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

CAN UNCLASSIFIED

DRDC | RDDC
technologysciencetechnologie

Halifax Class Modernization Automation Impact on Crew (3) Validation Trial Plan—HMCS Montreal

Agata Lawrynczyk
CAE Canada

Prepared by:
CAE Inc., 1135 Innovation Drive, Ottawa, ON,
K2K 3G7, Canada
PSPC Contract Number: W7707-145734 - T12
Technical Authority: Geoffrey Ho
Contractor's date of publication: September 2018

Defence Research and Development Canada

Contract Report
DRDC-RDDC-2018-C187
November 2018

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

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

CAE

CAE Inc.

350 Legget Drive, Suite 200
Ottawa, ON, Canada K2K 2W7
Tel: 613-247-0342
Fax: 613-271-0963

HALIFAX CLASS MODERNIZATION
AUTOMATION IMPACT ON CREW (3)
VALIDATION TRIAL PLAN - HMCS MONTREAL

W7707-145734 TASK 12

FOR

DEFENCE RESEARCH AND DEVELOPMENT CANADA - TORONTO

1133 Sheppard Ave West, Toronto, ON, M3K 2C9

31 March 2 0 17

Document No. 6010-008

RESTRICTION ON USE, DUPLICATION, OR DISCLOSURE OF DATA

APPROVAL SHEET

Document No. 6010-008 Version 03

Document Name: HALIFAX Class Modernization
Automation Impact on Crew (3)
Validation Trial Plan - HMCS Montreal

Primary Author

Name **Dr. Geneviève Dubé for Agata Lawrynczyk**

Position Human Factors Consultant, Defence & Security, CAE Canada

Reviewer

Name **Dr. Geneviève Dubé**

Position Human Factors Consultant, Human Factors Group, Defence & Security, CAE Canada

Approval

Name **Dr. Geneviève Dubé for Tab Lamoureux**

Position Senior Consultant, Human Factors, Defence & Security, CAE Canada

REVISION HISTORY

Revision	Reason for Change	Origin Date
Version 01 DRAFT A	Initial document issued for comment.	10 March 2017
Version 01	Final document issued.	31 March 2017
Version 02	Review comments from the client	21 September 2018
Version 03	Update of the reference list following discussion with the client	14 November 2018

This document has been updated to Version 03 and changes are denoted by a black line in the left-hand margin.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND DEFINITIONS

AOI	Area of Interest
ARRO	Air Raid Reporting Officer
ASPO	Anti-Submarine Plot Operator
ASWC	Anti-Submarine Warfare Coordinator
BARS	Behaviorally Anchored Rating Scales
CFCD	Canadian Forces Combat Doctrine
CFFS(A)	Canadian Forces Fleet School Atlantic
CFJP	Canadian Forces Joint Publication
CFMWC	Canadian Forces Maritime Warfare Centre
CIO	Communication Intercept Operator
CMS	Combat Management System
COI	Contact of Interest
CSC	Canadian Surface Combatant
CSNI	Consolidated Secret Network Infrastructure
CSTC	Combat Systems Training Center
D2E	Detect-to-Engage
D Nav P&T	Directorate of Naval Personnel and Training
DF	Data Fusion
DND	Department of National Defence
DRDC	Defence Research & Development Canada
ESM	Electronic Support Measures
EW	Electronic Warfare
EWS	Electronic Warfare Supervisor
FCS	Fire Control Supervisor
HCM	Halifax Class Modernization
HF	Human Factors
ID	Identity
IMD	Information Management Director
ISA	Instantaneous Self-Assessment
km	Kilometer
KPE	Key Performance Event
LED	Light Emitting Diode
LMC	Lockheed Martin Canada

MARS	Maritime Surface and Sub-Surface
MARTI	Maritime Tactical Instructions
MFW	Multi Functions Workstation
MOE	Measure of Effectiveness
MON	HMCS Montreal
MOP	Measure of Performance
NASA TLX	NASA Task Load Index
NCIOP	Naval Combat Information Officer Operator
NESOP	Naval Electronic Support Operator
OOI	Object of Interest
Ops Room	Operations Room
ORO	Operations Room Officer
ORS	Operations Room Supervisor
OTT	Operational Team Training
QL	Qualification Level
RCN	Royal Canadian Navy
RMP	Recognised Maritime Picture
SA	Situation Awareness
SAC	Shipborne Air Controller
SAFSIM	Semi-Automated Forces Simulation
SCS	Sonar Control Supervisor
SEAWOLF	Synthetic Environment Advanced Warfare Operations and Leadership Facility
SME	Subject Matter Expert
SONAR ops	Sonar Operator
SWC	Surface Warfare Coordinator
TA	Task Analysis
TacPlot	Tactical Display
TE	Threat Evaluation
TN	Track Number
TS	Track Supervisor
UN	United Nations
UWW	Under Water Warfare
WA	Weapons Assignment
X-Ship	Experimental Ship

EXECUTIVE SUMMARY

This document describes a validation trial plan to investigate the research question:

> *Can a reduced operator complement permit the required objective level of performance for the combat department to be attained during all missions and mission phases on the post-HCM HALIFAX Class frigate?*

The validation trial plan describes the hypotheses that will be considered to support the investigation of the research question as well as the Measures of Performance (MOPs) and Measures of Effectiveness (MOEs) that will permit acceptance or rejection of the hypotheses. To support the generation of the MOPs and MOEs a combination of objective and subjective data will be collected during the trial.

The validation trial plan describes the facilities required to run the trial. In particular, a ship with a full crew (i.e., two watches in the combat department) is required that will permit a defined scenario to be presented to a mature team of operators in a dynamic and responsive manner. The plan also describes the type of operators required with respect to their qualifications, their base of operations and the number of teams.

It is proposed that the operations room will be manned by the minimum number of operators necessary, currently believed to be seven (not including any operators required to use legacy systems). According to performance and subjective report, additional operators will be added selectively according to demand to ensure that performance remains acceptable. The operators will be presented with simulation scenarios through the training mode of the combat management system, including those that represent the most challenging situations that the Royal Canadian Navy expects its crews to encounter. This trial will evaluate the number of operators required to defeat (i.e. perform at the required level) the most challenging scenarios and determine whether this level of sustained activity can be maintained when faced with the additional demands of being at sea.

The trial is tentatively scheduled for autumn 2019. A schedule of activities and further work that must be completed prior to 2019 is provided. Among these are activities in 2017 intended to confirm that scenarios can be run in the identified simulation facilities and that objective data can be obtained, followed by an experiment using the Synthetic Environment Advanced Warfare Operations and Leadership Facility (SEAWOLF) simulation facilities in 2018. Further, the plan includes the development of detailed data analysis procedures that combine scenario ground truth and system-recorded data from the simulation to result in the MOPs.

This validation trial plan draws heavily from work already carried out by the Department of National Defence (DND) and the Royal Canadian Navy (RCN), specifically with respect to the scenario development and the MOPs and MOEs. This achieves three objectives: (a) by relying on previous work, it avoids re-creating work that has already been done; (b) it ensures that the scenario realistically represents the greatest demand that a RCN frigate can reasonably be

expected to successfully manage; and (c) it encourages the greatest acceptance of the eventual results of this trial.

The data arising from the trial will be compared with the RCN's existing standards of combat department performance, as documented in Canadian Forces Combat Doctrine (CFCD)106.

The expectation is that the follow-on activities described in this document will be continued in a subsequent contract.

1 INTRODUCTION

This document is one of two deliverables for Project W7707-145734/001/HAL Task 12- *HALIFAX Class Modernization – Automation Impact on Crew (3)*. The report describes the initial plan for an upcoming validation trial using the HALIFAX Class Frigate Her Majesty's Canadian Ship (HMCS) MONTREAL (MON) to investigate the possibility of augmenting the Operations Room (Ops room) crew size. This document is a living document and refinement continues as some activities described in this document are carried out. This document provides an overall framework of the work that will be performed over the next year (e.g. ground truth analysis, pilot studies and tabletop exercises). The document can also be used as a starting point for drafting a detailed ethics protocol, a maritime evaluation, and reporting of the validation trial. This trial is tentatively scheduled for 2019 subject to availability of the ship and her crew.

The present document is primarily meant to be an internal project document for planning purposes and draws largely from Dube, Lawrynczyk & Lamoureux (2018a) which documents the initial plan for conducting a simulation experiment onshore using a similar protocol. Indeed, the Defense Research and Development Canada (DRDC) project plan was to develop a simulation-based study that could then be also verified at sea onboard HMCS MON. As such, much of the material in this document overlaps with (and is identical to) Dube et al., (2018a).

Background

The twelve HALIFAX-class frigates are the foundational ships of the Royal Canadian Navy (RCN). A refit, the HALIFAX Class Modernization (HCM) programme, has recently been completed to advance the ship's world-class capabilities. HCM includes a new communications system, new and enhanced sensors, advanced armament, and a new combat management system.

The development of the Combat Management System (CMS) 330 introduced a variety of automation to the HALIFAX Class frigate combat department[1]. In particular, significant automation was introduced to support data fusion (DF), sensor and weapons configuration (a.k.a. Doctrine[2]), threat evaluation (TE) and weapons assignment (WA). This automation supports the overall goals of the combat department with respect to picture compilation, threat evaluation and weapons assignment (these latter two collectively referred to as TEWA). The automation may permit the combat department to maintain the required objective level of combat performance with a reduced complement of combat operators. Defence Research & Development Canada (DRDC) has been requested to support the RCN to carry out scientific investigations concerning this possibility.

[1] The combat department carries out its work primarily in the Operations Room. The Operations Room is referred to as the Ops Room in this report and is used to indicate the full complement of working operators at a given time (i.e., not those who are resting or eating or on other duties). The combat department will be used to refer to the complete complement of combat operators on the ship across both watches and is independent of whether or not they are on duty.

[2] Doctrine is the name of the function within CMS-330 that allows the crew to select pre-determined settings (based on geographic areas) for sensors and weapons.

As part of this research programme, DRDC plans to conduct some large-scale experiments to determine the viability of a reduced crew complement in the combat department. The first of these experiments will be carried out in a high-fidelity simulator. The second of these large-scale experiments will build on the findings from the first by conducting what is probably more properly considered a validation trial using a ship at sea with a full crew. HMCS MONTREAL has been designated by the RCN as the preferred ship with which to conduct experiments and trials to investigate new capabilities and processes[3]. Thus, it is expected that MON will support the validation trial concerning the complement of operators in the combat department.

