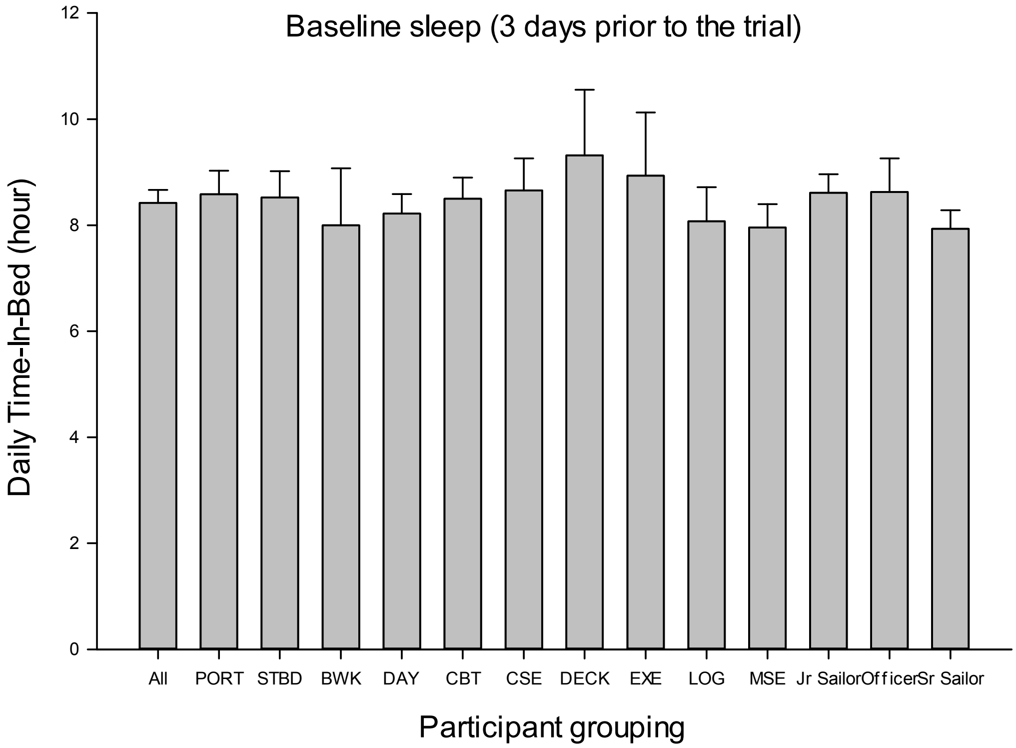


Figure 8: *Daily Time-In-Bed duration for three days prior to the start of the trial.*

3.6 Daily sleep duration

In this study, participants' daily sleep duration was measured in two ways: by using either self-report daily logs or a wrist actigraph. Results from these two measures are described separately in two different subsections. We suggest that self-report sleep durations be interpreted as TIB and actigraphically measured durations be a more precise indication of sleep time.

3.6.1 Self-report sleep time (i.e., Time-In-Bed)

Daily TIB was defined as the total amount of sleep a participant self-reported in a calendar day. Results of TIB were analyzed based on daily logs obtained from 160 participants. Logs from the first and the last day of the trial were removed from the analysis due to incomplete data, consequently results from eight trial dates are presented in this report, from Day 2 to Day 9.

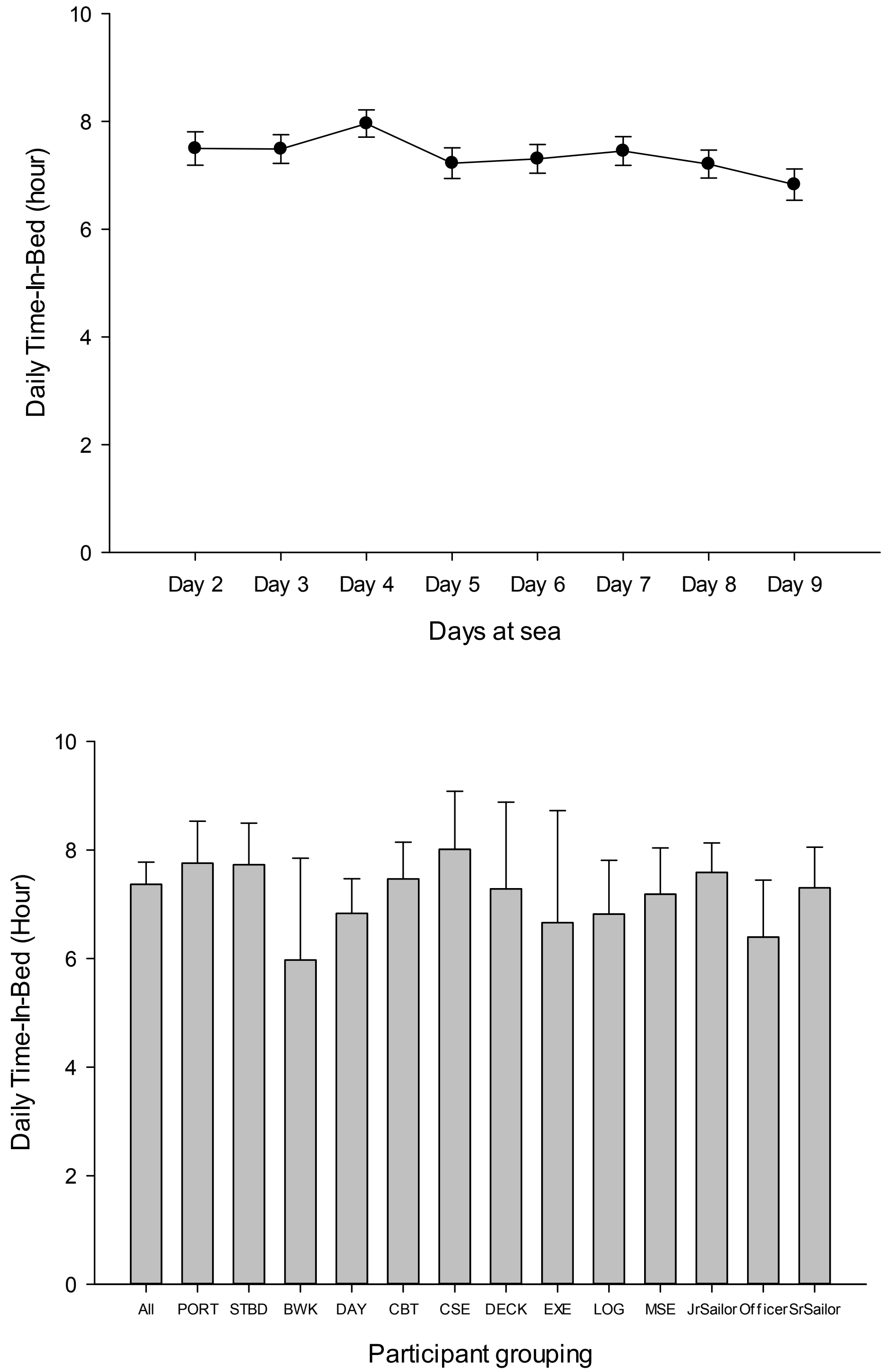


Figure 9: *Average daily Time-In-Bed (i.e., sleep time recorded by self-report logs) over eight trial dates (top) and comparison among watch, department, and rank groupings (bottom).*

Overall on average, a participant's daily TIB was 7.4 h in this trial. The TIB duration fluctuated significantly across trial date, as shown in Figure 9 and confirmed by ANOVA which indicated a main effect of date, $F(7, 1250) = 5.307$, $p < .001$, $\eta_p^2 = .029$. Post hoc pair-wise comparison using Tukey's HSD test confirmed that the average TIB on Day 4 (M=8.0, SD=1.6) was the longest during this study and it was significantly longer than four other days (Days 5, 6, 8, 9). On the other hand, the shortest daily TIB was observed on Day 9 (M=6.8, SD=1.9) which was significantly shorter than four other days (Days 2, 3, 4, 7).

Departmental comparison

A mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of department was conducted. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 77.871$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .858$). Main effects of date, $F(6.008, 859.131) = 5.332$, $p < .001$, $\eta_p^2 = .036$, and department, $F(5, 143) = 3.046$, $p = .012$, $\eta_p^2 = .096$, were qualified by an interaction between date and department, $F(30.040, 859.131) = 1.478$, $p=.048$, $\eta_p^2 = .049$.

A simple effects analysis was conducted to examine the significant interaction. The results confirmed that a significant difference in TIB duration among the following departments:

- On three days in this study (Days 2, 6, 9), the TIB duration of CSE was longer than that of LOG.
- On one day (Day 2), the TIB duration of CSE was also longer than that of DECK.
- On the last day (Day 9), two other departments (CBT and DECK) had a longer TIB than LOG.

Watch comparison

Data were analyzed using a mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of watch. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 70.178$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .873$). Main effects of date, $F(6.113, 886.318) = 3.315$, $p = .003$, $\eta_p^2 = .022$, and watch, $F(3, 145) = 9.224$, $p < .001$, $\eta_p^2 = .160$, were qualified by an interaction between date and watch, $F(18.338, 886.318) = 3.184$, $p < .001$, $\eta_p^2 = .062$.

A simple effects analysis was conducted to examine the significant interaction. The results confirmed that a significant difference in sleep duration among the following watch groups:

- The daily TIB duration of the Dayworkers was shorter than those of the 1-in-2 watch keepers. In this study, the Dayworker's TIB was significantly shorter than STBD on 6 days, PORT on 5 days.
- BWK's TIB was less than 1-in-2 watch keepers. The difference was significant for 4 days compared to STBD, 2 days compared to PORT.
- On one day, the TIB of STBD was shorter than that of PORT.

Rank comparison

Data were analyzed using a mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of rank. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 79.957$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .859$). The result showed a main effect of date, $F(6.016, 878.301) = 4.499$, $p < .001$, $\eta_p^2 = .030$, and rank, $F(2, 146) = 8.951$, $p < .001$, $\eta_p^2 = .109$. Officers' TIB (M=6.4, SD=1.4) was shorter than either Jr Sailors (M=7.6, SD=1.8) or Sr Sailors (M=7.3, SD=1.8).

3.6.2 Actigraphically measured sleep time

Eighty-five participants were provided a wrist actigraph in this study. The recordings from 14 participants were not analyzed due to missing data, mostly because the actigraph was not consistently worn during the trial. The results presented in this subsection were based on data from the remaining 71 participants.

On average, a participant's actigraphically measured sleep time was 6.3 h/day. The sleep duration fluctuated significantly across trial date, as shown in Figure 10 and confirmed by ANOVA which indicated a main effect of date, $F(7, 527) = 7.479$, $p < .001$, $\eta_p^2 = .009$. Post hoc pair-wise comparison using Tukey's HSD test confirmed that the average sleep time on Day 2 (M=7.3, SD=2.2) was the longest during this study and it was significantly longer than three other days (Days 5, 8, 9). On the other hand, the shortest daily sleep was observed on Day 9 (M=4.9, SD=2.2) which was significantly shorter than five other days (Days 2, 3, 4, 6, 7).

Departmental comparison

A mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of department was conducted. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 46.034$, $p = .013$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .762$). A main effect of date, $F(5.337, 229.471) = 7.099$, $p < .001$, $\eta_p^2 = .142$, was identified, together with a significant interaction between date and department, $F(26.683, 229.471) = 1.559$, $p = .045$, $\eta_p^2 = .153$. However, the between-subjects effect of department was not significantly different, $F(5, 43) = 1.289$, $p = .286$, $\eta_p^2 = .130$.

Watch comparison

To compare watch groups, a mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of watch was conducted. Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2(27) = 33.796$, $p = .174$). A main effect of date, $F(7, 315) = 10.294$, $p < .001$, $\eta_p^2 = .186$, was identified, together with a significant interaction between date and watch, $F(21, 315) = 3.705$, $p < .001$, $\eta_p^2 = .198$. However, the between-subjects effect of watch was not significantly different, $F(3, 45) = .561$, $p = .643$, $\eta_p^2 = .036$.

Rank comparison

A mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of rank was conducted. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(27) = 43.888$, $p = .022$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .779$). A main effect of date, $F(5.453, 250.847) = 8.226$, $p < .001$, $\eta_p^2 = .152$, was identified, together with a significant interaction between date and department, $F(10.906, 250.847) = 2.452$, $p = .006$, $\eta_p^2 = .0.96$. However, the between-subjects effect of rank was not significantly different, $F(2, 46) = 3.147$, $p = .052$, $\eta_p^2 = .120$.

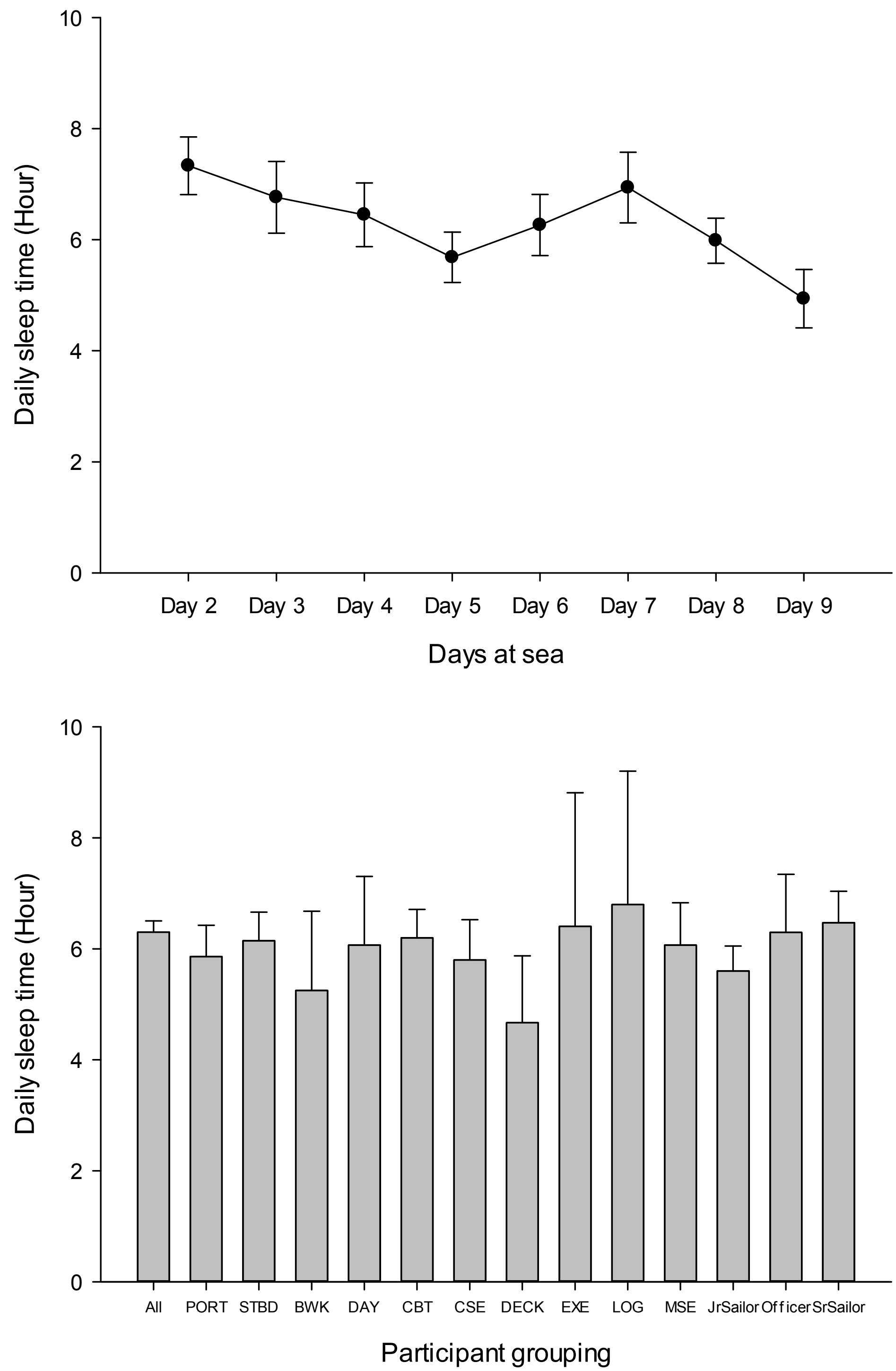


Figure 10: *Average daily sleep duration (i.e., measured by actigraph) over eight trial dates (top) and comparison among watch, department, and rank groupings (bottom).*

3.6.3 TIB versus actigraphically measured sleep

A further analysis was performed to examine two sleep measures by comparing the actigraphically measured sleep durations of 71 participants to their own self-report TIB durations using a paired-samples t-test. The results indicated that self-report TIB duration was significantly longer (M=7.5, SD=2.6) than actigraphically measured sleep duration (M=6.3, SD=2.4), $t(534) = 11.877$, $p < .001$, $d = 0.02$. The average difference between these two measures was 1.2 h per day.

A difference between these two measures is expected. TIB reflects the total length of time that a participant stays in bed to sleep, the duration includes periods of wakefulness both before and after sleep onset. In the parlance of sleep research, the length of time that a participant takes to accomplish the transition from full wakefulness to sleep is commonly referred to as Sleep Onset Latency (SOL), and the length of time a participant spends awake between sleep onset and fully wake-up is called Wake-Time After Sleep Onset (WASO). While both SOL and WASO were included in TIB, these periods of wakefulness were not considered in the actigraphically measured sleep duration.

In the current study, these two measures were both useful and they served different modelling purposes. The measure TIB is appropriate in SCORE modelling for analyzing the crew's work and rest schedules, where the design concern is to provide proper windows of opportunities for a crew member to sleep. On the other hand, for DFM's performance prediction, the actigraphically measured sleep patterns were used as model input since the data represents the amount of sleep that the crew obtains more realistically.