The current contract follows from previous contracts (Task 1: Lawrynczyk, Lamoureux, & Dube, 2015; Task 7: Dubé & Lamoureux, 2016a). These contracts (and anticipated follow-on contracts) are focused on the development and running of experiments in simulation facilities and aboard HALIFAX Class frigates at sea. This document only describes the plan for carrying out the trial aboard MON. A previous report (Dubé & Lamoureux, 2017) describes the simulation-based experiment on which the MON trial will build.

This validation trial seeks to investigate the viability of reducing the number of combat operators required to attain a desired objective level of combat performance. In particular, this trial will seek to validate the results obtained from qualified HALIFAX Class combat departments in simulation experiments by determining whether the desired level of objective performance can be sustained by a reduced crew complement in the presence of the additional demands imposed on the combat department while at sea. The results of this project will support the ongoing improvements to the frigates as well as future decisions for the design and development of the Canadian Surface Combatant (CSC), which is in the process of making design decisions regarding the makeup of its crew relative to the onboard automation and technologies.

1.1 Objective and Research Question

The objective of the current project is to complete detailed planning for two large-scale experiments; one in the Synthetic Environment Advanced Warfare Operations and Leadership Facility (SEAWOLF) (or other suitable simulation facility) and one aboard MON. This document focuses on the validation trial aboard the ship. The trial on X-Ship builds upon the outputs of the experiment carried out with SEAWOLF. This trial is currently in the planning stages. Therefore, the present plan is subject to change.

The X-Ship validation trial intends to answer the following research question:

> *Can a reduced operator complement permit the required objective level of performance for the combat department to be attained during all missions and mission phases on the post-HCM HALIFAX Class frigate?*

The trial aboard MON adds one specific component to the research question above: assuming simulation experiments have shown that a reduced crew can achieve the required performance

[3] Accordingly, MON is also referred to as the X-Ship (short for 'Experimental Ship').

standard, can they maintain it for the typical duration spent at sea between port visits. This period is 10 – 14 days. The specific objective of this report is to describe the detailed protocol that will be used in X-Ship in 2019 to answer this research question.

1.2 High Level Approach to Validation Trial

To satisfactorily investigate the research question presented above the following assumptions provided high-level guidance to the development of the validation trial approach:

1. Changes investigated in this study are relative to the current trades organization in the combat operations room (Ops room). This includes the number of operators, the warfare domains and the different combat operator trades.

2. The current implementation of the CMS330 at the time of the trial is assumed. It is further assumed that automation will only improve subsequent to the trial, thus ensuring conclusions from the trial remain valid.

3. The current CMS330 permits the presentation of scenarios that have been developed to run on Semi-Automated Forces Simulation (SAFSIM) – a simulation environment.

4. The trial will be a within groups design with counterbalancing applied between watches with respect to what order they see scenarios.

5. The combat Ops Room will be manned by the minimum number of operators necessary. This number is presumed to be seven: Track Supervisor, Ops Room Supervisor, Shipborne Air Controller, Surface Warfare Controller, Anti-Submarine Warfare Controller, Ops Room Officer and Information Warfare Director. According to performance and subjective report, additional operators will be added selectively according to demand to ensure that performance remains acceptable. This approach is described in more detail in Section 3.1. Performance and subjective report will be monitored constantly to ensure that there is never any threat to the safety of the ship.

6. To answer the question about number of combat operators it is not necessary to consider the means by which performance is achieved, only that the required level of performance is achieved reliably. Consequently, potential metrics concerning operator processes do not form part of the measurement and analysis plan, only performance outcomes are included.

7. The objective level of performance required by the RCN is represented by the task demands imposed by the scenario. If the combat department successfully deals with a scenario judged by the RCN to impose the greatest expected demand that a frigate can reasonably be expected to deal with, the combat department are judged to have achieved the objective level of performance required.

8. It is necessary to define what 'success' means. Where possible, objective measures will be used to determine success or failure. The RCN have developed objective performance standards for different activities in the Ops Room. These standards are documented in Canadian Forces Combat Doctrine (CFCD) 106 and associated Maritime Tactical

Instructions (MARTI). These documents will be used to set specific success criteria for the combat teams.

9. Where possible, scenarios that have already been created by the RCN will be used as a basis or in their entirety for this trial.

10. The crew configuration and mission intensity are the only independent variables that will be manipulated during the trial.

11. Dependent variables will be measured to answer the research question. Where possible, objective measures will be given precedence in order to manage the workload associated with data collection and to ensure that subjective biases do not undermine the validity and generalisability of the conclusions.

12. Objective data can be recorded from the CMS330.

13. Where possible, metrics that have already been developed by the RCN will be used as a basis or in their entirety for this trial.

14. The RCN and the Department of National Defence (DND) will grant permission to contractors to use:

    a. their master scenario developed by CFFS(A);

    b. CMS330 data logs;

    c. CFCD106 to derive performance standards; and

    d. CMS330 simulation facilities.

15. The conclusions will be limited to the combat department. No conclusions will be drawn relating to other departments (e.g., Deck, Engineering).

16. This trial can only result in conclusions for operators using CMS330 workstations. Operators using legacy systems (e.g., towed array and hull-mounted sonars) cannot be reduced at this time because their systems are not integrated into CMS330 and they are required to operate those systems and feed data into the CMS330 in a manual fashion.

17. The ship to be used for this trial is of sufficient complexity that full control over all possible sources of error may not be possible.

18. The trial will last a minimum of 10 days and a maximum of 14 days.

Note that this document attempts to build upon work done by various groups within the DND. This saves this project from having to reproduce work that has already been carried out, thus realising savings in effort. It also helps to ensure that the trial is based on the latest thinking of the RCN regarding threats, capabilities and level of performance required from a HALIFAX Class combat department. In some sections of this report, text is reproduced from other reports and technical documents, including the SEAWOLF Experimental Plan (Dubé & Lamoureux, 2017). To rephrase this text in the authors' own words would not achieve any benefit to the current work. This text is referenced and the original author credited with the effort.

1.3 This Document

This document is an interim report and is organized into the following sections:

- Section 1 – Introduction: Describes the background to this work and the objectives for this document.
- Section 2 – Hypotheses: Presents the hypotheses to be tested as well as the dependant and independent variables of the study.
- Section 3 – Validation Trial Design: Describes the trial settings, the scenarios, the schedule, the participants, the trial team and the procedure of this validation trial.
- Section 4 – Measures: Describes the objective and subjective measures of performance that will be used in this study.
- Section 5 – Analysis: Presents the analysis that will need to be performed.
- Section 6 – Conclusion: Summarises the document and outlines the next steps.

Additionally, this document includes a list of acronyms and abbreviations, references, and appendices presenting the detailed list of hypotheses, the incremental crew configurations and background concerning the Detect-To-Engage (D2E) cycle.

2 HYPOTHESES

This section presents the hypotheses that will be tested during the X-Ship validation trial. The main hypothesis (from which the sub-hypotheses were derived) was developed during the preceding contract, reported in Dubé & Lamoureux (2016a). One of the outcomes of the preceding contract was a list of potential research questions that were presented to DRDC Toronto and their RCN client from Directorate of Naval Personnel and Training (D Nav P&T). Following discussion, the DND stakeholders (i.e., DRDC and D Nav P&T) chose to focus on the following research question:

> *Can a reduced operator complement permit the required objective level of performance for the combat department to be attained during all missions and mission phases on the post-HCM HALIFAX Class frigate?*

The research question led to the following main hypothesis:

> *H1: A reduced combat crew can effectively meet the objective operational performance requirement for a HALIFAX Class Frigate.*

As noted in the introduction to this report, the validation trial aboard MON seeks to validate the results of the experiment performed in SEAWOLF and, assuming the combat crew can meet the performance requirement, determine whether they can maintain this level of performance for up to two weeks. The maintenance of performance over a prolonged period during which personnel face additional challenges (e.g., rest, nutrition, hygiene, secondary duties) is a key driver for using MON and evaluation of this trial will specifically considered whether performance degrades over the period of the voyage. This measurement focus is congruent with the existing research question and hypotheses.

Table 2-1 presents the dependent and independent variables for this study.

Table 2-1: List of Dependent and Independent Variables

Independent Variable	Dependent Variables
Crew configuration (crewing levels)	Performance (objective and subjective measures)
Mission/scenario intensity	Workload
	Subjective Data

H_1 was decomposed into sub-hypotheses aimed at covering every aspect of the task by considering operator trades, high-level operator goals and performance measurement outcomes. These broad categories informed the development of sub-hypotheses, listed in Table 2-2. The full list of detailed sub-sub-hypotheses is presented in Appendix A.

Table 2-2: List of Sub-Hypotheses (Level 1) (Dubé, Lawrynczyk, & Lamoureux, 2016)

Hypotheses Category	Sub-Hypotheses
Naval Electronic Support Operator (NESOP)-based	H(NESOP)1.1: The Surface Warfare Coordinator (SWC) can meet the objective operational performance requirement for sensor employment (Electronic Support Measures (ESM) / Communication Intercept Operator (CIO)) with no assistance (L1)
	H(NESOP)2.1: The SWC can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with no assistance (L1)
Naval Combat Information Officer Operator (NCIOP)-based	H(NCIOP)1.1: The TS can meet the objective operational performance requirement for radar utilisation with no assistance (L1)
	H(NCIOP)3: The SAC can effectively manage the helicopter resources (L1)
	H(NCIOP)4.1: The ORS and IWD can effectively facilitate the timely and accurate flow of information around the Ops Room with no assistance (L1)
SONAR-based	H(UW)1: The Anti-Submarine Warfare Coordinator (ASWC) can meet the objective operational performance requirement for Under Water Warfare (UWW) (sensors and weapons) with no assistance from subordinates (L1)
Director-based	H(Dir)1.1: The SWC has the information required to support effective planning and decision making (L1)
	H(Dir)2.1: The SWC has the means to act upon their decision (L1)
Identification-based	H(I)1.1: Sensors are used effectively to detect contacts in the area of responsibility
	H(I)2.1: Sensors are used effectively to detect contacts in the area of responsibility
	H(I)3.1: Sensors are used effectively to detect contacts in the area of responsibility
Countermeasure-based	H(C)1: Countermeasures are used in a timely, effective and accurate manner to defeat an air threat
	H(C)2: Countermeasures are used in a timely, effective and accurate manner to defeat a surface threat
	H(C)3: Countermeasures are used in a timely, effective and accurate manner to defeat a sub-surface threat.
Time-based	H(T)1: Observed performance times for identified tasks/task sequences will meet or improve upon RCN standards
	H(T)2: The time required to perform tasks, respond to contacts, respond to threats will not increase with any level of intensity or complexity
	H(T)3: The time required to perform tasks, respond to contacts, respond to threats will not increase over the duration of a deployment
Error-based	H(E)1: The observed error rate does not increase with any level of intensity or complexity
	H(E)2: The observed error rate does not increase over the duration of a deployment

Hypotheses Category	Sub-Hypotheses
Workload-based	H(WL)1: With the crew configuration that permitted the objective level of performance to be attained, workload was not judged to be excessive
	H(WL)2: The level of workload associated with the objective level of performance can be maintained for the full duration of a deployment

The sub-hypotheses related to the trades (NESOP-based, NCIOP-based, SONAR-based and Director-based hypotheses) were based on the incremental crewing levels presented in the interim report (Dubé, Lawrynczyk, & Lamoureux, 2016). The incremental crewing levels support the trial objectives by permitting the incremental addition of personnel to the Ops Room in reaction to the level of workload, where higher workload may require more operators to be present in the Ops Room. Appendix B presents the details of the crewing levels. Given that an incremental crew complement is central to the validation trial, many of the hypotheses are founded on the presence or absence of specific crew roles.

Although there are many sub-hypotheses, in practice most will be answered by the same data. What the sub-hypotheses do permit, however, is the consideration of the fundamental hypothesis, that a reduced combat crew can effectively meet the objective operational performance requirements for a HALIFAX Class frigate, from many different perspectives in order that the conclusions from the experimental programme are as valid and reliable as possible.

3 VALIDATION TRIAL DESIGN

This section presents the details of the trial design: the trial approach (Section 3.1), the trial environment (Section 3.2), the scenarios (Section 3.3), the schedule (Section 3.4), the participants (Section 3.5), and the trial team (Section 3.6).

This validation trial will require a full crew (i.e., two watches in the combat department). The combat department will be asked to operate with a minimal manning (seven operators). This reduced operator complement will carry out operations as normal but will occasionally be presented with scenarios of varying levels of intensity over the two-week period. The performance of the reduced operator complement will be monitored and, if necessary, additional operators will be added selectively to maintain the required level of performance (this approach is described in greater detail in Section 3.1). The trial team will measure subjective (e.g., workload, ratings of performance) and objective indices to determine the success of the crew in maintaining the required level of performance. Periods of low/no intensity (i.e. no scenario being presented) will serve as a baseline measure.

The specific procedure to be adopted while onboard MON is continuing to be refined, in collaboration with RCN and MON planning staffs, as well as following such activities as the ground truth analysis, the pilot study and tabletop exercises.

3.1 Validation Trial Approach

The general approach to the validation trial is to begin with a minimum crew complement in the Ops Room. It is expected that this minimum complement will be sufficient to maintain the required level of performance during routine picture compilation and threat evaluation (i.e. primarily civilian shipping, commercial airliners, cooperative contacts with no threatening intent). MON may also be tasked with a mission, for instance, a fisheries patrol. This may increase the intensity and demand upon the operators without the addition of a Shipboard Embedded Team Trainer (SETT) scenario. It is still expected, however, that the minimum crew complement will be sufficient to maintain the required level of performance.

According to a pre-determined schedule, identified workstations will be put into training mode and presented with a scenario that increases the work intensity. These scenarios are listed in Section 3.3.1. Up to eight Multi-Function Workstations (MFWs) can be put into training mode at once. Depending upon the crew complement in the Ops Room, this may require the addition of operators to continue the operational picture compilation and navigational safety task[4].

The scenarios to which operators will be exposed represent the most challenging situations that the RCN expects its crews to encounter. This trial will evaluate the number of operators required to defeat (i.e. perform at the required level) the most challenging scenarios and determine

[4] For instance, if the minimum crew of six is able to maintain the required level of performance, then all six operators can be put into training mode. This leaves no one in the Ops Room to carry out operational picture compilation and safety-related surveillance. Therefore, additional operator will need to be called into the Ops Room to open additional MFWs to carry out these live tasks while the 'onwatch' crew participate in the SETT scenario.

whether this level of sustained activity can be maintained when faced with the additional demands of being at sea. Operator performance will be monitored at all times, in particular via subjective evaluation and self-report. If it is deemed that an operator is having difficulty maintaining the required level of performance, an additional operator will be added to assist. This will continue, if-and-as required, to the point that the current crew complement is reached. If this occurs, the overall hypothesis will be rejected.

In order to permit the addition of operator to a SETT scenario, every scenario will have eight MFWs dedicated to it, even though, initially, it is not expected that all MFWs in training mode will have crew assigned to them. These MFWs will be identified during planning for each scenario.

Reducing the crew complement will necessarily require reallocation of the work done in the Ops Room. The smallest crew complement is seven operators: Track Supervisor, Ops Room Supervisor, Shipborne Air Controller, Surface Warfare Controller, Anti-Submarine Warfare Controller, Information Warfare Director and Ops Room Officer. This configuration therefore requires the following changes from the existing duties of each of these roles:

- Track Supervisor: responsible for all surface tracks, including all sensors (radar, EW) that comprise those tracks.
- Shipborne Air Controller: responsible for all air and sub-surface tracks, including all sensors (radar, EW) that comprise those tracks.
- Ops Room Supervisor: no change.
- Surface Warfare Controller: responsible for TEWA in air and surface domains; responsible to assist Track Supervisor and Shipborne Air Controller to carry out air and surface picture compilation using radar and EW sensors.
- Anti-Submarine Warfare Controller: responsible for TEWA in subsurface domain; responsible to assist Shipborne Air Controller to carry out sub-surface picture compilation using radar and EW sensors; responsible to compile picture from acoustic sources (during subsurface operations sonar operators will be present using legacy systems).
- Information Warfare Director: responsible to support Ops Room Officer in maintaining flow of information into and out of Ops Room; responsible to assist in picture compilation if so directed by SWC via ORO.
- Ops Room Officer: responsible to support SWC and ASWC carry out TEWA duties when they are assisting in picture compilation.

These additional responsibilities/lines of support are depicted in red in Figure 3-1. Also depicted in this is the existing information sharing responsibility of the ORS, and the tertiary support responsibilities of the IWD (in purple).

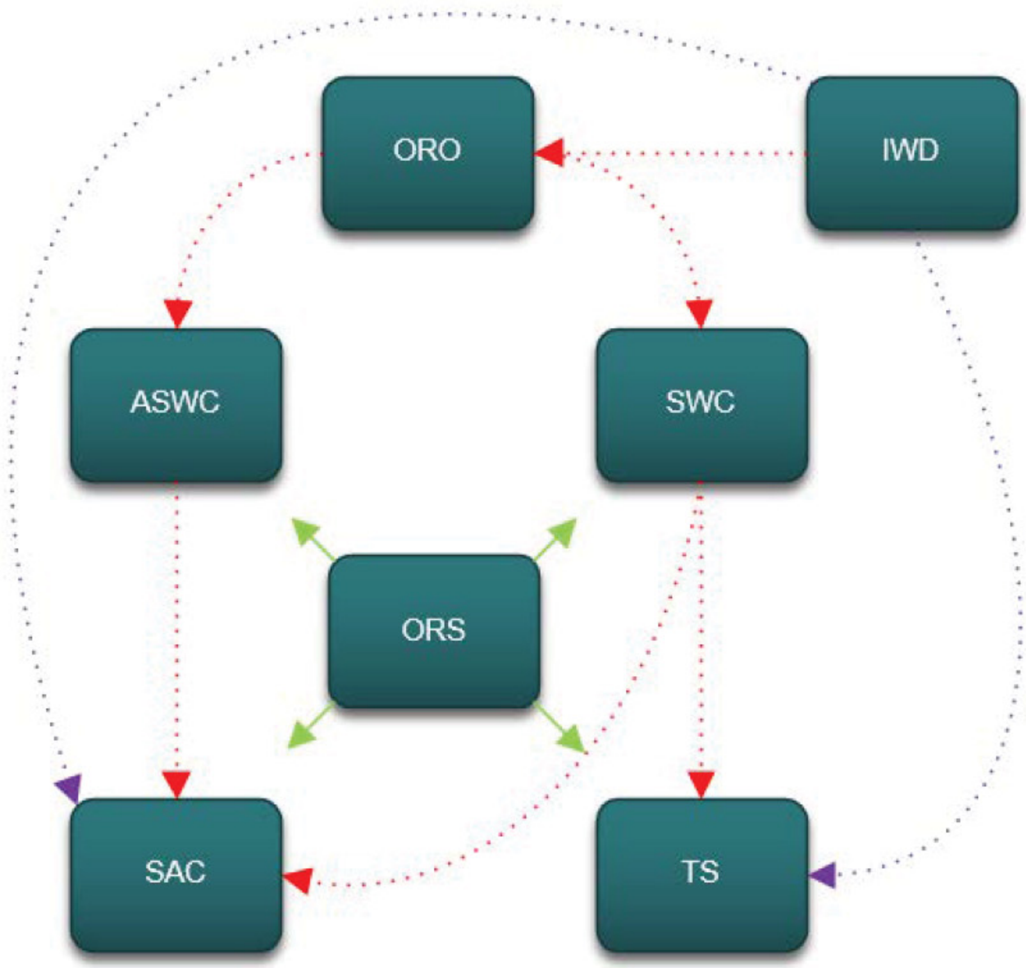


Figure 3-1: Lines of Support in Reduced Crew Configuration

Implicit in this revised concept of operator responsibilities is that the director will be responsible in the first instance to support their picture compilation resource(s). Only when the director requests additional personnel will an additional operator be added to the crew complement for that sensor grouping, in accordance with the incremental plan for that grouping.

All scenarios to be presented during this validation trial represent the objective level of performance the RCN expects its HALIFAX Class combat teams to manage successfully. The trial scenarios will differ according to the specific mix of contacts and threats that contribute to intensity and complexity, but the intensity and complexity that will be presented will be constant. Lower intensity and demand will be represented by normal operations, when the ship is performing its day-to-day operations.