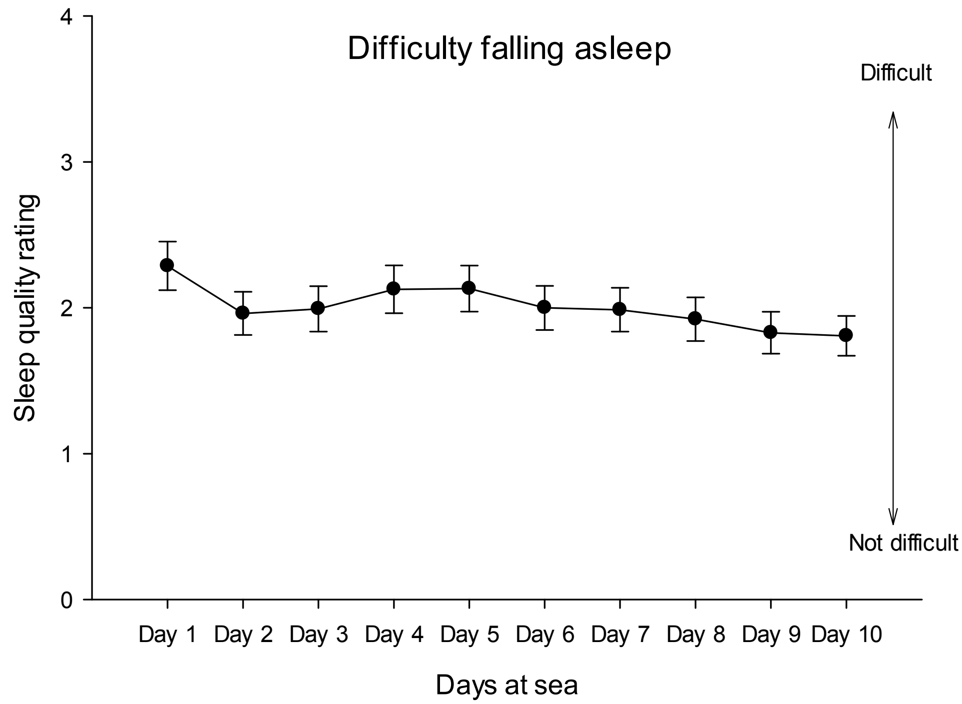
3.7 Sleep quality rating

Quality of sleep was measured subjectively using a daily survey that asked a participant to rate the difficulty of falling asleep (Question 1) and arising (Question 3), whether the sleep was deep (Question 2) and whether the participant felt rested (Question 4). All questions were rated on a 4-point scale with a lower score indicating a higher sleep quality.

Responses from 159 participants over the entire 10-day trial were collected. There was a small amount of missing data, with the worst case for Question 4 on Day 10 where a total of 136 valid responses were available for analysis. Results for each question are presented separately in the subsequent subsections.

3.7.1 "Difficulty falling asleep"

Figure 11 summarizes the results for the first sleep quality question. On average, participants provided a rating of 2.0 which was on the lower (therefore better) end of the 4-point scale, indicating that falling asleep in general was not too difficult.



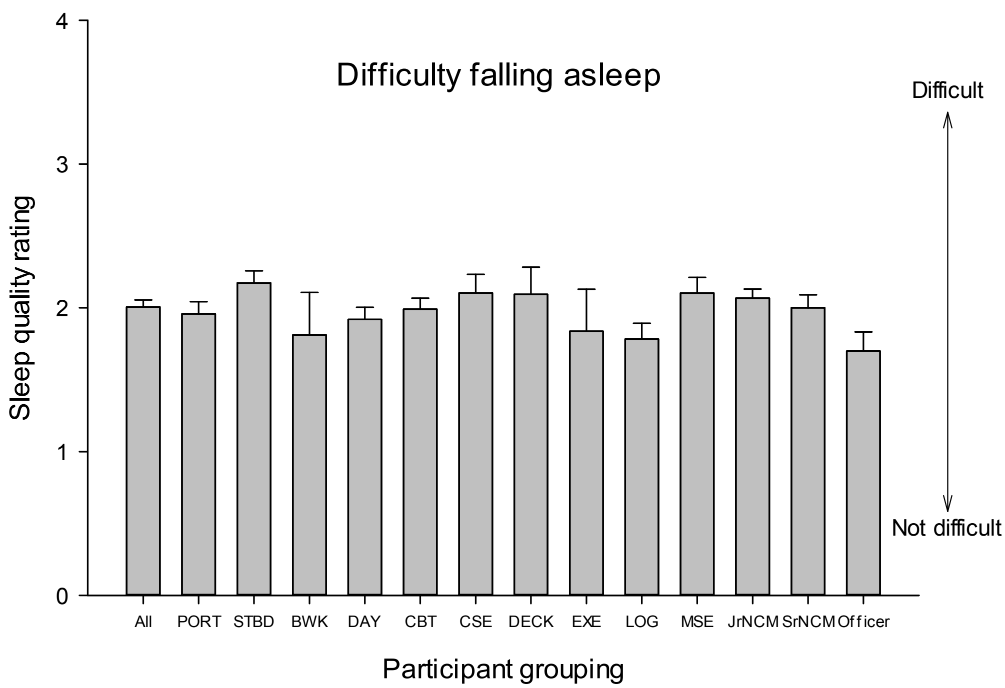


Figure 11: *Results for question 1 of the sleep quality survey "Difficulty falling asleep": ratings comparison across trial date (top) and participant groupings (bottom).*

The difficulty level however changed from one day to another. A Friedman's one-way ANOVA confirmed a main effect of date, χ^2 (9) = 34.747, p < .001. Post-hoc pair-wise comparison using the Wilcoxon Signed Ranks test reveals significant differences in rating for the following dates,[4] where a higher rating indicates more difficulty falling asleep.

- Day 1 > Days 2, 3, 6–10
- Day 3 > Day 10
- Days 4, 5 > Days 8–10
- Day 6 > Day 10
- Day 7 > Day 9, 10

Kruskal-Wallis tests were performed to analyze differences across participant groupings based on either department, watch or rank. The results indicated that there was not a main effect of department for any of the trial dates. A main effect of watch was detected for three days (i.e., Days 1, 3, and 7), and a main effect of rank for one day (i.e., Day 4).

Post-hoc pair-wise comparison using the Mann-Whitney U test confirmed that a significantly higher rating (i.e., more difficulty) was reported by STBD than the other three watch teams. More specifically STBD reported an elevated difficulty falling asleep than (1) Dayworkers on three days (Days 1, 3, 7), (2) PORT on two days (Days 1 and 3), and (3) BWK on one day (Day 7). For one day in this trial (Day 4), the average rating by Officers was lower (i.e., less difficulty) than either Jr or Sr Sailors.

Summary: participants did not have much difficulty falling asleep as the mean daily ratings were lower than neutral on the scale. As the trial unfolded, there was a general trend with a reduction of mean rating, indicating that participants found it easier falling asleep at a later trial date. Notably though, the trend was not monotonic with two exceptions for Days 4 and 5 when elevated ratings were observed.

Across watch teams, STBD appeared to have more difficulty falling asleep than other watch teams. For three days in this trial (that is, 30% of the time), STBD reported a higher rating than one or more other watch teams. Comparison among rank levels showed a significant difference in rating for one day only with both Jr and Sr Sailors providing a higher rating than Officers. There was no difference in rating among six ship departments.

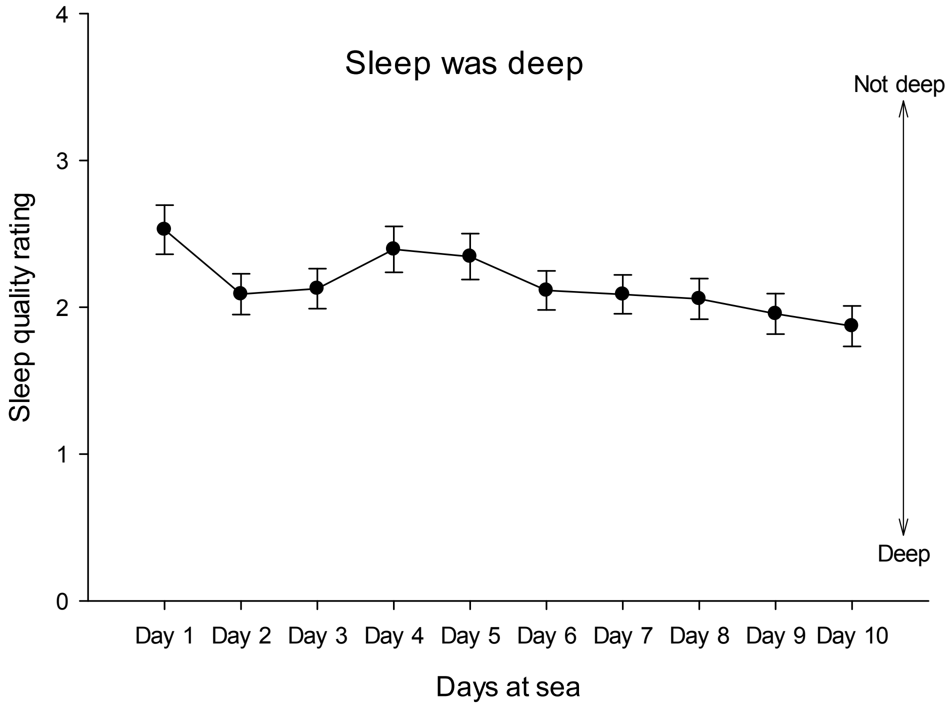
3.7.2 "Sleep was deep"

Figure 12 summarizes the results for the second sleep quality question. The average rating from all participants over ten days was 2.2, with daily mean ratings all slightly better than the half scale, indicating the sleep quality was somewhat deep.

A Friedman's one-way ANOVA showed a main effect of date, χ^2 (9) = 63.559, p < .001. The Wilcoxon Signed Ranks test confirmed a significant difference in rating for the following dates:

- Day 1, 4, 5 > Days 2, 3, 6–10
- Days 2, 3, 6, 7 > Day 10
- Day 3 > Day 9

[4] Note, a convention is followed in this report that uses either greater than (>) or less than (<) to indicate rating comparison between different dates or between different participant groupings. Date or grouping on the different side of the sign indicates a difference in rating that is statistically significant. Date or grouping on the same side of the sign does not have a significantly different rating.



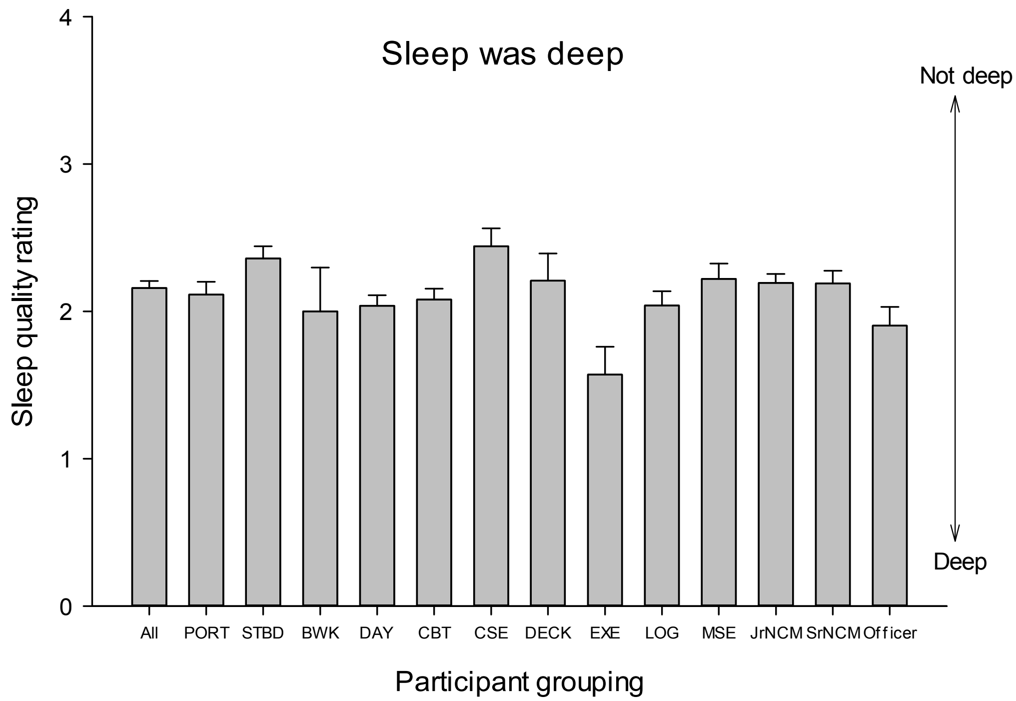


Figure 12: *Results for question 2 of the sleep quality survey "Sleep was deep": ratings comparison across trial date (top) and participant groupings (bottom).*

Kruskal-Wallis tests were performed to analyze differences across participant groupings. The results indicated that there was a main effect of department on Day 4, a main effect of watch on three days (Days 1, 3, and 4), and a main effect of rank on Day 6. Results from the post-hoc pair-wise Mann-Whitney tests are summarized as the following:

For three days (Days 1, 3, 4) in the trial, the sleep quality of STBD was less deep than either PORT or Dayworkers. On one day (Day 3), STBD's sleep was less deep than BWK. EXE and LOG had a lower rating (i.e., a deeper sleep) than the other four departments. However, their ratings were significantly better only for one day (Day 4). Across rank levels, the Officers' sleep was significantly deeper than either Jr or Sr Sailors on one day only (Day 6).

Summary: on average over this ten-day trial, participants rated the deepness of their daily sleep better than neutral, which can be interpreted as somewhat deep. The mean daily rating became lower (indicating an improvement in sleep quality) as the trial progressed, with an exception for Days 4 and 5 where the trend was disrupted.

The sleep was significantly less deep by the PORT watch keepers than STBD and Dayworkers on three different days (i.e., 30% of trial time). However, significant contrasts among all other participant groupings were identified only for one day. Further investigation with a longer trial duration is needed to confirm such differences.

3.7.3 "Arising was difficult"

Figure 13 summarizes the results for the third question in the sleep quality survey. The daily mean ratings were all lower than the midpoint of the scale, except on Day 7 when the mean rating was 2.6. Overall, the average rating over ten days was 2.4.

A Friedman's one-way ANOVA showed a main effect of date, χ^2 (9) = 29.102, p = .001. The Wilcoxon Signed Ranks test confirmed a significant difference in rating for the following dates:

- Day 1 < Days 3, 4, 6–8, 10
- Day 2 < Days 4, 7, 8
- Day 4 > Days 5, 9
- Day 5 < Day 7
- Days 6–8 > Day 9

Kruskal-Wallis tests were performed to analyze differences across participant groupings. The results revealed a main effect of department for four trial dates (Days 1, 3, 7, 8), a main effect of watch for five dates (Days 3–6, 10), and a main effect of rank for two days (Days 4, 8). Post hoc pair-wise comparison using Mann-Whitney tests confirmed a significant difference in rating for the following participant groupings:

- CBT > LOG on two days
- CBT < MSE on three days
- CSE > LOG on two days

- CSE < MSE on two days
- LOG < MSE on four days
- STBD> Dayworkers on three days
- STBD>PORT on two days
- STBD> BWK on one day
- Jr Sailors > Sr Sailors on two days.

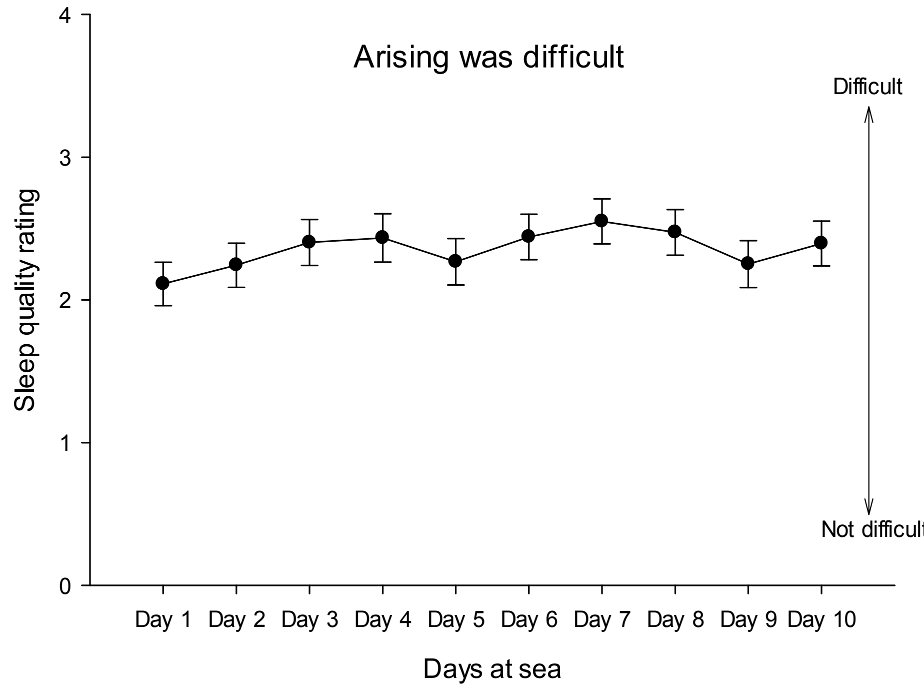
Summary: participants' mean rating for this question was close to the midpoint of the response scale, indicating arising from sleep was not too difficult. As the sail progressed, the difficulty level appeared to increase, with two exceptions on Days 5 and 9 when the trend was reversed.

Across ship departments, MSE and LOG were on two extreme ends of the response scale, with MSE reporting a significantly elevated difficulty to arise from sleep than LOG during 40% of the trial. Ratings from CBT and CSE were between those of MSE and LOG, and the difference was significant (i.e., less than MSE, greater than LOG) for 20% of the trial.

Among four watch groups, STBD found it more challenging to arise from sleep than the other three groups, specifically 30% of trial time than Dayworkers, 30% than STBD, and 10% than BWK.

Among three rank levels, Jr Sailors reported a higher rating. For 20% of time, their rating was significantly higher (i.e., more difficult to arise from sleep) than Sr sailors.

Notably, the mean ratings from BWK, EXE and Officers were among the highest for this question. However, due to the large variances that existed in the data, the ratings were not found to be significantly different from other comparable participant groupings. One way to address this issue in a future study is to increase the number of participants, consequently enhancing the power in statistical analysis.