Each scenario will start with a reduced crew (Crewing Level 1, seven operators, described above). If the reduced crew is able to manage the scenario with a satisfactory level of performance (based on self-report, workload and SME assessment; objective performance data will necessarily be analysed later and will be used to support or refute the crewing decisions of the team) there will be no increment of the crewing level. However, if the reduced crew is unable to maintain a satisfactory level of performance (again, based on self-report, workload and SME assessment), the crewing level will be incremented in accordance with the incremental crewing plan (Appendix B) by adding an operator to support the operator grouping (EW, Radar or Sonar) that is not maintaining adequate levels of performance. Additional operators may be added to more than one grouping at the same time.

Since the director is responsible to support picture compilation it is most likely that they will be the ones to request additional support for a particular operator grouping (i.e. EW, Radar, or Sonar). This self report will be shadowed and validated in two ways: SMEs will observe performance and suggest adding personnel if, in their judgment, this is required; and workload will be monitored and if the self-reported level of workload (on the Instantaneous Self-Assessment) is four or more for four consecutive cycles, an additional operator will be added to the crew complement for that sensor grouping, in accordance with the incremental plan for that grouping (see Appendix B). Note that the specific behavioural and performance indicators that the crew complement needs to be incremented have not yet been determined. This activity will occur over the course of 2017/2018.

If the need for crewing increment comes from an operator (i.e. an operator presents a level of workload of four or more for four consecutive cycles) rather than a director, then an operator from the same sensor grouping will be added. However, in the case where the operator experiencing a high workload is responsible for sensors from two different grouping, input from an SME will be required to identify what sensor grouping is problematic and needs to be incremented. For example, if the Track Supervisor reports a workload of four for four consecutive cycles, a SME will assess the situation and recommend that either an EW or a Radar operator be added, in accordance with the incremental plan for that sensor grouping.

The CMS330 implementation is such that there is little additional operator support if directors exhibit high workload. The ASWC provides little assistance to the picture compilation process because most underwater sensors are legacy systems. The SWC carries a great deal of responsibility for picture compilation in the first crew level (mostly for EW sensors), but this is removed from him at the second crewing level. After these points, the directors are the consumers of the picture and may demand additional operators to assist in developing a good picture, but the directors will have no assistance in their primary tasks of threat evaluation and weapons assignment, beyond that which can be provided by the Operations Room Operator (ORO).

The incremental crewing plan (Appendix B) indicates what operators are responsible for what sensors and/or weapons. When a new operator is added to the combat team, they will not be integrated slowly or deliberately. They will receive an in-brief from the operator who was previously responsible for their sensors and/or weapons, directing their attention to particular contacts, tasks or objectives. They will then need to rapidly build their SA and contribute to the actions of the team. This in-brief may take place at the MFW of the existing operator, or over the radio net once the incoming operator has sat at his MFW. The specific handover procedure will be determined in due course through the table-top exercise, the pilot test and discussions with SMEs.

Once a crew member has been added, he or she will remain with the crew for the rest of the scenario. There can be more than one increment per scenario.

3.2 Trial Environment

This section describes the environment and the system that will be used for this validation trial. Section 3.2.1 presents CMS330, the system on which the trial is predicated, Section 3.2.2 describes the workstation that will be used, and Section 3.2.5 the presents X-Ship, the HALIFAX Class frigate that will be used for the validation trial.

3.2.1 CMS330

The CMS330 is the system via which combat operators on the HALIFAX Class achieve their objectives. Its advanced automation is the reason reductions in crew complement may be possible in the Ops Room. The CMS330 is a highly integrated CMS comprising of a broad array of sensors, weapons and other information sources, integrated at the combat operator's workstation. With the new CMS330, the system assembles and fuses all of the sensor data to deliver a processed operational picture that supports Situation Awareness (SA) and information processing. CMS330 can fully automate sensor settings, threat evaluation, engagement planning and engagement itself (Baker & Banbury, 2011). This integration (contrasted with the old way of doing it with separate control units and, in some cases, display units), is a key enabler (alongside automation) for crew reduction. The present study will use the CMS330 to run simulations of operational missions. The functionalities that will be used are identical to those used by the crew onboard ships.

Four components of the CMS330 are of interest for this project: 1) DF; 2) DOCTRINE; 3) TE; and 4) WA that are presented on the TacPlot Display and the Auxiliary Display. Figure 3-2 illustrates a typical view of the TacPlot and Auxiliary Displays (other panels can be displayed depending on the task being performed). These components support a warship's goals of picture compilation, defensive actions and delivery of effects (both kinetic and non-kinetic).

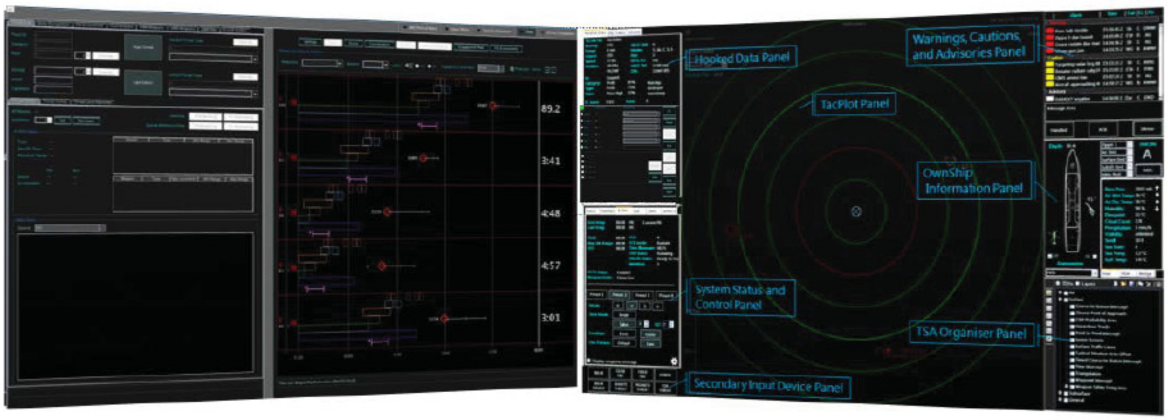


Figure 3-2: Example of the Auxiliary Display (left) and the TacPlot Display (right) from Baker and Banbury, 2011

3.2.2 Multi-Function Workstation

The trial will be run on MFWs (Figure 3-3) only; not on legacy workstations. The MFW is comprised of three displays: 1) Isolated Display on the right hand of the console, 2) Tactical Display (TacPlot) in the center, and 3) Auxiliary Display on the left hand side of the console. The CMS330 is presented via the TacPlot and Auxiliary display. The majority of picture compilation and TEWA work carried out by the combat department is carried out on the TacPlot and the Auxiliary Display. The Isolated Display is isolated from the CMS330 system and it is used only to access the Consolidated Secret Network Infrastructure (CSNI). It is not part of the automation; therefore, it will not be discussed further here.

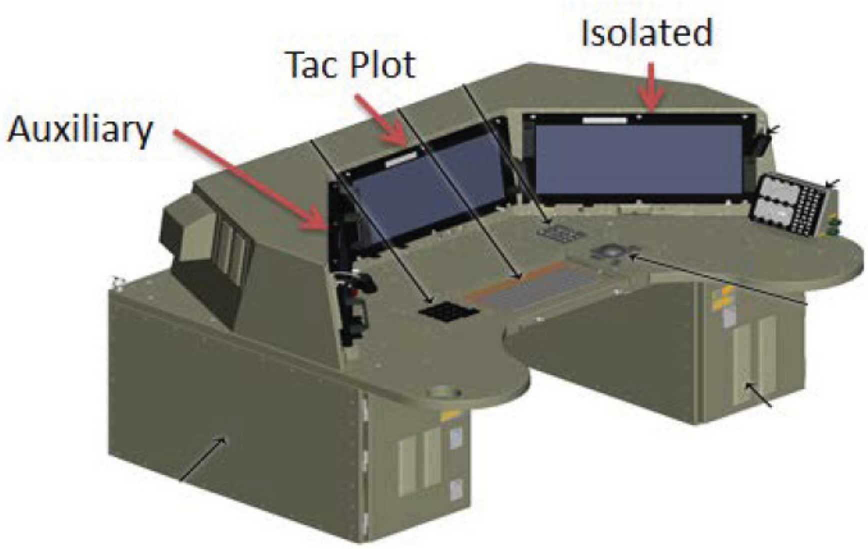


Figure 3-3: MFW Console (Figure from Baker and Banbury, 2011)

The MFW is standard for all operators using the standard CMS330 system (there are also special CMS330 implementations for the Fire Control Operators and the Sonobuoy Processing Supervisor), and any MFW can be configured for any of these roles, no matter where it is located in the Ops Room. In addition, an MFW allows for more than one user configuration at a time (e.g., configuration for Air Raid Reporting Officer (ARRO), configuration for Track Supervisor (TS)) per console. For example, both the ARRO and the Anti-Submarine Plot Operator (ASPO) configuration could be running on the same MFW. Concretely, this means that, in early stages (Crewing levels 1, 2 and 3), operators from different trades could be sitting beside each other in the front row, with subsequent reconfiguration of seating arrangements as additional operators are added and the respective trades choose to sit in groups.

Up to eight MFWs can be configured to run in SETT mode simultaneously. When SETT mode is enabled, a green bar appears across the TacPlot and Auxiliary screens with the word ‘training.’

3.2.3 Shipboard Embedded Team Trainer

The Shipboard Embedded Team Trainer (SETT) is the simulation facility available within the CMS330. SETT is the training mode that the CMS330 can be put into (as opposed to operational mode) and permits combat operators to train, individually and in teams, while onboard the ship. The Multi-Function Workstations (MFWs) indicate that they are in training mode by the presence of a prominent green border across the top of the tactical and auxiliary displays and between the picture compilation/threat evaluation panel and the settings panels. As SETT can only accommodate eight simultaneous users, the trial team will have to determine which MFWs should be placed in training mode to match the most pressing demands of the scenario.

SETT has the same capabilities as SEAWOLF does ashore, including objective data collection (non-audio/video and audio/video), and uses the same scenario format (i.e., Semi-Automated Forces Simulation (SAFSIM)). SETT only has the capacity for one interactor (i.e., a person who controls other entities in the scenario and may communicate with operators), however, making it difficult to present a scenario to trainees with many manoeuvring parts (e.g., aircraft flying evasively, small boats moving quickly and erratically). Also, SETT includes the countermeasure and sensor simulations that are present as separate simulations in SEAWOLF ashore. This means that the countermeasure and sensor simulations exhibit less fidelity in SETT than the operators would see in the simulator ashore. At this point in time, it is unclear whether this will be apparent to operators and what effect it will have on the execution of the scenario.