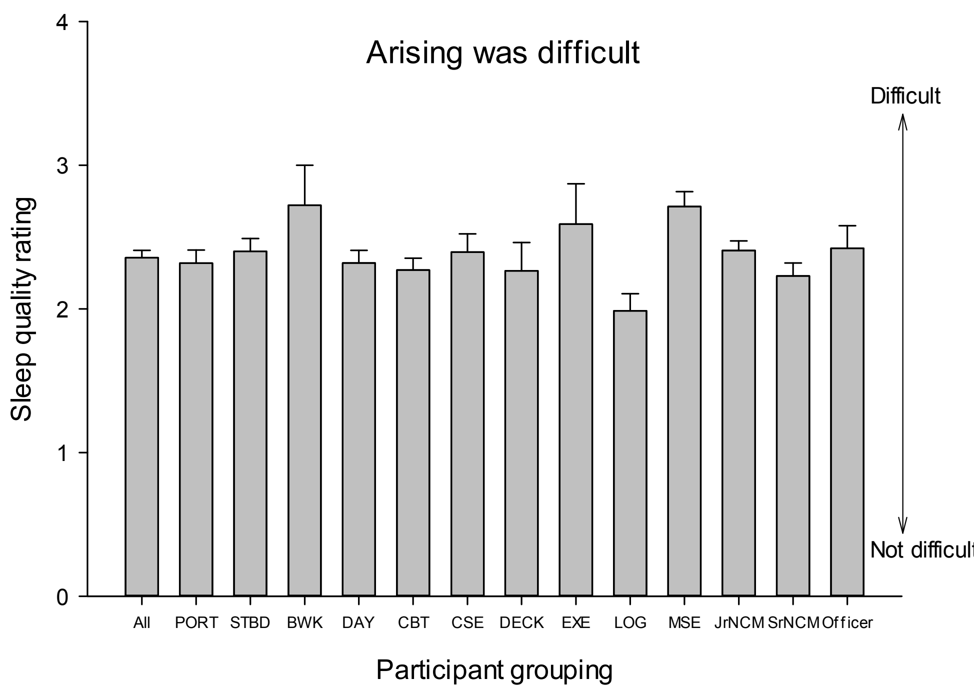


Figure 13: *Results for question 3 of the sleep quality survey "Arising was difficult": ratings comparison across trial date (top) and participant groupings (bottom).*

3.7.4 “Feel rested”

Figure 14 summarizes the results for the last question in the sleep quality survey. The daily mean ratings were all higher than the midpoint of the scale, except for Days 2 and 10. Overall, the average rating over ten days was 2.6, indicating participants’ feeling of rested from sleep was slightly worse than a neutral response.

A Friedman’s one-way ANOVA showed a main effect of date, χ^2 (9) = 30.098, $p < .001$. The Wilcoxon Signed Ranks test confirmed a significant difference in rating for the following dates:

- Day 1 > Day 2
- Day 2 < Days 4–9
- Day 3 < Days 4, 5, 9
- Day 5 > Day 10
- Day 9 > Days 8, 10

Kruskal-Wallis tests were performed to analyze differences across participant groupings. The results indicated that there was a main effect of watch for four trial days (1, 3–5). There was not a main effect for the factor of either department or rank. Post-hoc pair-wise comparison using the Mann-Whitney test confirmed the following results:

- STBD > Dayworkers on 4 days
- STBD > PORT on 2 days
- STBD > BWK on 1 day
- PORT > Dayworkers on 1 day

Summary: participants’ average rating on this sleep quality question was worse than neutral, indicating a general feeling that they did not feel very well rested from daily sleep. Elevated ratings (i.e., poorer sleep quality) were reported for Days 1, 4, 5, 9. Across watch teams, the quality of sleep appeared to be worst for the STBD watch team. For 40% of time, STBD reported a higher rating (i.e., worse sleep) than Dayworkers. For 10% of time, they reported a higher rating than either PORT or BWK. No difference was obtained across either rank levels or departments.

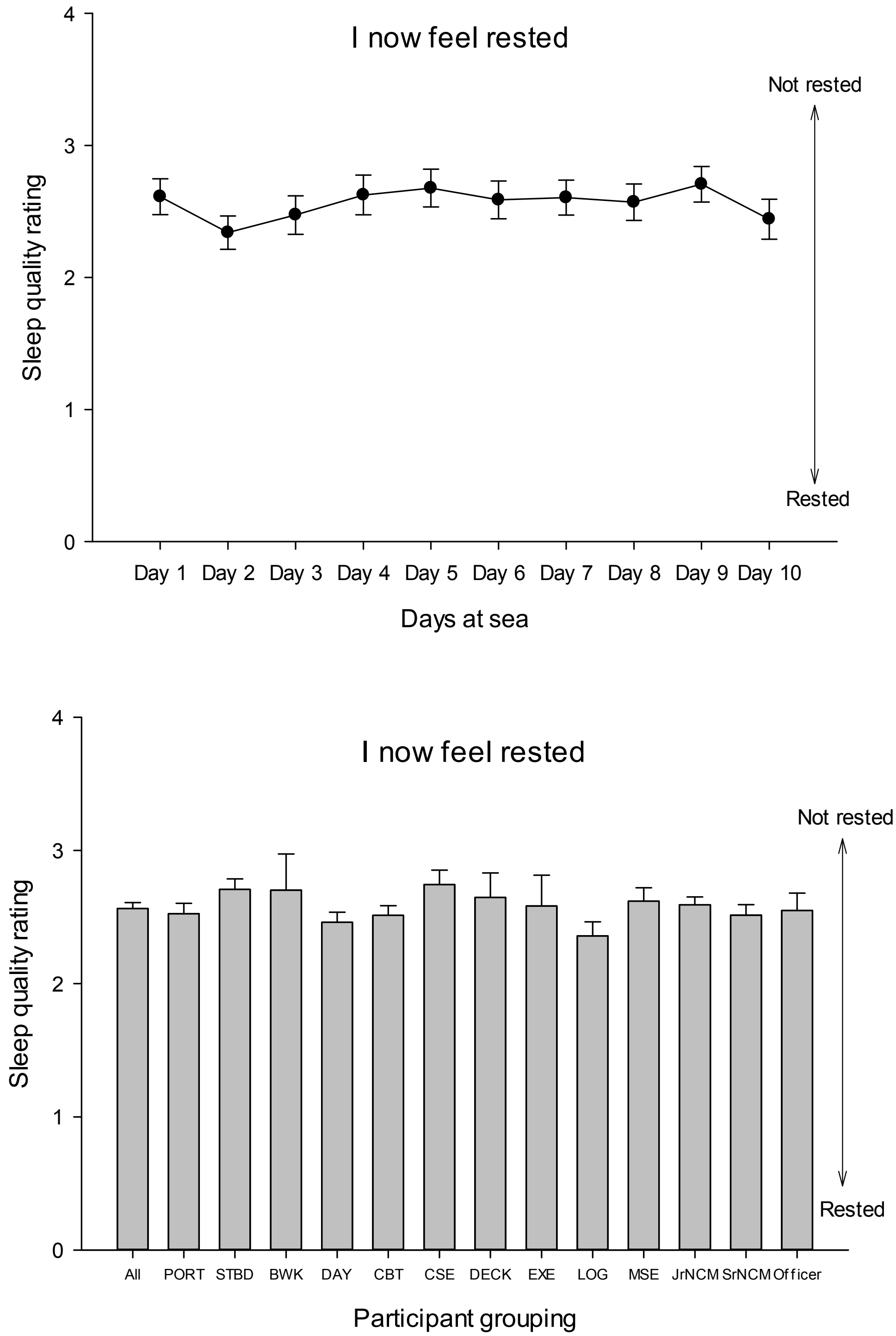


Figure 14: *Results for question 4 of the sleep quality survey "I now feel rested": ratings comparison across trial date (top) and participant groupings (bottom).*

3.8 Sleep disruption and reason

Participants were asked to record daily disruptions to their sleep, including the timing, duration and cause for each disruption. The responses from 160 participants over 10 trial dates were examined. The analysis was focused on both the amount of sleep disruption an average participant experienced every day and the main causes of disruption.

Key results are presented in Figure 15. Overall, participants reported an average of 28.8 minute sleep disruption each day, with a significant difference across trial dates, as confirmed by ANOVA, which revealed a main effect of date, $F(9, 1590) = 11.311$, $p < .001$, $\eta_p^2 = .060$. Post hoc pair-wise comparison using Tukey's HSD test confirmed that disruption was the longest on Day 9 (M=50.9, SD=59.8), than the rest of the trial dates except Day 5. Day 5 (M=46.4, SD=54.9) ranked the second with disruption longer than the rest except Days 1 and 6.

To compare across department, a mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of department was performed. Mauchly's test indicated that the assumption of sphericity had been violated (χ^2 (44) = 187.940, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .798$). The analysis only indicated a main effect of date, $F(7.185, 1106.546) = 9.316$, $p < .001$, $\eta_p^2 = .057$. The main effect of department was not significant, $F(5, 154) = 1.076$, $p = .376$, $\eta_p^2 = .034$.

For contrast among four watch groupings, a mixed-design ANOVA was conducted with a within-subjects factor of trial date and a between-subjects factor of watch. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(44) = 183.963$, $p < .001$), consequently degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .801$). The analysis indicated a main effect of date, $F(7.209, 1124.608) = 8.516$, $p < .001$, $\eta_p^2 = .052$, a significant interaction between date and watch, $F(21.627, 1124.608) = 4.216$, $p < .001$, $\eta_p^2 = .074$. The main effect of watch was not significant, $F(3, 156) = 1.815$, $p = .147$, $\eta_p^2 = .034$.

Across three rank levels, a mixed-design ANOVA was also conducted with a within-subjects factor of trial date and a between-subjects factor of rank. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(44) = 186.075$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .803$). The analysis indicated a main effect of date, $F(7.230, 1135.174) = 9.039$, $p < .001$, $\eta_p^2 = .054$ and the difference was not significant for the main factor of rank, $F(2, 157) = 1.216$, $p = .299$, $\eta_p^2 = .015$.

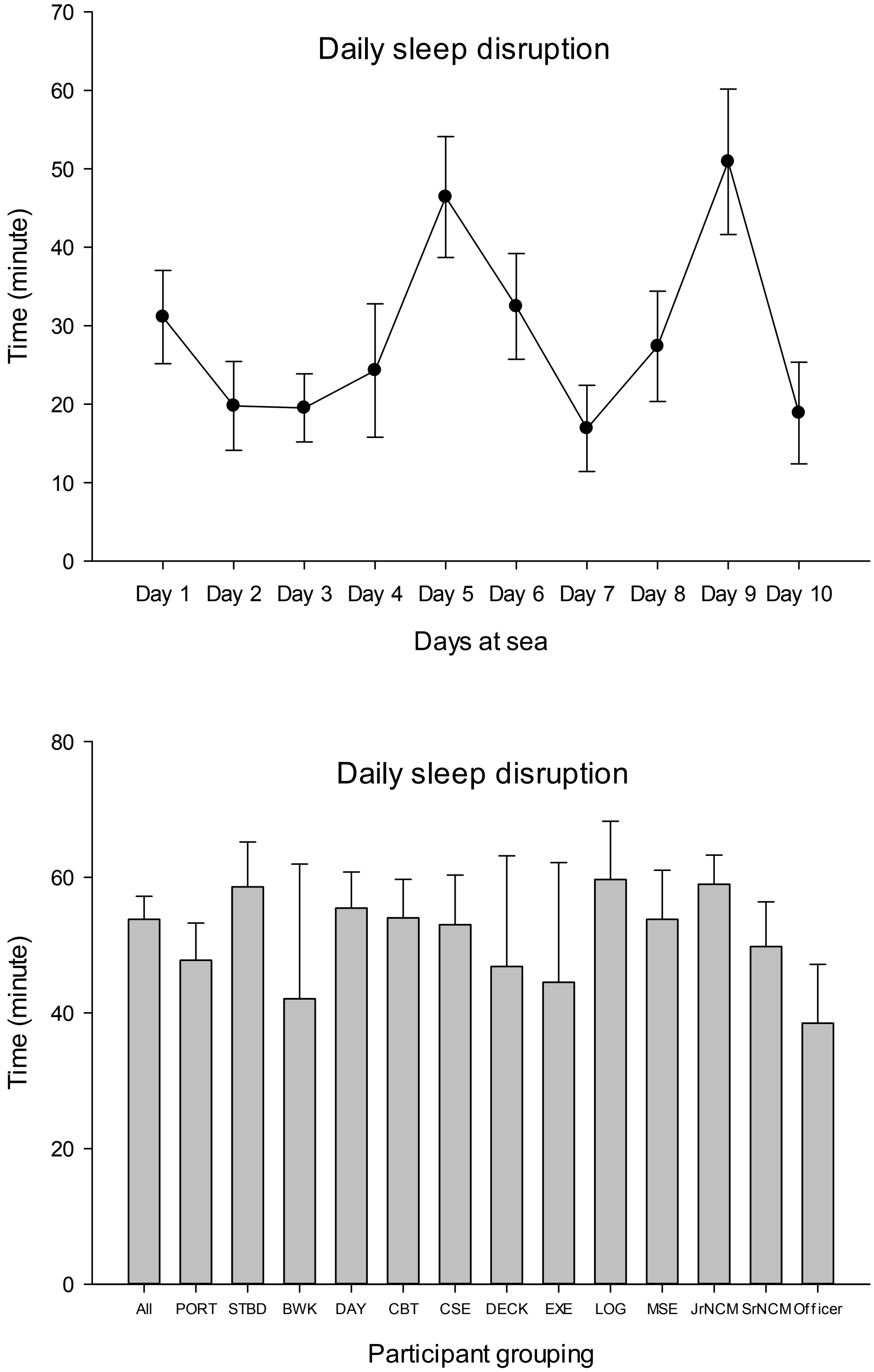


Figure 15: *Comparing daily sleep disruption across trial date (top) and participant groupings (bottom).*

Cause of sleep disruption

Over the ten days, a total of 1317 disruptions were recorded by 160 participants. On average, a participant's sleep was disrupted 2.2 times per day. Approximately half of all disruptions were caused by work related activities such as exercises and meetings. The remainder half was accountable by non-work related reasons, including noise from both environmental and operational sources as two major causes, followed by ship movement, bathroom and several other factors, as summarized in Table 12.

Table 12: *Main causes of sleep disruption and their frequency.*

Cause of sleep disruption	Frequency	Percentage
Work-related disruption (e.g., drills, meetings)	666	50.6%
Noise (environmental)	192	14.6%
Noise (operational, i.e., Alarm/pipe)	169	12.8%
Ship movement	70	5.3%
Bathroom	68	5.2%
Sick	21	1.6%
Temperature, smell, light	17	1.3%
Others	37	2.8%
Unknown	77	5.8%
Total	*1317*	*100%*

Summary: participants reported an average of 28.8 min disruption to their daily sleep. Such disruption was more severe on certain days such as the second last day of the trial (Day 9) when the average disruption reached 50.9 min. Sleep disruptions were experienced equally by all participants as comparison across department, watch or rank did not show a significant difference. Nearly half of all disruptions were work related. Noise from both environmental and operational sources were two other major categories of cause, which together were accountable for 27.4%.

3.9 Fatigue impact survey

MFIS was included in the pre-trial and post-trial surveys, and thus was filled out twice in this study on Day 1 and Day 10 respectively. In the pre-trial survey, a total of 159 responses were collected, 5 with incomplete data, therefore the total valid response number was 154. In the post-trial survey, a total of 154 responses were collected, 18 with incomplete data, resulting a total valid response number of 136. Among the valid surveys, 131 participants had responses available from both pre- and post-trial responses. Their data were analyzed using a matched-pairs statistical method.

Table 13: *MFIS ratings from pre-trial and post-trial surveys.*

MFIS Scales	N	Pre-trial survey		Post-trial survey		statistical significance
		M	SD	M	SD	
Physical subscale	131	10.0	5.5	14.6	6.8	Z = -7.031, p < .001
Cognitive subscale	131	12.4	6.4	16.9	6.8	Z = -6.270, p < .001
Psychosocial subscale	131	2.5	1.8	3.6	2.1	Z = -5.416, p < .001
Total MFIS rating	131	24.9	12.5	35.2	14.0	Z = -7.019, p < .001

Table 13 shows the average MFIS ratings, including ratings from its three subscales, obtained from two surveys. A Wilcoxon Signed-ranks test indicated a significantly higher rating in the post-trial survey than the pre-trial survey. Compared to Day 1, participants' average MFIS rating increased 41% by the end of the study. Such an elevation in fatigue was observed for three subscales as well, with an increase of 46% for physical, 36% for cognitive, and 44% for psychosocial subscales.

A further analysis of MFIS results is presented in Figure 16, in which the ratings were broken down according to participant's department, watch and rank groupings.

Across ship departments, there was no difference in MFIS rating in the pre-trial survey, $\chi^2(5) = 2.282$, $p = .809$. A significant difference however was found in post-trial survey, $\chi^2 (5) = 12.436$, $p = .024$, indicating the level of fatigue elevation differed across departments. A post-hoc pairwise comparison using Mann-Whitney U test revealed that MSE reported a higher level of fatigue than EXE, $Z = 1.145$, $p = .02$, and LOG, $Z = 2.69$, $p = .007$. CBT reported a higher level of fatigue than LOG, $Z = 2.52$, $p = .01$. In other words, the fatigue level as measured by FMIS was not different across department at the beginning of the study. After the 10-day trial, all departments reported an increased level of fatigue, with more pronounced increase by the engineering and combat departments (CBT 43.5%, CSE 51.0%, MSE 57.8%) than EXE (5.4%) and LOG (8.9%), and the difference was statistically significant between MSE and EXE, LOG, between CBT and LOG.