As illustrated in Figure 3-4, SETT includes a ‘game’ component that controls entities in the simulation with respect to their locations, course, speed, sensors and weapons. The SETT game component stimulates the sensors component that injects information into the CMS330, presented to operators via the MFW displays. The operators then process the data being presented to them and make decisions about how to react (e.g., give course and speed instructions to their ownship, contact the entities in the simulation, employ countermeasures to interact with the entities). The CMS330 then sends feedback to the scenario through simulated weapons component of SETT. The CMS330 is therefore not aware that it is acting upon simulated data. The scenario is executed and rendered dynamically responsive to operator actions by an interactor (i.e., role player), who simulates hostile entities, allies, etc. The operators may have an influence on the scenario through their interactions with the interactor, which they conduct through verbal communication. The system also allows the trial staff to stop the scenarios whenever they need to in order to gather external data.

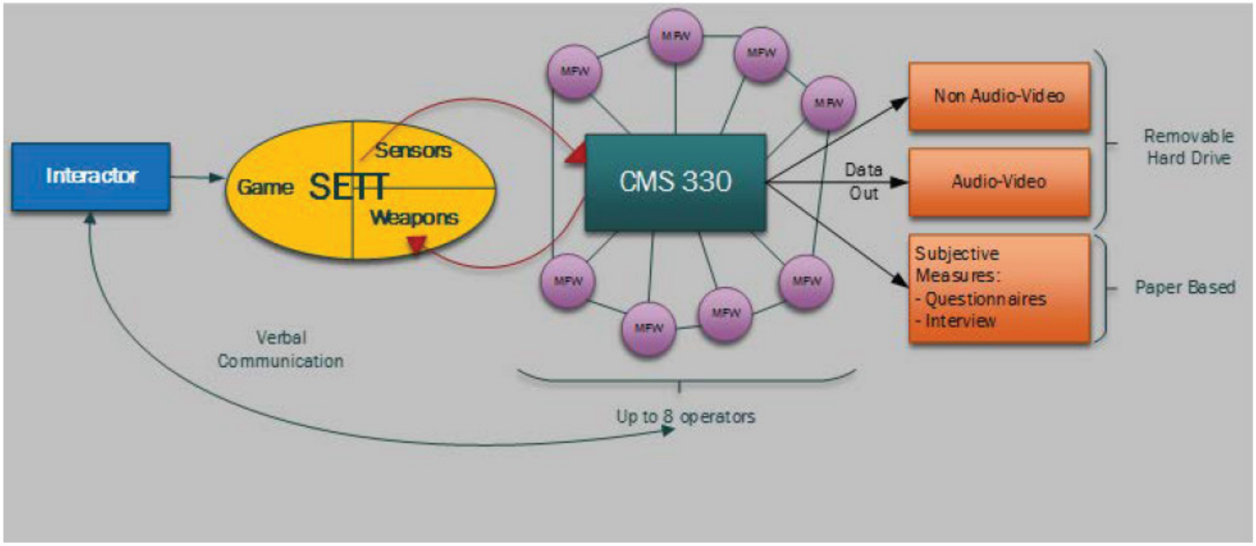


Figure 3-4: System Requirements

3.2.4 Data Recording

The system must support recording of audio-video and non audio-video data. Audio-video data records a replayable video of operator actions on the system display and sound that passes through the communication system. Audio-video data must be recorded via the workstation, rather than via an external camera. Non audio-video data records the internal system messages (i.e. the software commands) representing user input and system responses. Non audio-video data must be extracted directly from the system's log (see Section 4.2). Non audio-video data permit detailed and unambiguous analysis of user interactions with the CMS330 and associated timings. The system will be required to capture and record screen content and users' actions on a time base accurate to 1/10 seconds or less (Matthews at al., 2002). The audio-video and the non audio-video data must be copied from the CMS330 system to a removable hard drive after the simulation. Non-audio-video data should be in a format that can be read on a standard Window-based computer.

When replaying the audio-video data, the trial team will need to be able to fast forward, rewind, slow down and halt a scenario recording as necessary. This will facilitate Subject Matter Expert (SME) ratings of performance and accuracy-based measures that won't need to be captured live.

3.2.5 X-Ship

The validation trial will take place on MON, which is dedicated to research & development between January 2017 and December 2019. The objective of this ship is to serve as a platform to conduct experiments and observations that will contribute to requirements and designs for future ship classes, tactics, techniques and procedures. The X-Ship offers the advantage of an available crew facing real operating demands. The layout of the Ops Rooms is identical to all other frigate Ops Rooms, which increases the face validity of the results. A full crew complement will conduct all ship activities so this trial will be realistic to the extent that any conclusions drawn from it can reasonably be applied to the realities of the RCN.

Figure 3-5: HMCS Montreal: The X-Ship (www.combatcamera.forces.gc.ca)

As indicated in the previous section, the simulation devices can present the scenario only to those operators using the MFW; some Electronic Warfare (EW) and the Sonar operator positions are not integrated in CMS330 because they use legacy systems or systems. The CMS330 does, however, include functionality to support the Electronic Warfare Supervisor (EWS) and Sonar Control Supervisor (SCS) to feed information realistically into the Ops Room from EW and Sonar sources.

The intent is to take non audio-video data from the CMS330 in X-Ship. This data will be classified and appropriate permissions and procedures need to be sought and followed. The non audio-video data will be a significant amount of data, especially for longer scenarios. It is important to determine in advance exactly what data is needed for what period of time. This way, filters can be applied to the log so only the relevant information is transferred to the hard drive. Opportunities are being requested to carry out this preparatory work in advance of the actual validation trial and analysis.

It is expected that, during this trial, MON will be carrying out other operational and/or training tasks, to ensure that, outside of periods where operator work is stimulated by the SETT scenario, the ship is engaging in its normal activities and following its normal routines. The ship is expected to have no deficiencies or maintenance issues that would otherwise affect a mission and all shipboard sensors and weapons should be functional and available during the entire voyage. The trials team will work with the MON's command team to develop a schedule that integrates the trial with the activities of the ship, including any specific evolutions that require broader employment of the ship's company, such as deploying the towed array or flying stations.

3.3 Scenarios

The main objective of this project is to determine if a reduced combat crew, with support from the automation of CMS330, can achieve the objective level of performance required from the Combat Department. Different crew sizes will be tested in the simulation facility through simulation of representative operational scenarios. After discussions with the Canadian Forces Maritime Warfare Center (CFMWC) and CFFS(A), it was decided to derive our set of scenarios from a single master scenario developed by CFFS(A). This permits a smaller set of trial scenarios to be executed by a combat department that are tied together with a common theme and geographic area. This section presents the rationale behind the selection of scenarios (see Section 3.3.1), as well as the outline of the chosen scenarios (see Section 3.3.2).

3.3.1 Selection of Scenarios

The approach to be taken in this trial is that the combat department will carry out their regular tasks during the voyage. At pre-determined periods, identified MFWs will be placed in training mode and presented with one of the selected scenarios. Our trial scenarios will evolve realistically, therefore scenarios will last between 15 min (for a probe by an aircraft) and 240 min (for an anti-submarine scenario). Specific duration and specific segments will be chosen based on the ground truth analysis. Measurements during periods of lower intensity (e.g., just sailing) will also be taken to serve as a baseline. To be representative of the objective level of performance required by the RCN from the frigates, our trial scenarios need to simulate a complicated geopolitical situation where Canada is part of an alliance and there are hostile and neutral forces in the area, and where the neutral and hostile forces are hostile to each other and probably use similar platforms. This results in ambiguity, shifting alliances, and the need for the frigate to reorient to new or emerging threats. Our trial scenarios should begin with a simple transit, and then escalate in intensity (i.e., the number of contacts and their speed) and complexity (i.e., increased number of relationships between pairs of contacts, mix of domains, and ambiguity of information). Our trial scenarios should have 'layers', corresponding to domains (air, surface, subsurface) and identification (ID: friend, assumed friend, neutral, unknown, suspect, hostile). Finally, our trial scenarios should contain varying periods of intensity, representing:

1. routine Recognised Maritime Picture (RMP) compilation;

2. a surface problem;

3. an air problem;

4. a sub-surface problem;

5. a surface and air problem;

6. a sub-surface and air problem;

7. a sub-surface and surface problem; and

8. a surface, sub-surface and air problem.

Building scenarios is time and resource intensive. Scenarios have already been developed by CFMWC and CFFS(A) that are suitable for the purpose of this project with minor modifications. After discussions with CFMWC and CFFS(A), it was decided that the master scenario developed by CFFS(A) (which involved discussions with CFMWC) was most applicable for this trial. The master scenario developed by CFFS(A) was designed to train the operators to the level of competence required (as per CFCD106) in the detect-to-engage (D2E; see Appendix C for more information concerning D2E) sequence for different operational situations. Thus, the CFFS(A) master scenario corresponds to the most up to date Navy standards:

- It simulates the level of difficulty the Navy expects its crews to be able to deal with efficiently;
- The master scenario is used for both students and mature combat departments during Operational Team Training (OTT);
- The master scenario is used for training, they are well established;
- The script is rich and the sequence of events is well anchored;
- The master scenario is composed of smaller operational scenarios which last approximately 90-120 minutes; this duration corresponds to our requirement stated above; and
- The master scenario is already in SAFSIM and can be run in SETT with minor changes, if any.

3.3.2 Description of Scenarios

The CFFS(A) master scenario is based on current events and could plausibly develop in the real-world. The master scenario accords with standards from the Canadian Forces Joint Publication (CFJP) 5.2 concerning acceptable risk for the RCN (i.e., the scenario is challenging but a combat department could reasonably be expected to successfully complete it). The master scenario involves a complex geopolitical situation in which there is territory that is claimed by multiple nations. In support of a United Nations (UN) resolution, the RCN is present to ensure the peaceful conduct by all territorial claimants. The master scenario is supported by actual intelligence that is fed to the trainees over the course of the simulation. Indeed, students are provided with their mission in advance of the simulation so they have the opportunity to build their tactical pack (consisting of information they would need for the mission). Students are informed of the main goals and the key elements of the adversary's capability but tasks are not defined; these are passed in daily intentions messages. In practice, the actual tasks and missions are presented in accordance with the training curriculum.

The master scenario is complex and multi-faceted and supports the following mission types (of which each can constitute a smaller operational scenario):

- Humanitarian Relief;
- Strike;
- Special forces support;
- Non-combatant evacuation;
- Refugee support;
- Anti-piracy;
- Paramilitary action; and
- Human smuggling.

The master scenario has been in use for approximately eight months and a library of missions and scenario injects are being developed. Currently nine taskings have been developed, of which five are used with OTT crews. The taskings can be given sequentially to crews to represent a chronology of taskings wherein the situation has evolved from the tasking previously completed. CFFS(A) has developed narratives to explain how a tasking plays out to bring the crew to the context of a new tasking.

A requirement for our trial scenarios is that they must permit a single interactor to control all entities. This may require the actions of specific contacts in the scenario to be scripted to the greatest extent possible to minimise the workload on the interactor.

Due to training commitments for the RCN, it is not currently possible to do a detailed analysis of the master scenario to select and describe scenario elements that will be used to stimulate the different functions of the combat department. In a subsequent contract, a matrix will be developed to provide an overview of contacts, their affiliation and their timings to help the trial team understand what is happening in the scenario, when the periods of highest demand will occur and thus develop expectations for performance and problems with the crew.

3.3.3 Scenario Validation

Our trial scenario development will need to be validated in SETT on MON to ensure that it presents correctly and effectively to the participants. This will be a critical step for the success of the X-Ship validation trial as a complex scenario has not yet been played through SETT. A pilot study is proposed for November 2017 to validate this scenario. This activity will be carried in concert with validation of the data recording and data analysis.

3.4 Schedule

This section presents the schedule for the trial. Scenario development, data recording and analysis validation tasks are not included in this schedule and should take place as soon as possible to validate key assumptions in this trial plan:

- the trial scenarios chosen are appropriate for trial purposes;
- the trials scenarios display correctly in the selected simulation platform;
- data can be recorded from the simulation device; and
- data can be analysed by the trial team.

Scenario, data recording and analysis validation is described Section 3.4.1.

It is anticipated that this trial will occur in the second half of 2019 (i.e., sometime between July and December). Prior to the voyage, the trial team will work with the MON's command team to develop a schedule that integrates the trial with the activities of the ship, including any specific evolutions that require broader employment of the ship's company, such as deploying the towed array or flying stations. This activity is likely to happen early in 2019.

In total, the trial will use the X-Ship for 2 weeks (12-14 days) with some time dedicated to scenario testing and rehearsal prior to the sailing (i.e., while the ship is alongside in Halifax). Briefings to the commanding officer (CO), executive officer (XO), coxswain (CXN), and combat chief will occur on the first day. Briefing each watch will occur in a manner that works best logistically (e.g., both at once, or individually, depending on their schedules).

Within each on-watch, there may be no, one, or several trial scenarios presented, intermingled with regular ship duties (e.g., sailing). The weeks of the validation must be contiguous, however the trial scenario testing/rehearsal may occur well in advance. A pilot test is planned to occur about 2 years in advance of the actual validation trial although additional testing and preparation will occur at a time closer to the actual trial.

The trial is a within groups design in which all operators are exposed to all conditions (i.e., trial scenarios) twice. Therefore, each trial scenario will occur four times over the two weeks (i.e., twice per watch). Counterbalancing will be applied to the scenarios, whereby each watch will experience the trial scenarios in a different order. Measurements will be restricted to periods when the trial scenario is running and pre-identified periods when a trial scenario is not running. These pre-identified periods will be equivalent in duration to the time spent in a trial scenario and will be calculated individually for each operator. For instance, if the sum total of time the Track Supervisor spends participating in trial scenarios is 600 minutes, 600 minutes of on-watch time, non-scenario time will also be measured for the Track Supervisor for comparison.

A detailed schedule for the duration of the trial will be developed after the ground truth analysis is completed.

3.4.1 Pilot Test

Pilot testing will occur in November 2018 on MON. The objective of the pilot testing is threefold: 1) ensure that the scenarios work properly on MON; 2) ensure that data can be downloaded from MON; and 3) ensure that the data from MON can be analyzed.

There are two parts to the pilot test on MON. First, the selected trial scenarios will be run to ensure they present properly to the operators. These trial scenarios will already have been run in SEAWOLF so the level of demand imposed on operators will already have been validated. The trial team will perform actions at the MFW while these trial scenarios are running and note the timing of those actions.

The second part involves a more extensive consideration of the data that can be collected. The trial team will take advantage of an exercise in which MON is participating with a great deal of scripted elements (e.g., a missile exercise). Because the event is scripted, much of the ground truth is known. The trial team will attend the exercise and make detailed notes concerning the contact of interest (e.g., a missile) and the operators actions on that missile.

At the end of the pilot test, data will be collected to ensure that the right data can be collected and that the format is readily usable by the team for further analysis. The data will be reviewed to identify all the actions and timings of the actions performed by the trial team to validate that the right data is recorded and extracted. Finally, data will be analysed to determine if:

- actions and events are logged appropriately;
- captured data can be analysed;
- captured data can be associated with ground truth; and
- there are any difficulties in the execution and measurement of the validation trial for which alternative plans need to be developed.

Based on the results of the pilot study, adjustments may be made to the trial scenarios.

3.5 Participants

For this trial, the full combat department of MON will be participating for 2 full weeks of trials (see Section 3.4 for a detailed schedule). The crew should be fully qualified (i.e., weapon certified). It is expected that all combat operators will be suitably qualified as indicated in Table 3-1. Note that the SWC and ASWC roles can be performed by either a Non-Commissioned Member (NCM) or an officer, hence they appear twice.

Table 3-1: Trades Required and Qualification Levels

	NESOP	NCIOP	SONAR	MARS
QL3	ESM	ASPO	HMS	
QL4	FC Operator	ARRO	Towed Array and HMS	
QL5a	FC Operator	TS	SPS	SAC
QL5b	FCS	C4i	Towed Array and HMS Supervisor	
QL6a	EWS	ORS	SCS	
QL6b	SWC	IMD	ASWC	SWC, ASWC, ORO

Subject to availability, one CFFS(A) staff with knowledge of the scenarios and how they should appear to operators will be required for the pilot test. During the validation trial itself, one NCM from Sea Training with knowledge and experience of picture compilation and one Maritime Surface and Subsurface (MARS) officer from Sea Training with knowledge and experience of threat evaluation and weapons employment will be required.

3.5.1 Participant Training

Although each participant will be fully qualified prior to attending the trial, there will still be two forms of training required: operational training and measurement training. Operational training will consist of a familiarisation briefing. All participants will be briefed on the purpose of the trial and the measurement philosophy, i.e., that their individual performance is not being assessed, data will be anonymized and their performance during the trial will not be related or otherwise impact their career progression and formal performance evaluations.

Measurement training will introduce the measurement instruments that participants will be exposed to during the trial. It is anticipated that most subjective measurements will take place following a simulation run. The exception is the Instantaneous Self-Assessment (ISA) which is a simple workload measure which will be collected at regular intervals (interval to be determined later) through a simulation run. Participants will be briefed on the different measurement instruments and the definitions to be used when responding to them. Participants will be given a simple game (Tetris) to play while being prompted for their ISA response to familiarise themselves with this measurement demand and seek any clarification required. At the conclusion of the training game they will complete the NASA Task Load Index (TLX) and be asked complete a short questionnaire about their workload and performance during the game.

The Tetris game was selected because task demand rises gradually over time as the pieces are presented more quickly and the game requires the player's constant attention. Because performance is scored, there is a multi-dimensional aspect to the task demand, with the player being able to choose between clearing blocks as expediently as possible thereby avoiding failure, or attempting to maximise their score by clearing multiple rows of blocks at once with the attendant risks of failing while waiting for the correct shape. The continuous demands of Tetris make it a good training game to familiarise with ISA, while the multi-dimensional components of task demand mean it permits the player to consider the different dimensions presented on the NASA TLX.

3.6 Validation Trial Team

The validation trial team will consist of three CAE Human Factors (HF) consultants, three DRDC defense scientists, two Subject Matter Experts (SMEs) from Sea Training and 1 interactor. Table 3-2 presents the team members and describes their role.

Table 3-2: Validation Trial Team

Organization	Position	Role
CAE	HF Consultant - Test director	Test director Brief participants Start/stop scenarios Observe Simulation (ratings, notes, etc.) Analyse Data
	HF Consultant (x2)	Set up trial material Administer questionnaires Observe Simulation (ratings, notes, etc.) Pick-up data Store Data Analyse Data
DRDC	Defence Scientists (x3)	Administer questionnaires Observe Simulation (ratings, notes, etc.)
Sea Training SME	NCM - Picture compilation (x1)	Subjective Performance Rating
	MARS Officer directors (TEWA) (x1)	Subjective Performance Rating
CFFS(A) SME	Interactor (x1)	Coordination with test director Directing execution of scenario Act as white force (i.e., non-combatant contacts)

Organization	Position	Role
		Act as blue force (i.e., allied contacts) Act as red force (i.e., enemy or suspect contacts)

During the measured sessions, the trials team will need to start and stop the scenarios at predefined times, pass questionnaires and gather data. Precise timings of pauses and the number of pauses will follow detailed analysis of the scenarios.

4 MEASURES

This section describes the measures that will be used to test the hypotheses presented in Section 2. It is necessary to collect data concerning measures of performance (MOPs) and measures of effectiveness (MOEs) in order to unequivocally accept or reject a hypothesis. MOEs are measures that are directed towards the way in which a system meets its stated operational goals. MOPs on the other hand are more directed toward the processes required to produce the system output (Helleur, Mathews, Kashyap, & Rafuse, 2007). Based on a literature review, Matthews, Webb and Keeble (2002) recommended to measure performance using accuracy/error and time on tasks measures supplemented by subjective ratings of performance. Therefore, as illustrated in Figure 4-1, a combination of objectives and subjective measures will be used in this trial. Objective measures are described in Section 4.3 and subjective measures in Section 4.4.