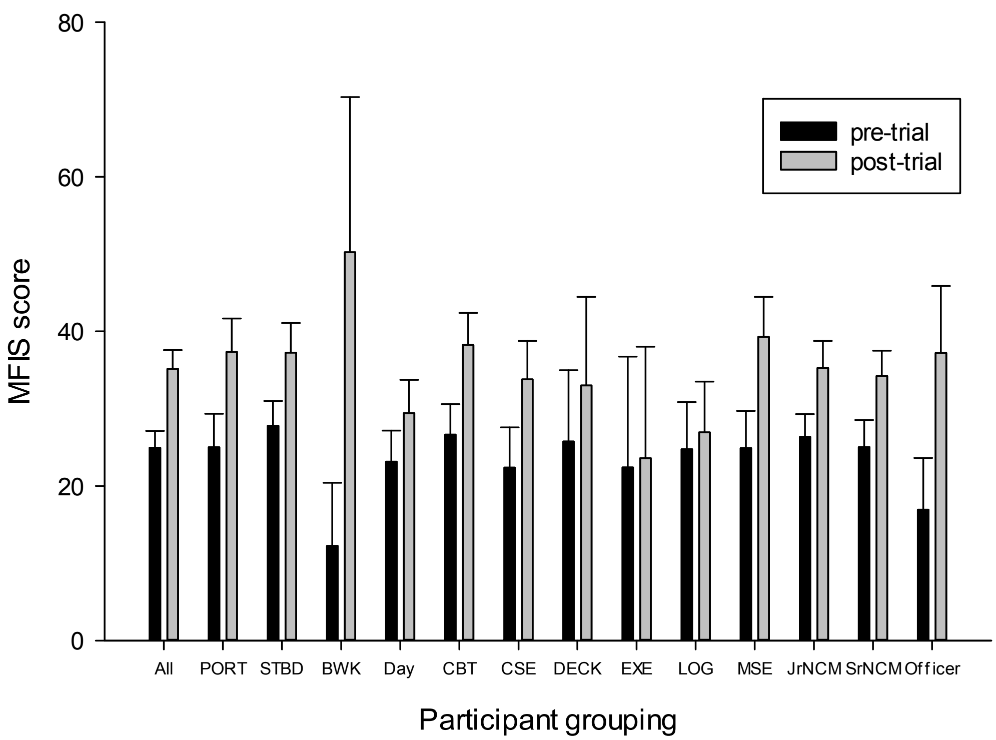


Figure 16: *MFIS ratings from pre-trial and post-trial surveys according to participants' department, watch, and rank.*

Across four watch groupings, a significant difference in MFIS rating was observed for both pre-trial ($\chi^2(3) = 8.513$, $p = .037$) and post-trial ($\chi^2(3) = 11.291$, $p = .010$) surveys. On the first day of trial, the reported MFIS ratings were lower by the BWK than either STBD watch keepers, $Z = 2.781$, $p = .002$, or PORT watch keepers, $Z = 2.076$, $p = .037$. On the last day of the study, the MFIS ratings by three watch groups were not different, but they were significantly higher than those reported by the Dayworkers ($Z = 2.146$, $p = .032$ for PORT; $Z = 2.425$, $p = .015$ for STBD; $Z = 2.438$, $p = .009$ for BWK). In other words, the small group of 1-in-3 BWK started this trial with a lower level of fatigue. After 10 days of sail, their fatigue level significantly increased and became equivalent with the larger group of 1-in-2 watch keepers. And compared to watch keepers, a significantly lower level of fatigue was reported by the Dayworkers at the end of the study.

Across three rank levels, a significant difference was detected in pre-trial survey ($\chi^2(2) = 6.996$, $p = .03$), with a lower level of fatigue reported by officers than both Jr sailors, $Z = 2.627$, $p = .009$, and Sr sailors, $Z = 2.202$, $p = .026$. In post-trial survey, there was no difference in reported MFIS ratings by three rank groups, $\chi^2(2) = 0.283$, $p = .868$. In other words, while officers reported a lower level of fatigue at the start of the trial, the elevated MFIS ratings provided by all three rank groups at the end of the study were the same.

3.10 Fatigue and Mood Survey (FMS)

FMS was administered each day during the trial. Responses from 159 participants were collected. Some participants did not complete this survey on the last day, as a result, a total of 112 responses for Day 10 was included in this analysis.

3.10.1 Alertness

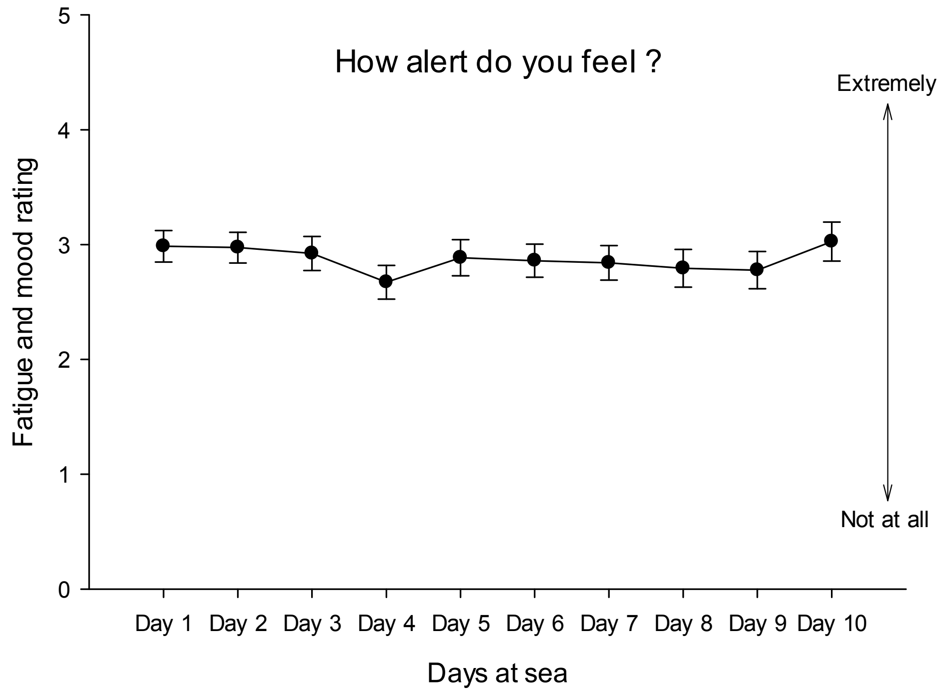
Participants' ratings on their daily alertness were analyzed and presented in Figure 17. Overall, an average rating of 2.9 was obtained for all participants over the ten-day trial, with daily average ratings fluctuating around the neutral score of 3 on the 5-point scale.

A Friedman's one-way ANOVA showed a main effect of date, $\chi^2(9) = 25.534$, $p = .002$. The Wilcoxon Signed Ranks test confirmed the following results:

- Day 4 < Days 1, 2, 3, 5, 6, 10
- Days 8, 9 < Days 1, 2
- Day 9 < Day 10

To compare alertness ratings among various participant groupings, the Kruskal-Wallis test did not identify a main effect of either watch ($\chi^2(3) = 4.448$, $p = .217$), department ($\chi^2(5) = 10.848$, $p = .054$), or rank ($\chi^2(2) = 3.982$, $p = .137$).

Summary: On average, participants provided a neutral response to this survey question and the mean alertness ratings were close to the midpoint of the scale. As the trial unfolded, a gradual reduction of mean alertness ratings was observed, with a severe degradation on Day 4 which had the lowest rating in this study. Compared to the first two days of the trial, the reduction of alertness became statistically significant on Days 8 and 9. On the last day, a recovery of alertness was observed. There was no difference in alertness ratings across either watch, department or rank groupings.



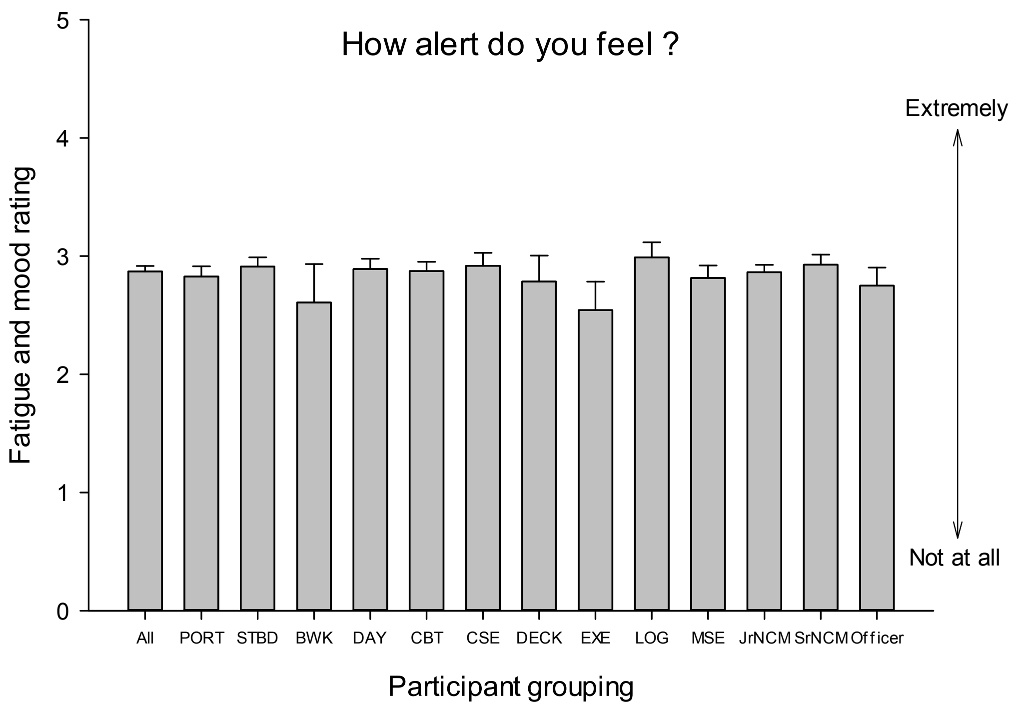


Figure 17: *Fatigue and mood survey: the daily alertness rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

3.10.2 Sadness

Participants rating on daily sadness are presented in Figure 18. Overall an average rating of 1.9 was obtained for the entire trial, and all daily mean ratings were lower than the midpoint of the scale, with the highest rating of 2.2 reported on Day 9.

A Friedman's one-way ANOVA showed a main effect of date, $\chi^2(9) = 65.448$, $p < .001$. The Wilcoxon Signed Ranks test confirmed the following results:

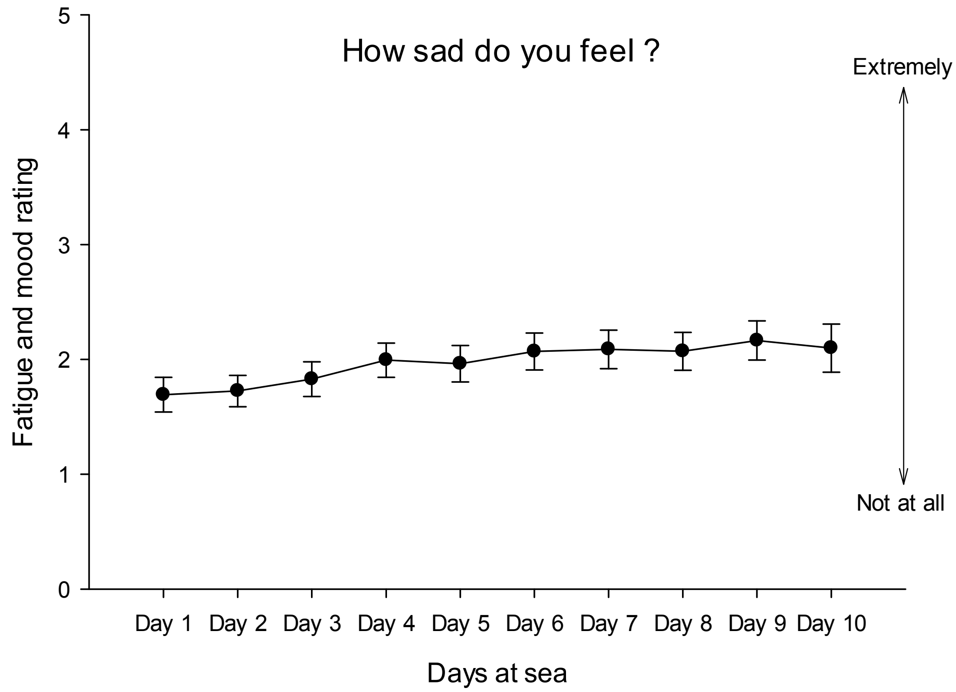
- Days 1, 2 < Days 4–10
- Day 5 < Day 9

To compare sadness ratings among various participant groupings, Kruskal-Wallis tests indicated a significant main effect of watch ($\chi^2(3) = 46.840$, $p < .001$), department ($\chi^2(5) = 46.230$, $p < .001$), and rank ($\chi^2(2) = 38.061$, $p < .001$). Results from the pairwise comparison using Mann-Whitney test are presented below:

- MSE > CBT, EXE, LOG, DECK
- CSE > EXE, LOG, DECK
- CBT > LOG, DECK
- STBD > PORT, Dayworker, BWK
- PORT > BWK
- Jr Sailors > Sr Sailors, Officers

Summary: On average, participants' daily ratings on sadness were lower than neutral, indicating an overall feeling of not very sad in this study. As the trial unfolded, an increase of rating was observed and the increase became statistically significant from Day 4 to Day 10 than the beginning of the trial (i.e., Days 1 and 2). The highest sadness rating was obtained on Day 9, followed by a slight reduction on the last day.

The sadness ratings differed significantly when participant responses were analyzed in different groupings. Participants from two engineering departments (MSE and CSE) and CBT reported a higher level of sadness than DECK and LOG. The 1-in-2 watch keepers (STBD and PORT) had a higher rating than either BWK or Dayworkers. Between the two 1-in-2 watch teams, ratings from STBD were higher than that of PORT. Jr Sailors felt sadder than Sr Sailors and Officers.



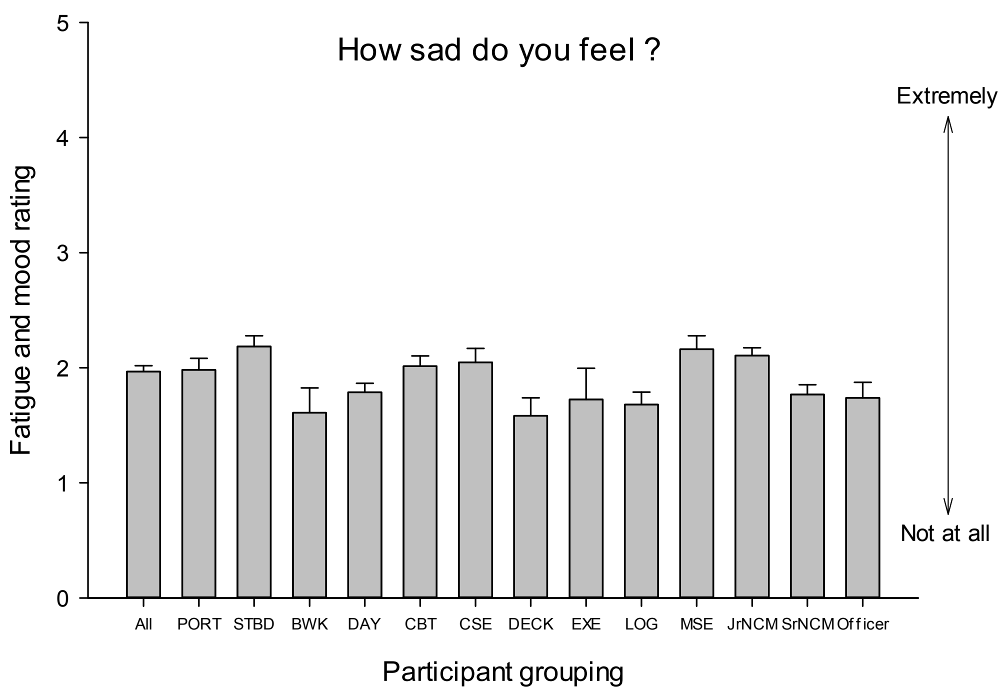


Figure 18: *Fatigue and mood survey: the daily sadness rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

3.10.3 Tenseness

Participants rating on the feeling of tenseness are analyzed and presented in Figure 19. Throughout the trial, the daily mean ratings were all below the neutral rating (i.e., 3), with an overall average score of 2.4.

A Friedman's one-way ANOVA did not show a main effect of date, $\chi^2(9) = 8.215$, $p = .513$. To compare various participant groupings, Kruskal-Wallis tests indicated a significant main effect of watch ($\chi^2(3) = 41.904$, $p < .001$), department ($\chi^2(5) = 71.006$, $p < .001$), and rank ($\chi^2(2) = 10.681$, $p = .005$). Results from the follow-up pairwise comparison using Mann-Whitney U tests revealed significant differences in the following groupings:

- CSE, MSE > EXE, DECK, LOG
- CBT > DECK, LOG
- STBD, PORT > BWK, Dayworker
- STBD > PORT
- Jr Sailors > Officer, Sr Sailors

Summary: On average, participants' daily ratings on tenseness were lower than neutral, indicating an overall feeling of not very tense throughout this study. The daily mean rating did not change significantly as the trial progressed. However, there existed a significant difference between participant groupings. In particular, a higher level of tenseness was reported by engineers and operators in CSE, MSE and CBT than those in DECK and LOG departments. 1-in-2 watch keepers also felt more tense than Dayworker and BWK. Among 1-in-2 watch keepers, those who stood STBD watch reported a higher level of tenseness than PORT.

3.10.4 Effort

Results from daily effort ratings are presented in Figure 20. Overall the average rating was 2.6, with the daily mean effort scores less than 3 (i.e., neutral) for all trial dates.

A Friedman's one-way ANOVA identified a main effect of date, $\chi^2(9) = 37.341$, $p < .001$. The Wilcoxon Signed Ranks Test confirmed the following pair-wise comparison results:

- Day 1 < Days 4–9
- Day 2 < Days 4, 5, 7, 8, 9
- Day 3 < Days 4–10
- Day 4 > Days 6, 10
- Day 6 < Day 8

To compare effort ratings among various participant groupings, Kruskal-Wallis tests indicated a significant main effect of watch ($\chi^2(3) = 10.740$, $p = .013$), and department ($\chi^2(5) = 27.142$, $p < .001$). No main effect was identified for the factor of rank ($\chi^2(2) = 2.259$, $p = .323$). Results from the pairwise comparison using Mann-Whitney tests revealed a significant difference in rating between the following groupings.