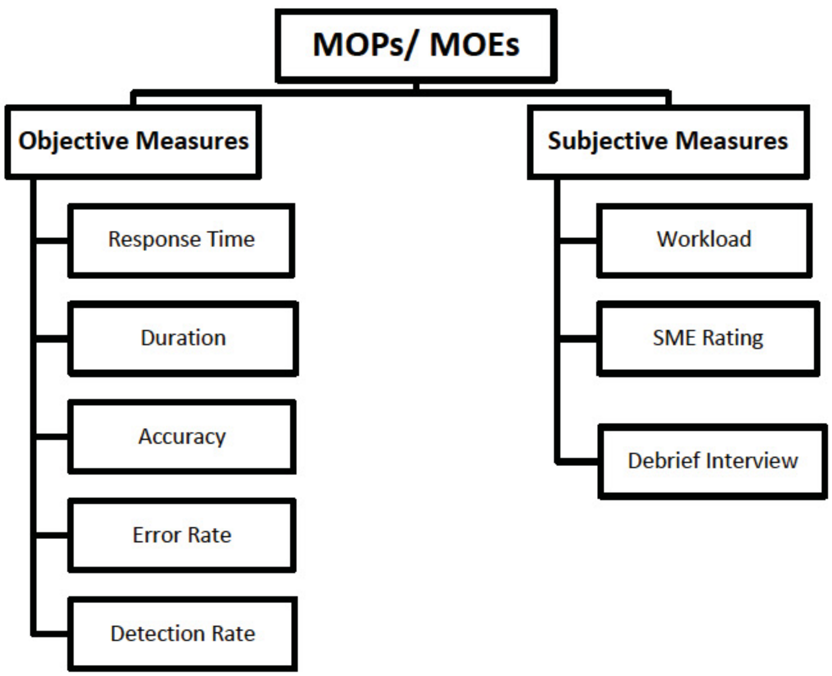


Figure 4-1: Measures of Performance

4.1 MOEs and MOPs

A MOE is concerned with assessing ability of a system to achieve its goal; its impact on the operational environment. In this respect, an MOE tends to be more complex than an MOP and may include two or more MOPs in its determination. An MOE focuses on a holistic assessment of how well a system performs overall and can be considered an overarching measurement that comprises more than one MOP. MOEs focus on overall mission goals such as whether or not the crew met their objectives, completed their mission, minimized casualties and collateral damage, etc.

An MOP is a quantifiable measurement that can take any number of forms, such as a simple count of something, a physical measurement, an average, a rate, a percentage, etc. An MOP may be stated as a declarative statement and indicates a system's achieved level of performance. An MOP does not provide an assessment of the overall impact of the measurement attribute on the goal of the system; It focuses on the absolute measurement of unidimensional qualities. MOPs, unlike MOEs, directly address the use of the mission system through absolute metrics (e.g., how long they spent using something, how many key presses were required) and through more intangible metrics such as workload and situation awareness.

Previous work by Therrien (2016) and Gauthier, Bourdon, Dore and Fong (2004) has identified MOEs and MOPs. MOPs will be dealt with in more detail in the following sections, but the MOEs should be discussed since they are similar to the hypotheses being investigated. Table 4-1 lists the MOEs from Therrien (2016) and Gauthier et al (2004) side-by-side. The MOEs do not map perfectly nor comprehensively, reflecting the purposes for which they were developed. The MOEs in Therrien (2016) reflect the full D2E cycle and are predisposed toward theatre-level missile defence; the MOEs in Gauthier et al (2004) were developed to evaluate the quality of the RMP (there are additional MOEs in Gauthier et al (2004) concerning data fusion and sensor performance).

Table 4-1: MOEs from Therrien (2016) and Gauthier et al. (2004)

Therrien (2016)	Gauthier et al (2004)
Planning and tasking	
Surveillance	Completeness
Defended area/asset	
Detection	
Tracking	
Identification	Correctness, Currency, Clarity, Extent of information
Correlation	
Degree of protection	
Defendable stress level	Prediction performance
Force tracking	Consistency
Event effector C2	
Effect delivery	
Number of leakers	
Intercept efficiency	
Blue casualty avoidance	
Collateral damage containment	

The MOPs presented below are mapped to the MOEs in Section 4.3. Note that only Therrien's (2016) MOEs are presented since they cover the D2E cycle more broadly.

4.2 Data Collection

The data required to assess performance will be collected in various ways. The objective data will be collected through audio-video recordings of the scenarios. The audio data will come from the communication system, where each communication made through the system will be automatically recorded. The video data comes from the playback function of the CMS330. Note that the audio-video data is gathered only to serve as a back-up if a member of the trial team needs to go back and verify some information. At the moment there is no plan to analyse the audio-video data. The non audio-video data will record software messages (actions and objects acted upon) and their timings. The non audio-video data will be analysed using Matlab. Finally, subjective data will be collected from operators via questionnaires and interviews during pauses in the scenarios, at the conclusion of each scenario run, during onwatch periods without a scenario, and at the conclusion of onwatch periods. Additional subjective data will be provided by SME observers who will make notes on performance that is attributable to the crew configuration and the demands imposed by the system (i.e., not due to the skill or competence of the operator themselves).

4.2.1 Data Recording and Analysis Validation

The ability of SETT to record data and correctly identify and attribute operator interactions with the CMS needs to be validated. This activity will be carried out in concert with scenario validation. Data recording and analysis validation will take place during the pilot test.

The general procedure for the data recording and analysis validation will be as follows:

1. place selected MFWs into training mode;
2. run the scenario;
3. hook a contact;
4. note the time the contact was hooked, the workstation identification and the track number of the contact;
5. hook a new contact;
6. note the time the new contact was hooked, the workstation identification and the track number of the contact;
7. repeat steps 2 to 5 for all workstations, ensuring that a different contact is hooked at each workstation;
8. download the data to a removable hard-drive;

9. review the data to find and understand:

    a. the time of each action;

    b. the interval between each action;

    c. the contact that was hooked; and

    d. the workstation at which the activity took place.

10. verify this data matches manual records made in the simulator.

Further task steps (e.g., weapons assignment, doctrine selection) will be added to this procedure on the basis of task analysis data (Dubé, G. & Lamoureux, T.,2016b) to identify the system messages that will need to be used to generate MOP data.

A pilot test will also be used to draft a detailed procedure to execute the scenario so the process proceeds with no mistakes during the actual trial.

The second part to the pilot test (observation and data collection concerning ship evolutions already scheduled, e.g., a missile exercise) will follow the same approach as the scenario validation test.

This data recording and analysis validation will provide early confidence that the simulator can provide the data required to answer the research question.

4.3 Objective Measures

The main objective of the automation (DF, Doctrine, TE and WA) is to support the operator in the detection of, management of and response to contacts. With this goal in mind, key performance events (KPE) were developed based the D2E cycle (see Appendix C for more information concerning D2E). KPEs are events that trigger a response by the system or the operator that is recorded in the system log and will be used to support the development of objective measures of performance. KPEs are based on general MOP frameworks such as that suggested by Matthews et al (2002) concerning timings, duration and accuracy. Seven KPEs were selected:

- detection of the contact by the system;
- detection of the contact by the operator;
- classification of the contact by the operator;
- recognition of the contact by the operator;
- identification of the contact by the operator;

- engagement of the contact by the operator; and
- contact is no longer in the system.

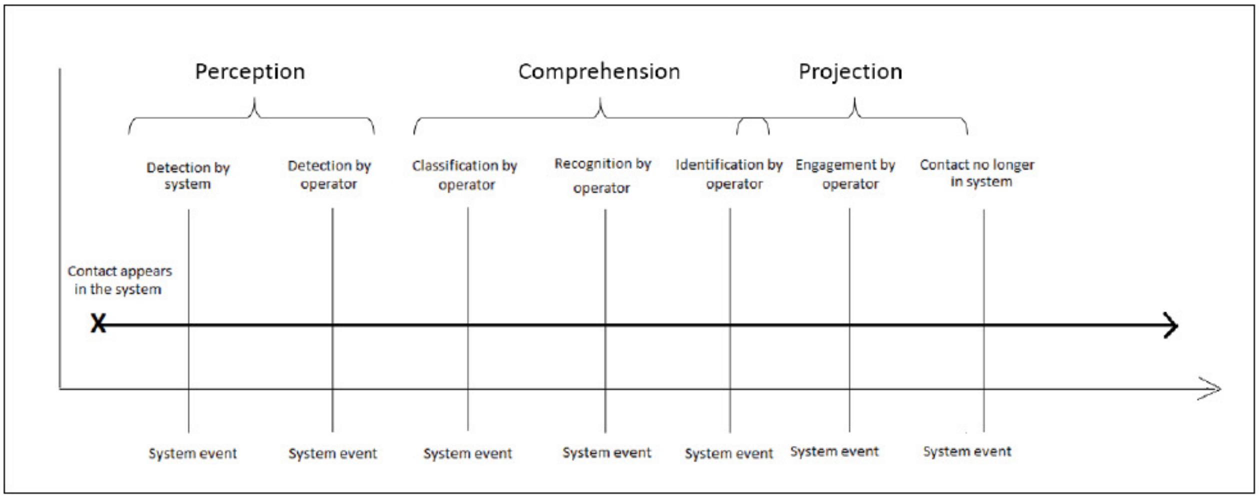


Figure 4-2: Operationalisation of Performance Measures

As illustrated in Figure 4-2, each of the KPEs can be associated to a 'system event' and/or a scenario event. A scenario event is one that is scripted to occur at a specified time and/or in response to a role player's trigger. Typically, a scenario event will also represent a planned pattern of actions. A scenario event is input via the *simulator* and will most likely result in a corresponding *system* event, typically a response of some sort. Initially this response will be a system (automated) response before the scenario event is communicated to the operator who may also make a response (e.g., hooking a new contact) which will be input via the *system* (i.e., the CMS330)[5]. Both the system and the operator response are recorded in the system log (non-audio-video data). The trials team controls the timing of the scenario event, but knows the timing of the system events because of the data log. Thus, the trials team knows 'ground truth' and can determine the true story of system and operator actions to evaluate overall system performance without having to rely solely on subjective data from operators and observers. KPEs will be used as markers to evaluate overall performance in the different crew configurations. An example of system events relating to each KPE can be found in Table 4-2 (in due course, a full list of system events that are expected to be recorded and analysed will be provided, following a deeper analysis of the trial scenarios). The KPEs and their related system events will be used as marker to calculate the performance.

[5] The reader should note the distinction between scenario events and system events: scenario events are mediated by the simulator and system events are mediated by the CMS330.