- MSE > CBT, CSE, DECK, LOG
- CBT, EXE > LOG
- BWK > STBD, PORT, Dayworker

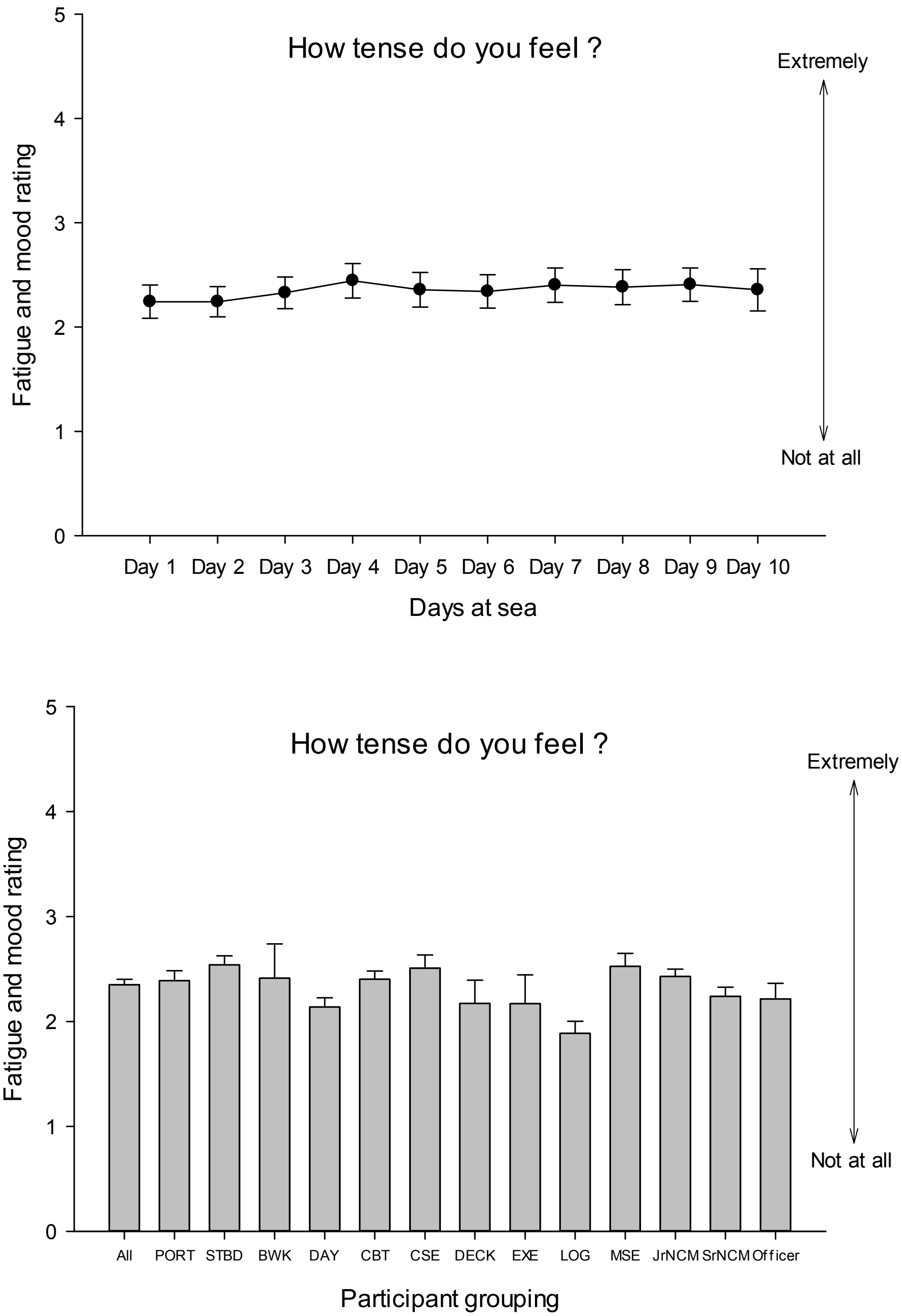
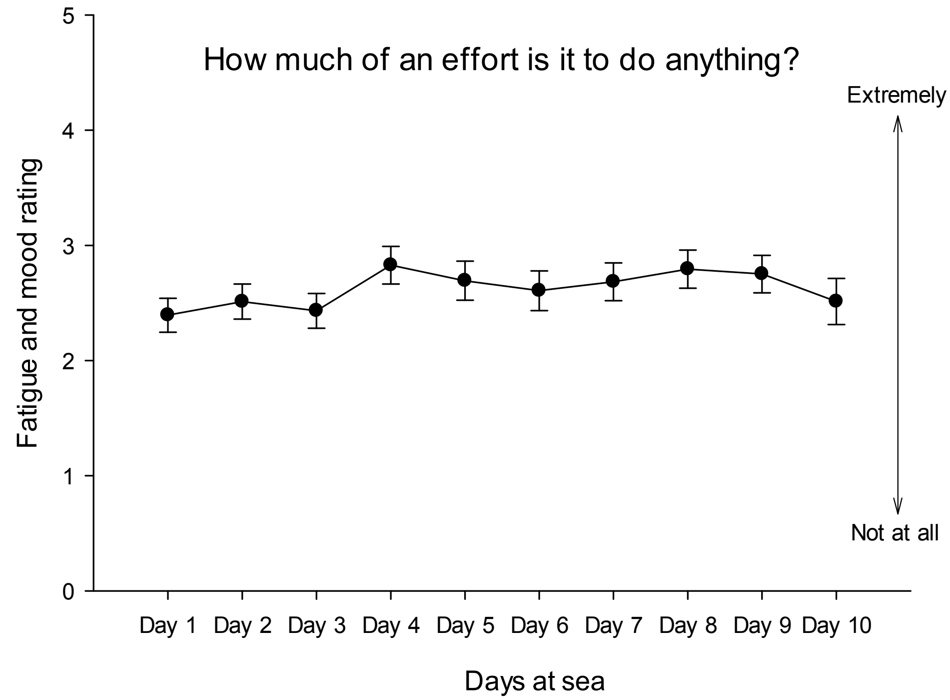


Figure 19: *Fatigue and mood survey: the daily tenseness rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*



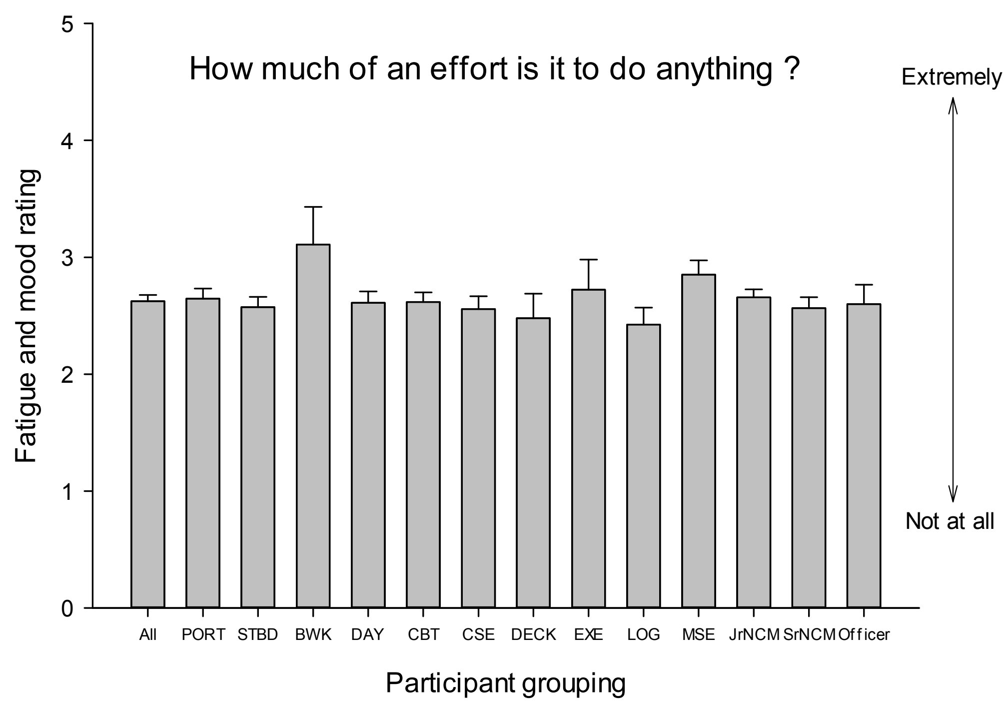


Figure 20: *Fatigue and mood survey: the daily effort rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

Summary: Mean daily effort ratings were better than neutral throughout the trial, with a significant increase of effort level in later trial dates, that is Day 4 and later than the first three days. A reduction of effort rating was observed on the last day. Between departments, MSE, EXE and CBT reported a higher level of effort, significantly higher than LOG. Among four watch groupings, the highest rating was obtained from BWK, indicating a significantly more effort was required by this group of participants than all others.

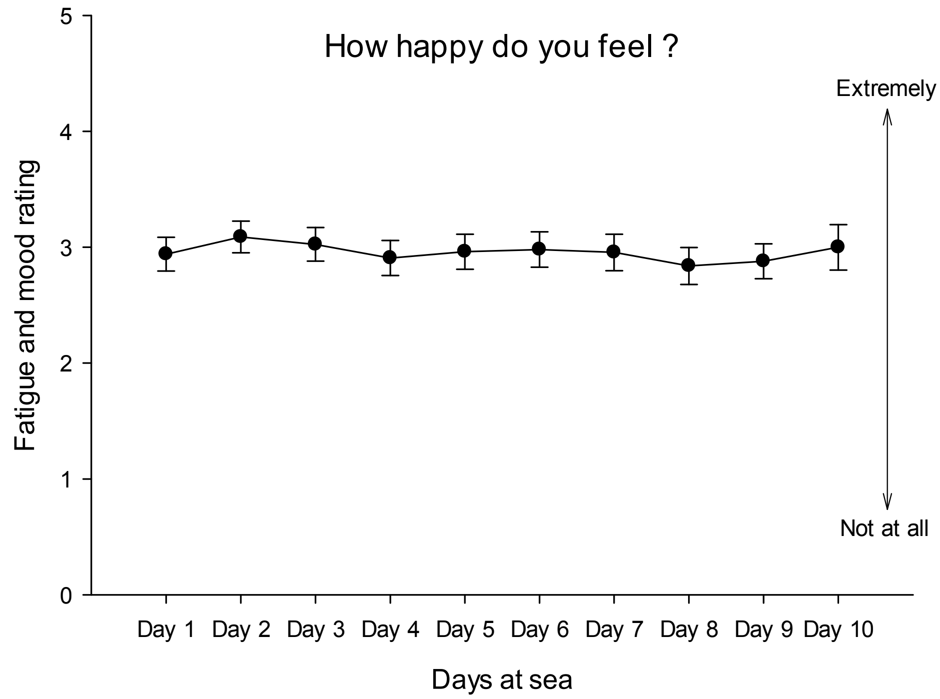
3.10.5 Happiness

Results from the daily happiness rating are summarized and presented in Figure 21. An average rating of 2.9 was obtained from all participants for the entire trial, with all daily mean ratings fluctuating around the midpoint of the rating scale.

A Friedman's one-way ANOVA did not indicate a main effect of date, $\chi^2(9) = 9.374$, $p = .404$. Kruskal-Wallis tests indicated a significant main effect of watch ($\chi^2(3) = 56.031$, $p < .001$), department ($\chi^2(5) = 81.584$, $p < .001$), and rank ($\chi^2(2) = 42.084$, $p < .001$). The pairwise comparison using Mann-Whitney U tests revealed the following results:

- EXE > all other five departments
- DECK, LOG > CBT, CSE, MSE
- CBT > MSE
- Dayworker, BWK > PORT, STBD
- PORT > STBD
- Officers, Sr Sailors > Jr Sailors

Summary: participants' happiness ratings were neutral and did not change significantly during this study. The small group of EXE participants was happier than those of all other departments. On the other end, the two engineering and the combat departments had lower happiness ratings, with an intermediate level of ratings reported by LOG and DECK. Among watch keepers, the Dayworkers and BWK were happier than either PORT or STBD. Between two 1-in-2 watch teams, STBD was less happy. Among three rank levels, Jr Sailors were less happy than Sr Sailors and Officers.



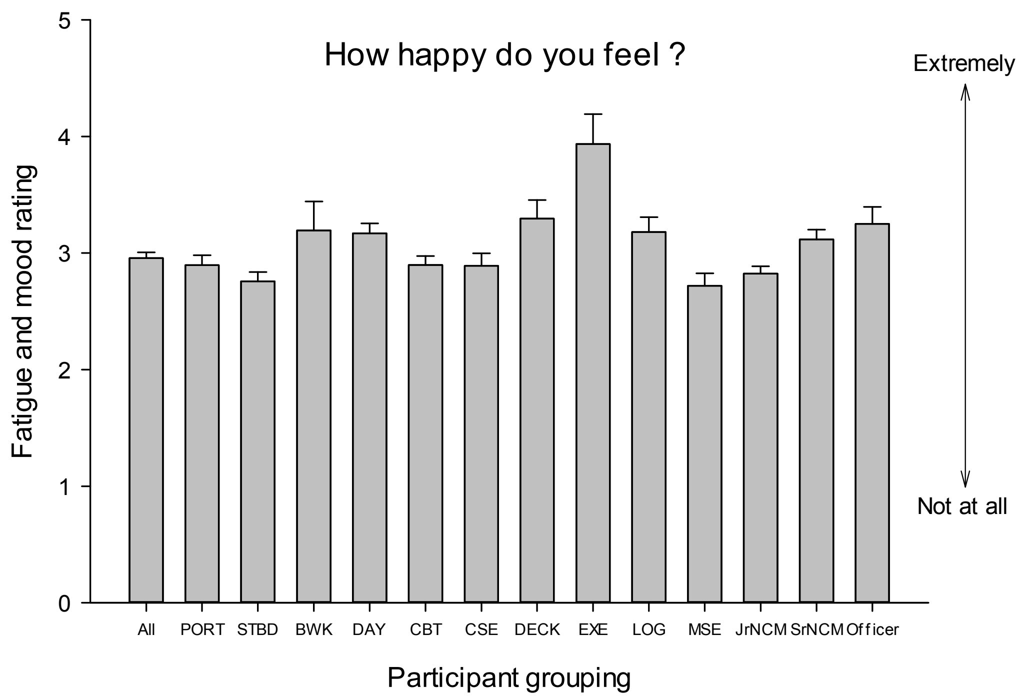


Figure 21: *Fatigue and mood survey: the daily happiness rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

3.10.6 Weariness

Results from daily ratings of weariness are summarized in Figure 22. Overall, an average rating of 2.8 was obtained for the entire trial, with daily mean ratings all below the midpoint of the scale.

A Friedman's one-way ANOVA showed a main effect of date, $\chi^2(9) = 20.512$, $p = .015$. The Wilcoxon Signed Ranks test confirmed the following results:

- Day 1 < Days 4, 8, 9
- Days 2, 5, 7 < Day 9
- Days 3, 5, 7 < Day 4

Kruskal-Wallis tests indicated a significant main effect of watch ($\chi^2(3) = 11.305$, $p = .010$), department ($\chi^2(5) = 40.570$, $p < .001$), and rank ($\chi^2(2) = 14.823$, $p = .001$). The pairwise comparison using Mann-Whitney U tests revealed significant difference in rating for the following participant groupings:

- DECK, LOG < EXE, MSE, CSE, CBT
- MSE < EXE
- PORT, Dayworker < STBD
- Jr Sailors < Officers, Sr Sailors

Summary: average ratings on weariness were better than a neutral response throughout this trial, indicating a general feeling of not too weary. A pronounced elevation of weariness was observed on Days 4 and 9. Across ship departments, a significantly lower level of weariness was reported by DECK and LOG. PORT watch keepers and Dayworkers were less weary than STBD, Jr Sailors less weary than either Officers or Sr Sailors.

3.10.7 Calmness

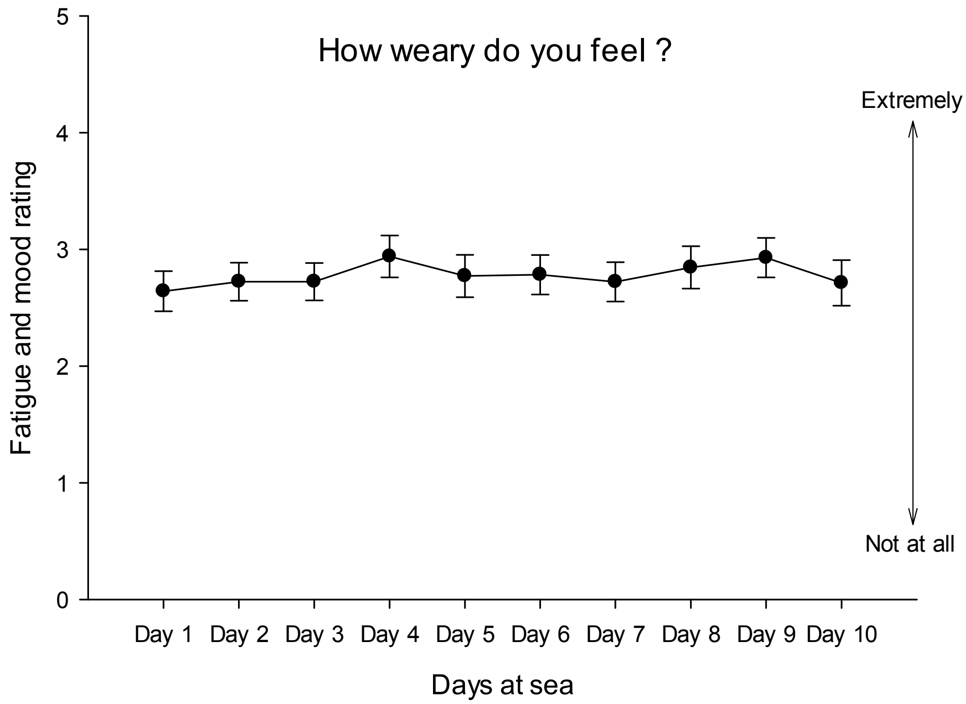
Results from daily ratings on calmness are presented in Figure 23. Overall an average rating of 3.2 was obtained from this study, with all daily ratings greater than neutral, i.e., the midpoint of the rating scale.

A Friedman's one-way ANOVA showed a main effect of date, $\chi^2(9) = 26.010$, $p = .002$. The Wilcoxon Signed Ranks test confirmed the following results:

- Day 2 > Days 5–10
- Day 3 > Days 5, 6, 7, 9, 10
- Day 4 > Days 5, 7, 8

Kruskal-Wallis tests indicated a significant main effect of watch ($\chi^2(3) = 55.961$, $p < .001$), department ($\chi^2(5) = 77.672$, $p < .001$), and rank ($\chi^2(2) = 22.490$, $p < .001$). The pairwise comparison using Mann-Whitney U tests revealed significant differences for the following participant groupings:

- EXE > LOG, DECK > CBT > CSE, MSE
- Dayworkers, BWK > PORT> STBD
- Officers > Sr Sailors > Jr Sailors



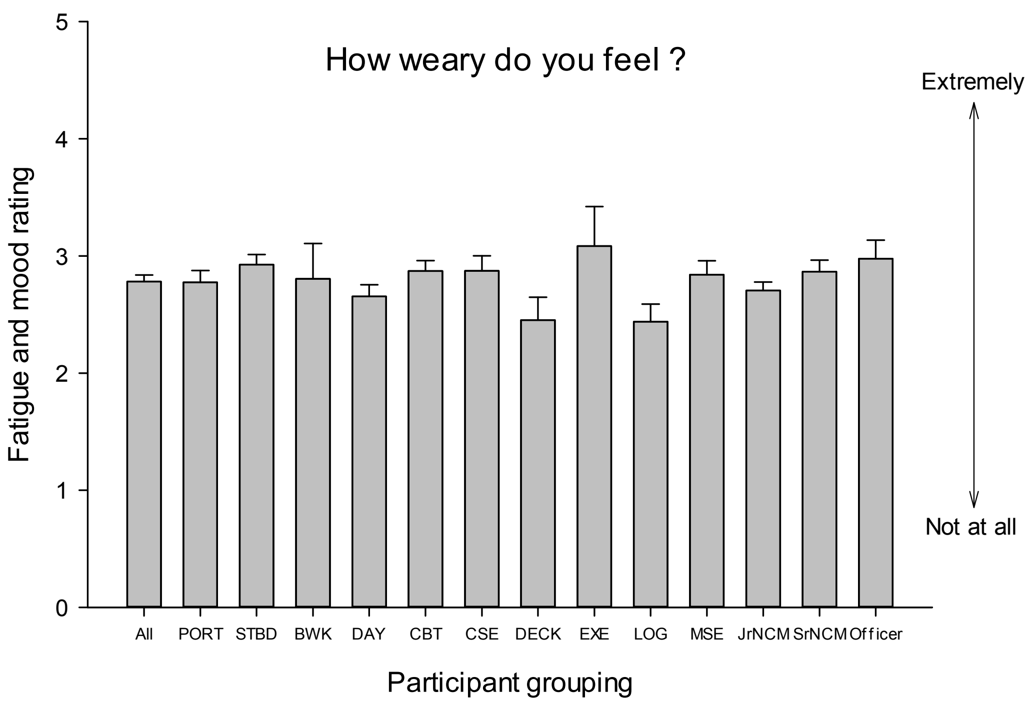
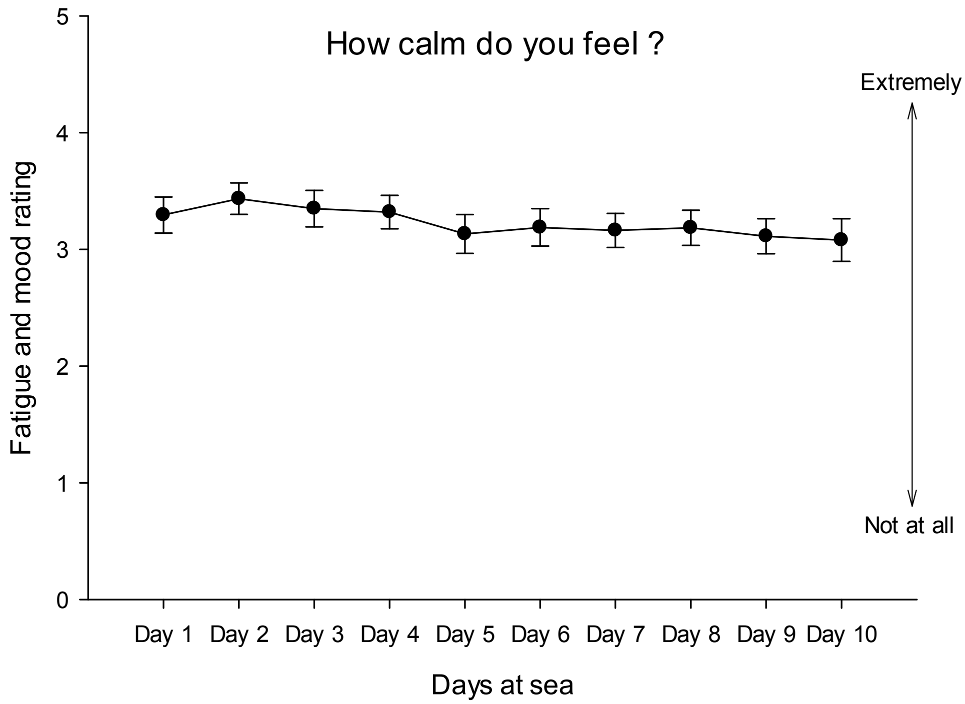


Figure 22: *Fatigue and mood survey: the daily weariness rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*



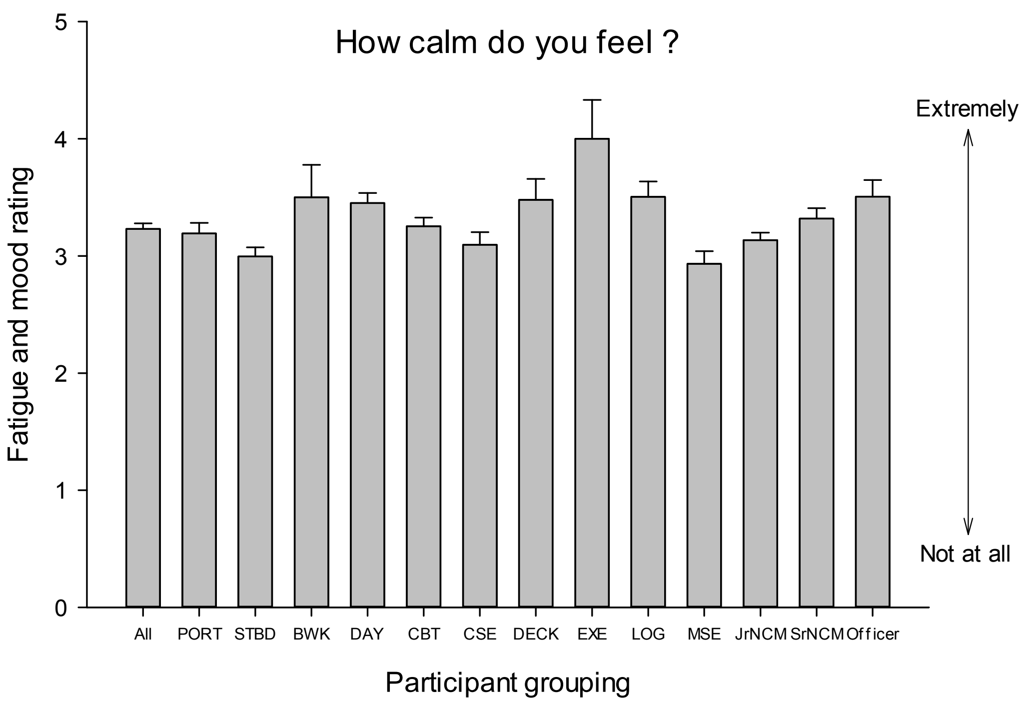


Figure 23: *Fatigue and mood survey: the daily calmness rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

Summary: daily ratings on calmness were all greater than neutral, indicating participants in general were quite calm during this study. The highest level of calmness appeared on Day 2, following by a general reduction of ratings as the trial progressed. Among ship departments, ratings from EXE was the highest (i.e., calmest), followed by DECK and LOG. Lower ratings were obtained from CBT, CSE and MSE, with the two engineering departments providing the lowest ratings. For watch groupings, Dayworkers and BWK were calmer than PORT and STBD. Between two 1-in-2 watch teams, PORT was calmer than STBD. Significant differences in rating existed among three rank levels, with Officers, Sr Sailors, and Jr Sailors ranked in that order.

3.10.8 Sleepiness

Lastly for the FMS, results from daily sleepiness ratings are presented in Figure 24. An average rating of 3.3 was obtained in this study, with daily ratings all greater than neutral on the scale.

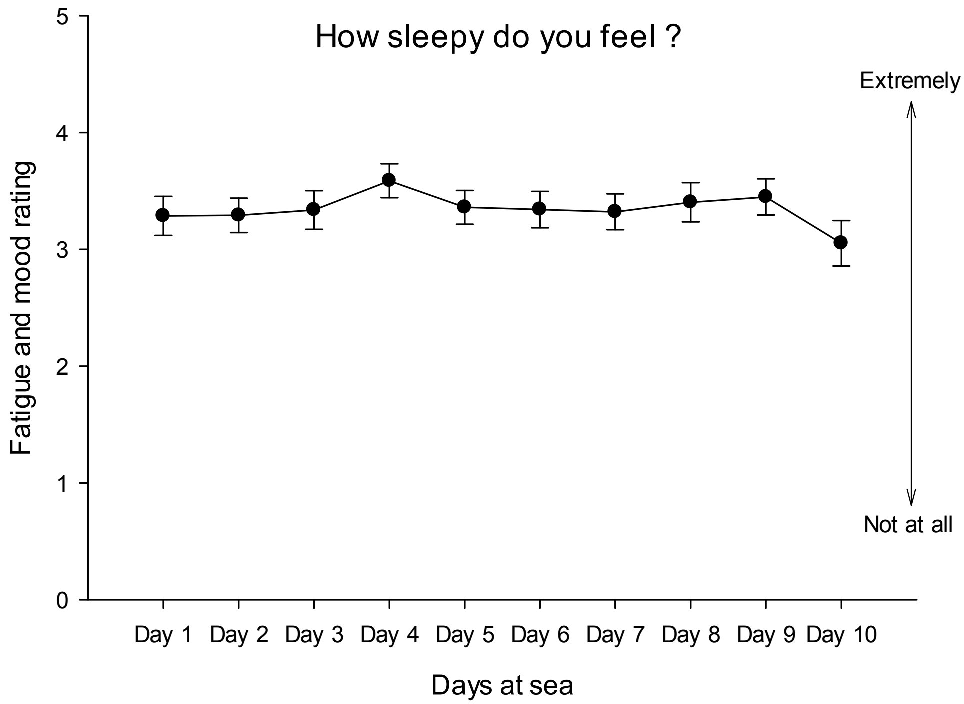
A Friedman's one-way ANOVA showed a main effect of date, $\chi^2(9) = 18.322$, $p = .019$. The Wilcoxon Signed Ranks test confirmed the following results:

- Day 4 > Days 1, 2, 3, 5, 6, 7, 10
- Days 6, 8, 9 > Day 10.

Kruskal-Wallis tests indicated a significant main effect of watch ($\chi^2(3) = 15.462$, $p = .001$), department ($\chi^2(5) = 34.385$, $p < .001$), and rank ($\chi^2(2) = 11.468$, $p = .003$). The pairwise comparison using Mann-Whitney test revealed significant differences in rating among the following participant groupings:

- EXE > CBT, CSE, MSE, LOG
- EXE, DECK, CBT, CSE, MSE > LOG
- BWK, PORT > STBD and Dayworkers
- Officers > Jr Sailors, Sr Sailors

Summary: Participants' daily ratings on sleepiness were worse than neutral, reflecting a moderately elevated level of sleepiness throughout the trial. Notably, this is the only indicator in the FMS where the average daily ratings resided on the negative side of the scale. The peak rating was observed on Day 4, and a decrease of sleepiness was found on the last day (Day 10). Among departments, EXE and LOG were on two different extreme ends with ratings significantly different from the other four departments. EXE felt sleepier than all other departments except DECK. For watch groupings, a higher level of sleepiness was reported by BWK and PORT than STBD and Dayworkers. Lastly, Officers were sleepier than either Jr or Sr Sailors.



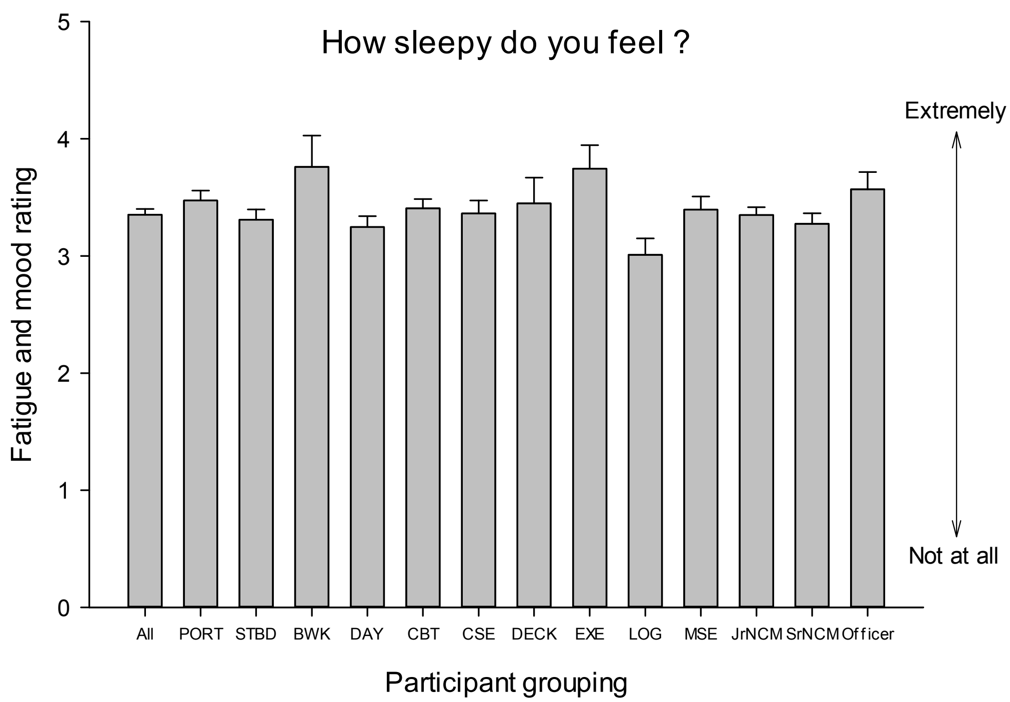


Figure 24: *Fatigue and mood survey: the daily sleepiness rating over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

3.11 Sleepiness rating from activity log

An alternative method to measure daily sleepiness, based on the Stanford sleepiness scale, was also conducted in this study. In the activity log, participants were asked to provide a rating (using a 7-point scale, as shown previously in Table 4) to indicate their sleepiness for each 30-min block of time when they were awake. Ratings for eight trial dates were included in the examination, with data for the first and the last day removed from analysis.

Results are presented in Figure 25. Overall, the average sleepiness rating was 2.7 which reflects a state between a rating of 2 ("Functioning at a high level, but not at peak, able to concentrate") and 3 ("Relaxed; awake; not at full alertness, responsive").

A Kruskal-Wallis test to compare average sleepiness rating across eight trial dates indicated a main effect of date, $\chi^2(7) = 27.744$, $p < .001$. Post-hoc pair-wise comparison using Mann-Whitney U tests confirmed a significant difference in rating for the following dates:

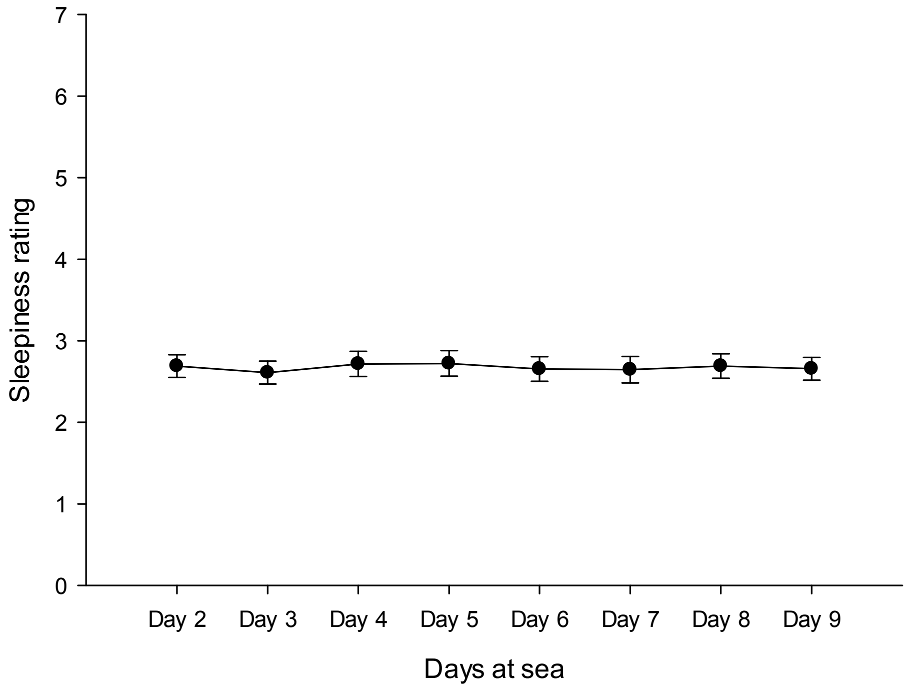
- Days 5, 8 > Days 2, 3, 6, 7
- Days 4, 9 > Day 3

To compare ratings across participant groupings, Kruskal-Wallis tests indicated a main effect of department ($\chi^2(5) = 900.012$, $p < .001$), watch ($\chi^2(3) = 1539.339$, $p < .001$), and rank ($\chi^2(2) = 213.248$, $p < .001$). Post-hoc pair-wise comparison revealed a significant difference in rating for the following groupings:

- LOG < EXE < CBT < CSE < DECK < MSE
- Dayworker < PORT < STBD, BWK
- Sr Sailors < Jr Sailors, Officers

A further analysis was conducted to examine the relationship between sleepiness rating and the time of day. This analysis was based on a grouping of participants by their watch syndicate. The results are illustrated in Figure 26, in which the duty periods of each watch syndicate are highlighted in red boxes. The following two insights were obtained from this analysis.

First, although the average sleepiness rating remained quite low in this study, there were periods that participants experienced a severe elevation of sleepiness. For most participants, the highest level of sleepiness was experienced at time that was scheduled as their regular sleep time. Based on operational requirements (including drills), participants sometimes were woken up from their sleep and consequently resulting in a period with a high rating on sleepiness.



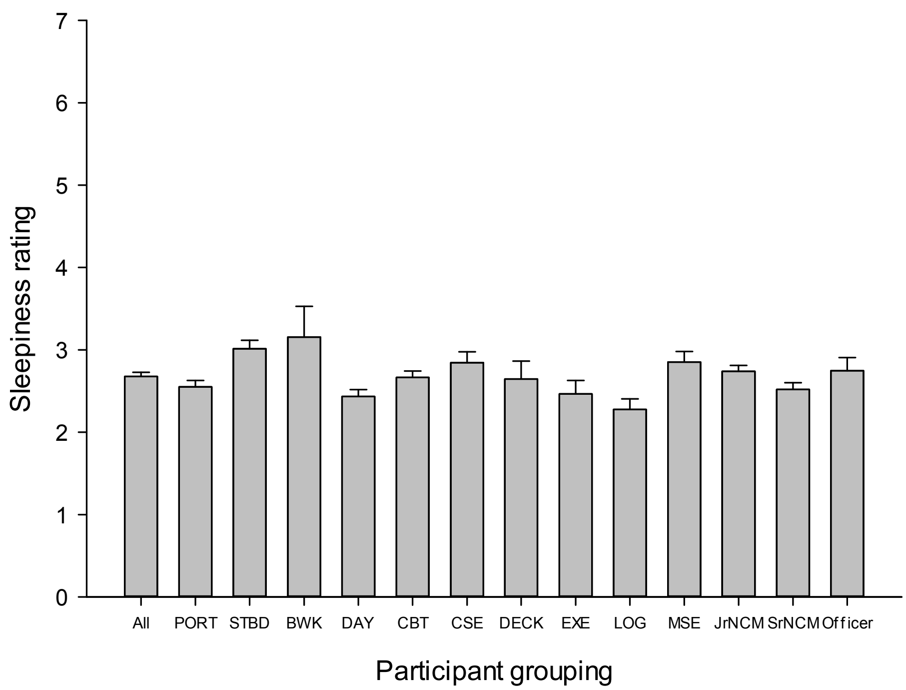


Figure 25: *Sleepiness ratings over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

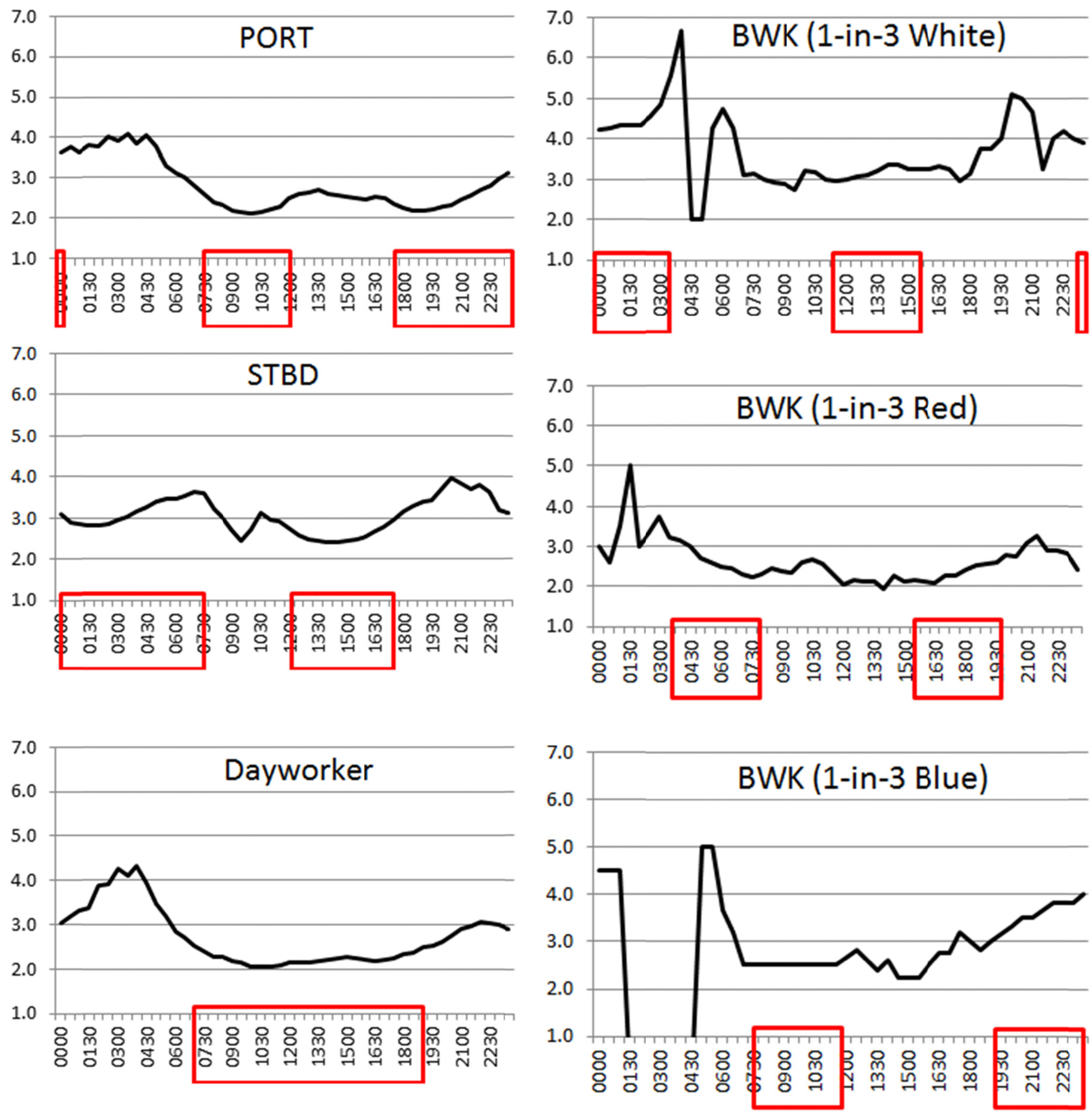


Figure 26: *Sleepiness rating profiles for watch groupings.*

Second and more importantly, an elevation of sleepiness was also observed for participants within their watch keeping duty periods. Generally, an increase of rating was observed towards the end of each duty period, in most cases the level of elevation was within an acceptable range. However, two BWK teams, that is 1-in-3 White team and 1-in-3 Blue team, need a closer examination particularly for their duty periods at night time. The ratings from the Blue team showed a rapid elevation of sleepiness towards midnight with the average rating approaching 4 (i.e., which indicates a feeling of "A little foggy; not at peak; let down"). The situation was worse for the White team, the average sleepiness ratings consistently exceeded 4 during the duty period between 23:30 and 03:30. A similar pattern exists for the 1-in-2 STBD team as well, with an elevated sleepiness during the second half of their night shift (i.e., roughly between 04:30 and 07:30), although their condition was less severe compared to BWK.

Summary: sleepiness ratings recorded in activity logs provided a more precise picture of participants' fatigue level. Overall, an average rating of 2.7 indicated that participants were able to maintain an adequate level of alertness in this trial. However, there were significant differences in rating among department, watch groups and rank levels. Across departments, higher ratings were obtained by MSE and DECK, with LOG reporting the least level of sleepiness. Among watch teams, those with a significant period of work time at night (e.g., BWK and STBD) reported a higher level of sleepiness. Across rank levels, a higher rating was provided by Officers and Jr Sailors than Sr Sailors.

3.12 Caffeine and sleep medication intake

The intake of caffeine and sleep medication was tracked in the daily log. The results revealed that the caffeine intake was primarily in the form of coffee and tea, plus a small amount of energy drinks. No sleep medication was recorded in survey responses. In this subsection, the results are based on an analysis of participants' daily consumption of either coffee or tea. More specifically, the amount of caffeine intake was estimated based on the quantity of fluid that was consumed.

Several different units were used in the log to indicate the amount of coffee or tea participant drank during the trial, including milliliter, ounce, and cup (or mug). In this analysis, we converted all units to milliliter and used an assumption that 1 cup equals 295 ml (i.e., 10 ounces).

The results revealed that a participant drank an average of 823 ml of coffee or tea each day. ANOVA did not show a main effect of date, $F(9, 905) = .413$, $p = .929$, $\eta_p^2 = .004$.

To compare across ship departments, a mixed-design ANOVA with a within-subjects factor of trial date and a between-subjects factor of department was performed. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(44) = 407.077$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .543$). The analysis indicated a main effect of date, $F(4.887, 752.532) = 6.212$, $p < .001$, $\eta_p^2 = .039$ and the difference was not significant for the main factor of department, $F(5, 154) = 1.552$, $p = .177$, $\eta_p^2 = .048$.

Based on the same analysis method, comparison across watch groupings was examined using a mixed-design ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(44) = 423.845$, $p < .001$), thus degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .530$). The analysis did not show a main effect of either date, $F(4.774, 744.760) = 1.526$, $p = .182$, $\eta_p^2 = .010$, or watch ($F(3, 156) = .737$, $p = .521$, $\eta_p^2 = .014$.

A mixed design ANOVA was conducted to compare three rank levels. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(44) = 426.873$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .529$). The analysis indicated a main effect of date, $F(4.763, 747.856) = 3.536$, $p = .004$, $\eta_p^2 = .022$, and rank ($F(2, 157) = 7.086$, $p = .001$, $\eta_p^2 = .083$. Post-hoc test using Tukey's HSD confirmed that Sr Sailors consumed more coffee or tea than Jr Sailors.

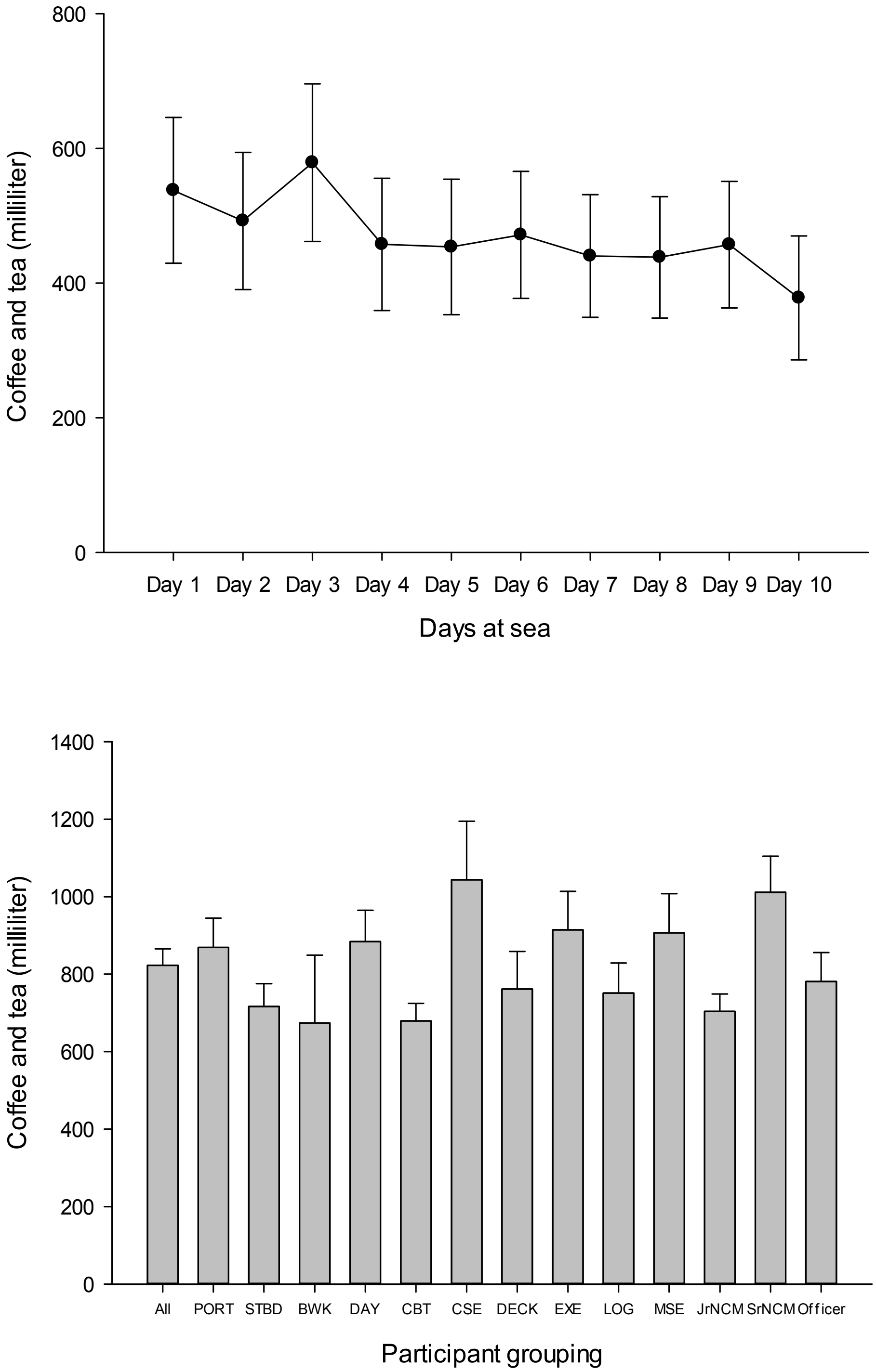


Figure 27: *Consumption of coffee and tea over trial dates (top) and comparison among watch, department and rank groupings (bottom).*

4 Discussion

Significant results from this study are discussed in this section, including findings on fatigue and mood (Section 4.1), work duration and workload (4.2), sleep duration and quality (4.3). A set of profiles is suggested for participant groupings, in Section 4.4, based on the similarity in their response to the fatigue and mood survey.

4.1 Evaluation of crew fatigue and mood

A significant elevation of fatigue was observed during this trial for all participants. The post-trial MFIS survey that was conducted on the tenth day of the sail indicated an increase of fatigue score by 41% compared to results from the first day.

This is generally confirmed by results from the daily fatigue and mood survey as well. Among the eight survey questions, four showed a change of ratings that reflected a gradual increase of fatigue or deterioration of mood, including alertness, sadness, effort, calmness. The difference became statistically significant when ratings at the end of trial were compared to those collected at the beginning of the trial. For the remaining four questions (i.e., tense, happiness, weariness, sleepiness), the daily mean ratings stayed flat throughout the trial.

There was a significant increase of fatigue level during this ten-day trial. The mean MFIS score from all participants increased from 24.9 on Day 1 to 35.2 on Day 10. To put these scores into perspective, the MFIS rating has a range from 0 to 84. Previous studies used a cutoff rating of 38 to discriminate fatigued from non-fatigued individuals [20]. Based on such a cutoff threshold, it can be concluded that participants on average were not fatigued on the tenth day of the sail. This conclusion was also generally supported by FMS results. Participants' ratings on all eight fatigue and mood questions were either close to or better than neutral (i.e., the midpoint on the scale), except the rating on sleepiness.

However, it is important to point out that elevation of fatigue was not equally experienced by all participants. Some groups witnessed a more pronounced increase of fatigue than others. Specifically, watch keepers were found to be more tired than Dayworkers at the end of the trial. In particular, the small group of 1-in-3 bridge watch keepers had a severe increase of fatigue. Their mean MFIS score on the tenth day was 50.3, significantly exceeding the cutoff threshold (38). The scores from PORT and STBD were 37.4 and 37.3 respectively, very close to the threshold as well. Across ship departments, MSE and CBT had a mean score greater than 38. Officers' mean score (37.2) was also close to the threshold. A closer examination of these participant groups is needed. Their operational sustainability becomes a concern since many naval operations last longer, and often much longer than ten days.

The mean ratings of daily fatigue and mood fluctuated, sometimes significantly, from day to day. In this study, an elevation of fatigue or a reduction of mood consistently appeared on day 4 and day 9 in ratings of multiple fatigue/mood attributes as well as in sleepiness rating from activity log. A recovery of fatigue and mood was also observed on the last day of the trial (Day 10). Such findings can likely be explained by poor sleep quality on Day 4 due to high sea state and a long work duration on Day 9. Although such findings were not the main objectives of this investigation, the results nonetheless demonstrated the sensitivity of the survey instruments.

Lastly, participants' fatigue as measured by sleepiness ratings indicated that there existed periods of time during a day when they experienced a severe elevation of sleepiness. Many such periods were not a major concern to operational performance as they occurred during participants' off-duty hours. However, for three watchkeeping teams, that is 1-in-2 STBD, 1-in-3 White, and 1-in-3 Blue, an elevated level of sleepiness was reported during their duty periods that encompassed evening and night hours. The results revealed a potential operational risk associated with reduced alertness by watchkeepers during these periods.

4.2 Evaluation of work duration and workload

On average, a participant worked 13.3 h per day during this study, which translates into a 93.1 h weekly total on-duty time. To provide a context for interpreting this result, the United States Navy (USN) uses a Navy Standard Workweek (NSWW) to determine required manning levels for USN ships [21]. NSWW has designated 81 hours as a crew member's weekly total on-duty time. In this trial, the shortest daily work time was recorded by the LOG department (12.3 h/day), even their weekly on-duty time was 86.1 h/week, longer than NSWW specification.

While long work duration is generally applicable to all crew, it is important to highlight the risk it imposes on two participant groups. Among watch teams, BWK reported a mean daily work time of 15.3 h. For 45% of the dates in this study, BWK's daily work time exceeded 16 hour per day. Across rank levels, Officers worked on average 15.2 hours per day. The work duration likely is a major contributor to the pronounced increase of fatigue experienced by these two groups of participants.

Besides work time, we also studied the nature of sailors' work in WRS and tracked task demand (i.e., workload) in daily activity logs. An emphasis on cognitive skills is apparent for most departments. Communication, problem-solving, and timesharing have in general been rated higher than physical skills like strength and dexterity. This set of data enables researchers to analyze performance risks due to fatigue based on the required skillsets. It also supports the development of experimental tests and performance metrics in future studies that represent sailors' work more precisely.

With respect to workload, the results from this study indicated a higher level of task demand reported by Dayworkers (than watch keepers), the EXE department (than all other departments), and Officers (than either Jr or Sr Sailors). However workload does not appear to be a major driver of fatigue, as daily average workload ratings obtained in this trial remained overall at a moderate level.

4.3 Evaluation of sleep duration and quality

In this study, participants reported an average daily TIB of 7.4 h, it was significantly shorter than the amount of TIB they received before the sail (8.4 h/day). Their actual daily sleep time, as measured by actigraph which removed SOL and WASO, was further shortened to 6.3 h/day. For a week, the total sleep time would be 44.1 hours. To help interpret this result, Table 14 is compiled to compare the finding with similar studies conducted in the past by allied navies. Overall given the variability in the data, the total weekly sleep times from these studies appear to be comparable, all are shorter than the 56 h weekly sleep time specified in NSWW.

Based on the TIB measure, BWK, EXE, and Officers were three groups that stood out. Not only were their mean TIB the shortest in their respective grouping, they also had the largest day-to-day variability. The average TIB for BWK was 6.0 h/day before the consideration of sleep disruptions. Participants of the EXE department had below-than-average daily TIB. So were the Officers, when their TIB was compared to those of Jr and Sr Sailors.

Table 14: *A comparison of weekly total sleep time.*

Type of platform	Weekly total sleep time
RCN frigate (current study)	M=44.1, SD=16.6
USN frigate [21]	M=47.0, SD=9.2
RAN frigate [22]	M=50.2, SD=10.4
USN destroyer [23]	M=51.1, SD=7.5
USN cruiser [24]	M=48.5, SD=7.2
USN NSWW specification	56 hour

When the actigraph measure were examined, a shorter daily mean sleep time was found for BWK, DECK and Jr Sailors in their respective grouping, with the DECK department as a group having the shortest duration of 4.7 h/day. However due to the large variability in the data, the comparisons across department, watch or rank did not reveal a statistically significant contrast. In terms of practical significance, a large sleep variability itself is a useful indicator as it often reveals poor sleep hygiene. Among all participant groups, the largest variabilities in sleep duration were observed for BWK, EXE, LOG, and Officers. Since a major cause of such variabilities was participants' work schedule, the results indicated that participants of these groups had a less structured work pattern and therefore should be examined more closely in future crewing trials and SCORE modelling.

Results from the sleep quality survey revealed that a consistent difference existed in comparisons among watch groups, indicating the time of sleep plays a primary role in affecting sleep quality. Among four watch groupings, the quality of sleep was poorer for STBD than all others.

The poor sleep quality reported by STBD watch keepers is likely affected by several factors, including circadian rhythm and sleep disruption. More than half of STBD participants had an intermediate chronotype, indicating their daily peak alertness naturally occurs during the day time (between morning and evening). This may contribute to a reduction of sleep quality as a significant portion of sleep by STBD takes place during the daytime. The mean sleep disruption time was the highest for STBD, however, due to the large variability in the data, the disruption time was not statistically significant from other watch groups.

Overall a significant amount of sleep disruption (2.2 times/day with a total duration of 28.8min) was observed in this study and it was experienced equally by all participants. The major cause was operational requirements which were accountable for nearly half of the time reported. For all other disruptions, noise was identified to be the leading factor. A degradation of sleep quality was also universally observed on Days 4 and 5 when the sea state was the highest in this trial.

4.4 Fatigue and mood profiles

The results obtained from this study allow us to create a fatigue and mood profile for each participant group. In this subsection, we reorganized the results, particularly from FMS, in a tabular form to better visualize differences across the three grouping factors of department (Table 15), watch (Table 16) and rank (Table 17). In these tables, each descriptor between a pair of participant groups indicates a statistically significant contrast. A judgement was made regarding the quality of each contrast from the perspective of the group shown in the first column, with desirable ones presented in the top half of the cell and undesirable ones in the bottom half.

A qualitative judgment on the similarity between different participant groupings can be made. For example, a fewer number of descriptors between a pair of groupings reflects a higher level of similarity in terms of their fatigue and mood; the difference in number between desirable and undesirable descriptors indicates which group has a more preferable profile.

Table 15: *Contrast of significant difference in FMS ratings across six ship departments.*

	CSE	DECK	EXE	LOG	MSE
CBT	CBT more calm		CBT less sleepy		CBT less sad Less effort More happy More calm
		CBT more sad More tense Less happy More weary Less calm	CBT less happy Less calm	CBT more sad More tense More effort Less happy More weary Less calm More sleepy	
CSE			CSE less sleepy		CSE less effort
		CSE more sad, More tense Less happy More weary Less calm	CSE more sad More tense Less happy Less calm	CSE more sad More tense Less happy More weary Less calm More sleepy	
DECK			Deck less weary		Deck less sad Less tense Less effort More happy Less weary More calm
			Deck less happy Less calm	Deck more sleepy	
EXE				EXE more happy More calm	EXE less sad Less tense More happy More calm
				More effort More weary More sleepy	More weary More sleepy
LOG					Log less sad Less tense Less effort More happy Less weary More calm Less sleepy

Table 16: *Contrast of significant difference in FMS ratings across four watch groups.*

	STBD	BWK	Dayworker
PORT	PORT less sad Less tense More happy Less weary More calm	PORT less effort	
	More sleepy	PORT more sad Less happy Less calm	PORT More sad More tense Less happy Less calm More sleepy
STBD		STBD Less effort Less sleepy	
		STBD more sad Less happy Less calm	STBD more sad More tense Less happy More weary Less calm
BWK			
			BWK more effort more sleepy

Table 17: *Contrast of significant difference in FMS ratings across three rank levels.*

	Officer	Sr NCM
Jr Sailors	Jr Sailors less weary Less sleepy	Jr Sailors less weary
	Jr Sailors more sad More tense Less happy Less calm	Jr Sailors more sad More tense Less happy Less calm
Officer		Officer more calm
		Officer more sleepy

Key insights obtained from this analysis are presented as following.

1. Among six ship departments, there existed three sets of fatigue and mood profiles. CSE, MSE and CBT shared a similar profile, as there was only one difference (i.e., one descriptor) in each pairing among these three departments. The same could be concluded for LOG and DECK. EXE had a unique profile that was different from the rest.

2. The profiles for MSE, CSE and CBT were worse than those of DECK and LOG. When either MSE, or CSE, or CBT was compared to either DECK or LOG, the descriptors were all undesirable and worse for MSE, CSE or CBT. A different pattern emerged when EXE was compared to another department. Typically the comparison showed a number of desirable and undesirable descriptors,

indicating a mixed result. A general pattern was that EXE reported a higher level of both effort and sleepiness (e.g., undesirable), at the same time a high level of calmness and happiness (desirable).

3. Among four watch groupings, each had a unique fatigue and mood profile. The Dayworkers' profile was clearly better than those of watch keepers, with all descriptors desirable in favor of the Dayworkers. Across three watch keeper groups, the profile of STBD was worse than those of PORT or BWK.

4. Comparing among three rank levels, the profiles of Officers and Sr Sailors were similar, with two descriptors, one desirable and one undesirable, between them. Their profiles in general were better off than that of Jr Sailors.

It is useful to point out that the profiles discussed above are approximate in nature. There are other important factors that affect a crew member's fatigue and mood, such as individual differences, that were not considered in this analysis. Detailed attributes in a profile also reflect the specific operation in which the data were collected. With a good understanding of these caveats, these profiles have some conceivable utility. They provide a relatively simple way to summarize, at a high level, the data collected in this study. For future crewing studies, these profiles provide some insight on the similarities or differences among members of the crew, and assist the identification of potential issues of concern, as well as corresponding solutions.

5 Conclusion

We studied the crew of HMCS Montreal during a ten-day at sea trial in which a high intensity work-up scenario was followed. The work and rest patterns of 160 crew members were analyzed, together with subjective measures of their fatigue and daily mood.

On average, the crew worked 13.3 hours and slept 6.3 hours every day in this trial. They experienced a significant increase of fatigue during the ten days of sail. Significantly different patterns of results emerged when the data were analyzed by grouping participants according to their department, watch or rank. A summary of key insights is presented as following.

The 1-in-3 bridge watch keepers were identified as a group with an elevated risk for fatigue. A combination of long work duration and short sleep time appeared to be the major cause of fatigue. Two BWK groups in particular, i.e., 1-in-3 White and Blue teams, need to be flagged, as the highest ratings of sleepiness observed in this study were reported by these members while they stood their night watch. With a severe elevation of fatigue after ten days of sail, BWK as a group has a high risk for burnout in a long deployment timeline.

The increase of fatigue was substantial for the 1-in-2 watch keepers, significantly larger than that of Dayworkers. Besides one obvious cause of sleep restriction (as the daily sleep of watch keepers was typically fragmented into two episodes), other sleep quality factors (such as circadian rhythm and sleep disruption) also played an important role. The STBD watch keepers in particular reported a significantly poorer quality in sleep than others, including their counterpart who stood PORT watch. Their opportunities for sleep, as afforded by the current work schedule, appear to be more susceptible to sleep disruption due to either circadian desynchrony or operational tempo.

Across rank levels, officers as a group reported long work duration and short daily sleep time. They faced task demand (i.e., workload) that was comparable to Sr Sailors, but significantly higher than Jr Sailors. Their short sleep was mitigated to some extent by a higher sleep quality. And a mixed result was obtained from their mood survey, that is, they felt more weary and sleepy, but at the same time were calmer and happier. The cause of such a seemingly conflicting result may be due to motivational factors that were not investigated in this study.

Across ship departments, three profiles were compiled for departmental groupings based on the similarity in response to the fatigue and mood survey. The two engineering departments and the combat department shared one profile that was in general less desirable than the one shared by the logistic and the deck department. Among all eight fatigue and mood attributes that were studied, none of the responses from engineering and combat departments was better than that of the logistic or the deck department. The profile for the small team of executives had a quality that was between the other two, with a mix of desirable and undesirable results when the executives' fatigue and mood ratings were compared to participants of other departments.

Results from this study also inform on-going research on naval crewing. The empirical data collected in this trial directly supports the validation of SCORE and its performance prediction module DFM. A separate analysis has been completed to examine the way SCORE is applied to model crew size and composition based on this X-Ship trial data [24]. For DFM, its prediction algorithms were developed based on past research mostly conducted in laboratory settings or the aviation domain (i.e., for airplane pilots). To apply DFM in naval operations, an urgent research task is to validate its underlying models using empirical data collected onboard of naval platforms. To our knowledge, this X-Ship trial was the first such a study with a purpose to improve fatigue prediction models for maritime operations. Results obtained from this study will support the selection of the most suitable model, as well as its parameter calibration, to improve fatigue prediction. More specifically, the study facilitates an improvement of DFM's ability to account for the impact of naval environmental stressors and predict fragmented daily sleeps by ship watchkeepers with an increased precision.

Naval platform crewing is a complex issue determined by an array of inter-connected factors. This X-Ship study provides a framework to analyze these factors, as well as a set of metrics suitable for data collection in at-sea trials. The study was the first of a series of crewing trials that have been planned. Results presented in this report provide a CPF crew's work-rest schedule benchmark based on an established complement of 217. In future trials, alternative crewing options will be examined and the results will be compared to the benchmark to assist decision-making. From the perspective of research and development, we will further investigate the use of standardized cognitive tests to examine the impact of fatigue on operator performance based on objective metrics. Such research will enhance the scientific rigor of the modelling tool and enable us to better assist RCN in addressing crewing related design challenges for future platforms.

References

[1] Chow, R., Wang, W., Lamb, M., Coates, C., Perlin, M., and McKay, P. (2013). Simulation for crew optimization and risk evaluation. In *TTCP HUM TP-18 Workshop on Crewing Systems for Maritime Warfare*: Portsdown West, UK.

[2] Chow, R., Burke, R., Labbe, D., Charest, G., and Lamb, M. (2015). Options analysis for the Arctic Offshore Patrol Ship (AOPS) crew: A SCORE case study. (DRDC-RDDC-2015-L100) Defence Research and Development Canada.

[3] Chow, R., Lamb, M., Charest, G., and Darragh, A. (2014). Crew validation for the Canadian Patrol Frigate: A SCORE case study. (DRDC-RDDC-2014-L50) Defence Research and Development Canada.

[4] Chow, R., Lamb, M., Smith, E., Woodburn, M., Collins, M., Coates, S., and Saltel, B. (2015). Preliminary crew analysis for Canadian Surface Combatant (CSC): Phase 2 crew validation. (DRDC-RDDC-2015-L187) Defence Research and Development Canada.

[5] Chow, R., Lamb, M., Smith, E., Woodburn, M., Collins, M., Foley, J., Coates, S., and Saltel, B. (2015). Preliminary crew analysis for Canadian Surface Combatant (CSC): Phase 1 crew generation. (DRDC-RDDC-2015-L407) Defence Research and Development Canada.

[6] Peng, H., Bouak, F., Wang, W., Lamb, M., and Chow, R. (2017). Development of Sleep Prediction Models for Crew Fatigue Analyses.

[7] Wang, W., Chow, R., Peng, H., Bouak, F., Lamb, M., and Ueno, K. (2017). Sailors' work/rest schedule and fatigue on a Canadian patrol frigate during exercise trident fury 2015. (DRDC-RDDC-2017-R048) Defence Research and Development Canada.

[8] International Maritime Organization (2001). Guidance on fatigue mitigation and management.

[9] Oldenburg, M., Hogan, B., and Jensen, H. J. (2013). Systematic review of maritime field studies about stress and strain in seafaring. *International Archives of Occupational and Environmental Health*, 86 (1), 1–15.

[10] Waterhouse, J., Fukuda, Y., and Morita, T. (2012). Daily rhythms of the sleep-wake cycle. *Journal of Physiological Anthropology*, 31 (1).

[11] Horne, J. A. and Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology* 4 (2), 97–110.

[12] Fisk, J. D., Ritvo, P. G., Ross, L., Haase, D. A., Marrie, T. J., and Schlech, W. F. (1994). Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. *Clinical Infectious Diseases*, 18 Suppl 1:, S79–83.

[13] Fleishman, E. A., Quaintance, M. K., and Broedling, L. A. (1984). *Taxonomies of human performance: The description of human tasks,* Orlando, FL Academic Press, Inc.

[14] Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophsiology*, 10, 431–6.

[15] Paul, M. A., Ebisuzaki, D., McHarg, J., Hursh, S. R., and Miller, J. C. (2012). An assessment of some watch schedule variants in Cdn Patrol Frigates: Op Nanook 2011. (DRDC Toronto TR 2012-078) Defence R&D Canada – Toronto.

[16] Paul, M. A., Love, R. J., Waggoner, L. B., Hursh, S. R., and Miller, J. C. (2016). An assessment of alternative 1-in-2 and 1-in-3 watch schedules on an RCN Patrol Frigate: Exercise Trident Fury 2015. (DRDC-RDDC-2016-R059) Defence Research and Development Canada.

[17] Pearson, P. G. and Byars, G. E. (1956). The development and validation of a checklist measuring subjective fatigue. USAF School of Aviation Medicine.

[18] Lamb, M., Wang, W., Versteeg, T., Chow, R., Townsend, J., Bisal, D., Saunders, D., Labbé, D., Peng, H., and Bouak, F. (2017). A study of crew work, rest, fatigue and performance on the X-Ship for SCORE validation: Data collection summary and lessons learned. (DRDC-RDDC-2017-D030) Defence Research and Development Canada.

[19] Vagias, W. M. (2006). Likert-type scale response anchors. Clemson International Institute for Tourism and Research Development, Department of Parks, Recreation and Tourism Management.

[20] Flachenecker, P., Kumpfel, T., Kallmann, B., Gottschalk, M., Grauer, O., Rieckmann, P., Trenkwalder, C., and Toyka, K. V. (2002). Fatigue in multiple sclerosis: a comparison of different rating scales and correlation to clinical parameters. *Multiple Sclerosis Journal*, 8, 523–526.

[21] Green, K. Y. (2009). A comparative analysis between the Navy standard workweek and the actual work/rest patterns of sailors aboard U.S. Navy frigates. Naval Postgraduate School: Monterey, CA.

[22] Roberts, W. S., Grech, M. R., and Cleary, A. (2014). Crew endurance at sea: An analysis of sleep- and work-hours across a deployment period. Defence Science and Technology Organisation.

[23] Haynes, L. E. (2007). A comparison between the Navy standard workweek and the actual work and rest patterns of U.S. Navy sailors. Naval Postgraduate School: Monterey, California.

[24] Mason, D. R. (2009). A comparative analysis between the Navy Standard Workweek and the work/rest patterns of sailors aboard U.S. Navy cruisers. Naval Postgraduate School: Monterey, California.

List of symbols/abbreviations/acronyms/initialisms

ANOVA	Analysis of Variance
AOPS	Arctic Offshore Patrol Ship
AUS	Australia
BSS	Baseline Sleep Survey
BWK	Bridge Watch Keeper
CBT	Combat department
CPF	Canadian Patrol Frigate
CSC	Canadian Surface Combatant
CSE	Combat Systems Engineering department
DECK	Deck department
DFM	DRDC Fatigue Model
DRDC	Defence Research and Development Canada
EXE	Executive department
HITL	Human-In-The-Loop
HMCS	Her Majesty's Canadian Ship
IMO	International Maritime Organization
JSS	Joint Support Ship
LOG	Logistics department
M	Mean
M&S	Modelling and Simulation
MEQ	Morningness-Eveningness Questionnaire
MFIS	Modified Fatigue Impact Scale
MSE	Marine Systems Engineering department
NSWW	Navy Standard Work Week
PORT	1-in-2 Port watch
RCN	Royal Canadian Navy
SCORE	Simulation for Crew Optimisation and Risk Evaluation
SD	Standard Deviation
SOL	Sleep Onset Latency
STBD	1-in-2 Starboard watch

TIB	Time-In-Bed
USN	United States Navy
WASO	Wake-up After Sleep Onset
WRS	Work Requirement Survey
X-Ship	Experimental Ship Programme

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or tasking agency, is entered in Section 8.) DRDC – Toronto Research Centre Defence Research and Development Canada 1133 Sheppard Avenue West P.O. Box 2000 Toronto, Ontario M3M 3B9 Canada	2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.) CAN UNCLASSIFIED 2b. CONTROLLED GOODS NON-CONTROLLED GOODS DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

An X-Ship Study on Crew Work and Rest Schedule and Fatigue aboard a Halifax-Class Patrol Frigate

4. AUTHORS (last name, followed by initials – ranks, titles, etc., not to be used)

Wang, W.; Lamb, M.; Chow, R.; Peng, H.; Bouak, F.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total references cited.)
March 2018	79	24

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter.)

Scientific Report

8. SPONSORING CENTRE (The name and address of the department project office or laboratory sponsoring the research and development.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9
Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. DRDC PUBLICATION NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2018-R023	

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Any limitations on further dissemination of the document, other than those imposed by security classification.)

12. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Sailors' work and rest schedules, and fatigue were studied onboard Her Majesty's Canadian Ship (HMCS) Montreal during a 10-day at-sea trial. We examined the crew's work hours and workload, sleep duration and quality, as well as daily fatigue and mood. The results indicated a significant increase of fatigue at the end of the trial, relative to the beginning of the trial. Significantly different patterns of results emerged when the data were analyzed by grouping participants according to their department, watch or rank. An elevated risk for fatigue was identified for watch keepers, particularly those on the bridge. Results from this study support the ongoing validation of a crewing analysis tool SCORE and provide a baseline for assessing crewing options in future X-Ship trials.

Les horaires de travail et de repos ainsi que la fatigue des marins ont été étudiés à bord du Navire canadien de Sa Majesté (NCSM) *Montréal* durant une période d'analyse en mer de dix jours. Nous avons examiné les heures et la charge de travail des membres d'équipage, la durée et la qualité de leur sommeil, ainsi que leur fatigue et leur humeur au quotidien. Les résultats ont révélé une augmentation notable de la fatigue à la fin de la période d'analyse, par rapport au début de la période. Des modèles considérablement différents se sont dégagés de l'analyse des données par groupe de participants en fonction de leur service, de leur quart de travail ou de leur grade. Il existe un risque élevé de fatigue pour le personnel de quart, particulièrement sur la passerelle. Les résultats de cette étude vont dans le sens de la validation actuelle de l'outil d'analyse d'affectation des équipages SCORE et constituent une assise pour évaluer les options en matière d'affectation des équipages lors des futurs essais du navire X.

13. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g., Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

work and rest schedule; fatigue; X-Ship