Table 4-2: System Events

	Detection by System	Detection by Operator	Classification by Operator	Recognition by Operator	Identification by Operator	Engagement by Operator	End
Picture Compilation (DF)	Appearance of the track in system	Mouse Over Hook	Category (Hostile, friend, etc.) assigned to track		ID assigned to track		
Threat Evaluation			Mouse Over Hook	Mouse Over Hook	Mouse Over Hook		
Weapon Assignment					Engagement plan window	Engagement plan approved	
Sensor Utilization					Doctrine Selected Doctrine Loaded		
Weapons Utilization						Weapon fire	

In order to translate the above KPEs into MOPs that are valid and representative of the operational tasks of the combat team, a Task Analysis (TA) was used (Dubé & Lamoureux, 2016b). The TA described the tasks to be performed that are mediated or required by the automation. The MOPs and MOEs derived from the TA are based on response time, duration of action or state, accuracy, detection rate and error. Table 4-3 presents the mapping between the MOEs and the MOPs selected for this study. For each MOE, a set of MOPs is identified and categorised based on the type of measure (response time, duration or accuracy). The MOPs include a cross reference to the corresponding descriptive section.

Table 4-3: MOEs vs MOPs

MOEs	MOPs				
	Response Time	Duration	Accuracy	Detection Rate	Error
Detection	Mean Time from Entry into Battlespace to Initial Detection (4.3.1.1)		Percentage of Objects Detected and Tracked (4.3.3.1)	Percentage of Tracks Detected (4.3.5.1)	

MOEs	MOPs				
	Response Time	**Duration**	**Accuracy**	**Detection Rate**	**Error**
Tracking			Percentage of False Tracks (4.3.3.2) Positional Accuracy (4.3.3.3) Track Completeness (4.3.3.4)		
Identification	Mean Time between Initial Detection and Correct Identification (4.3.1.2)		ID Completeness (4.3.3.5) Percentage of time a correct ID is applied to an object (4.3.3.6) Percentage of contact initially identify correctly which were later mis-identified (4.3.3.7) Percentage of time contacts were depicted with track ID conflicts (4.3.3.8)	Percentage of time an incorrect ID is applied to an object (4.3.3.6)	
Correlation		Duration of dual (multiple) designations (4.3.2.1)	Percentage of dual designations (4.3.3.9)		
Degree of protection	Mean time to engage threat (4.3.1.3) Mean Time to Apply Doctrine (4.3.1.4)		Mean range of target(s) from defended area or asset or ownship for initial engagement (4.3.3.10)		
Force Tracking			Mean number of track number changes per object (4.3.3.11)		
Event Effector C2			Number of threats engaged (4.3.3.12)		Number of invalid assignments (4.3.4.2)

MOEs	MOPs				
	Response Time	Duration	Accuracy	Detection Rate	Error
			Number of engagements ordered against each threat (4.3.3.13)		Number of Invalid Doctrine (4.3.4.3)
Effect Delivery			Percentage of successful engagements (4.3.3.14)		
Number of Leakers			Ratio of successful penetrations to total number of threats (4.3.3.15)		

4.3.1 Response Time

The response time is the time an operator (or the combat team as a whole) takes to react to a stimulus and execute a response. In other words, response time is the time elapsed between two events, at least one of which must be a system event (i.e. one can, but does not have to be, a scenario event). System event time stamps are extracted from the CMS330 event log and used to compute response time.

Three response time-based measures will me used in this study: 1) mean time from entry into battlespace to initial detection (Section 4.3.1.1); 2) mean time between initial detection and correct identification (Section 4.3.1.2); and 3) mean time to engage threat (Section 4.3.1.3). Those measures were taken from the *MTMD Forum Master Test Plan* (Therrien, 2016).

4.3.1.1 Mean Time from Entry into Battlespace to Initial Detection

This section was taken from *the MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.4.3. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.1.1.1 Description

This measure quantifies the difference in time when an object entered the defined battlespace and when the participant(s) first detected it at the CMS. This measure is comprised of two equations-time difference per object and unit level:

$$T_{j,m} = t_{j,m} - t_j \qquad (1)$$

Where,

$T_{j,m}$ = time difference of when participant m first detected object j and when object j first entered the battlespace at the CMS,

$t_{j,m}$ = time object j was first detected within an assigned battlespace by participant m, and

t_j = actual time that object j entered the assigned battlespace.

$$T_{J,m} = \frac{\sum_{j=1}^{J} T_{j,m}}{NT_{j,m}} \qquad (2)$$

Where,

$T_{J,m}$ = unit level time difference of when participant m first detected all objects J and when all objects J first entered battlespace (un-weighted mean of equation (3)),

$NT_{j,m}$ = total number of $T_{j,m}$ samples for participant m, and

J = total number of objects over the evaluation period.

4.3.1.1.2 Dimension of the Measure

This measure is a time measurement expressed in seconds. The variables of $t_{j,m}$ and t_J are greater than zero. The measures $T_{j,m}$, $T_{j,M}$, and $T_{J,M}$ are greater than zero.

4.3.1.1.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The object of interest (OOI) must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. The battlespace must be defined. The t_j time is based on truth data. Only objects that entered the battlespace will be assessed.

4.3.1.1.4 Assessment Methodology

In order to calculate the measure, $T_{j,m}$ and $T_{J,m}$ must be determined for each participating unit. If no track start time $t_{j,m}$ is found for an object, then $T_{j,m}$will not be calculated for that object j.

4.3.1.2 Mean Time between Initial Detection and Correct Identification

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.6.3. Force Level measures have been removed to reflect only the measures that will be used in this study.

4.3.1.2.1 Description

This metric quantifies the mean time it took from the initial detection of an object to correctly identifying that object at the CMS. The measure is comprised of two equations- time it takes to correctly identify tracks per object and unit level:

$$T_{j,m,ID} = t_{j,m,ID} - t_{j,m,T} \qquad (3)$$

Where,

$T_{j,m,ID}$ = time difference of when object j is initially detected to when it is correctly identified at participant m at the CMS,

$t_{j,m,ID}$ = the time object j was correctly identified at participant m, and

$t_{j,m,T}$ = the time object j was initially tracked at participant m.

If for the life of an object there are dual designations and the IDs are incorrect and correct, then object ID state is incorrect. If initially the object has dual designations and the IDs are incorrect and correct, at the time the dual designations seize to exist and the ID state is correct the $t_{j,m,ID}$ value is determined.

$$T_{J,m,ID} = \frac{\sum_{j=1}^{J} T_{j,m,ID}}{NT_{j,m,ID}} \qquad (4)$$

Where,

$T_{j,m,ID}$ = unit level measure of mean time from initial detected to correct ID for all objects J at participant m (un-weighted mean) at the CMS, and

$NT_{j,m,iD}$ = the total number of $T_{j,m,ID}$ for participant m.

4.3.1.2.2 Dimension of the Measure

This measure is a time measurement expressed in seconds.

The variables of $t_{j,m,ID}$ and $t_{j,m,T}$ are greater than zero. The measures $T_{j,m,ID}$ and $T_{J,m,ID}$, are greater than zero.

4.3.1.2.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI and respective ID assignments must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined. This measure is calculated at the CMS level.

The variables of $t_{j,m,ID}$ and $t_{j,m,T}$ are based on digital data from CMS. If variable $t_{j,m,T}$ is not found for an object, then $T_{j,m,ID}$ will not be calculated for that object.

4.3.1.2.4 Assessment Methodology

In order to calculate the measure, $T_{j,m,ID}$ and $T_{J,m,ID}$ must be determined for each participating unit. If no track start time $t_{j,m,T}$ is found for an object, then $T_{j,m,ID}$ will not be calculated for that object j.

4.3.1.3 Mean Time to Engage Threat

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 2.1.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.1.3.1 Description

This measure quantifies the mean time when a threat enters the AOI, to when the first engagement is ordered. The term *engagement is ordered* refers to the order issued at CMS to the weapons system for weapons engagement against a threat k. The measure is comprised of two equations:

$$\boldsymbol{T_{k,EO} = t_{k,EO} - t_k} \quad (5)$$

Where,

$T_{k,EO}$ = time to engage,

$t_{k,EO}$ = time first engagement was ordered against threat k, and

t_k = actual time that threat k entered the assigned AOI.

$$\boldsymbol{T_{K,EO} = \frac{\sum_{k=1}^{K} T_{k,EO}}{NT_{k,EO}}} \quad (6)$$

Where,

$T_{K,EO}$ = mean time to engage (unweighted mean of $T_{k,EO}$)

$NT_{k,EO}$ = total number of $T_{k,EO}$ samples, and

K = total number of threats.

4.3.1.3.2 Dimension of the Measure

The measure is a time measurement that is expressed in seconds.

The variable t_k is less than or equal to $t_{k,EO}$. The measures $T_{k,EO}$ and $T_{K,EO}$ are greater than or equal to zero.

4.3.1.3.3 Conditions of the Measure Assessment

This measure is calculated using digital data from the CMS and Truth. All threats k and AOI must be identified to calculate measure. The track reconstruction of all threats k must be completed. Threats that are unengaged are not included in this measure. Threats that are not tracked are not included in measure. Targets engaged prior to entering AOI will not be included in measure.

4.3.1.3.4 Assessment Methodology

In order to calculate the measure, $t_{k,EO}$ for each threat k at each unit is identified using local CMS digital data. The variable $t_{k,EO}$ is determined by using the initial time when the local CMS ordered weapons engagement on threat k and variable t_k is determined by using the truth time of when the threat k entered the AOI. $T_{k,EO}$ is calculated using variables $t_{k,EO}$ and t_k.

4.3.1.4 Mean Time to Apply Doctrine

This section was inspired from the *MTMD Forum Master Test Plan* (Therrien, 2016).

4.3.1.4.1 Description

This measure quantifies the mean time when a missile enters the AOI, to when the Doctrine is loaded. The measure is comprised of two equations:

$$T_{ms,D} = t_{ms,DS} - t_{ms} \qquad (7)$$

Where,

$T_{ms,D}$ = time to load,

T_{msDS} = actual time Doctrine was ordered (i.e. approved) against missile ms, and

T_{ms} = actual time that missile ms entered the assigned AOI.

$$T_{MS,D} = \frac{\sum_{ms=1}^{MS} T_{ms,D}}{NT_{ms,D}} \qquad (8)$$

Where,

$T_{MS,D}$ = mean time to load (unweighted mean of $T_{ms,D}$)

$NT_{ms,D}$ = total number of $T_{ms,D}$ samples, and

MS = total number of missiles.

4.3.1.4.2 Dimension of the Measure

The measure is a time measurement that is expressed in seconds.

The variable t_{MS} is less than or equal to $t_{ms,D}$. The measures $T_{ms,D}$ and $T_{MS,D}$ are greater than or equal to zero.

4.3.1.4.3 Conditions of the Measure Assessment

This measure is calculated using digital data from the CMS and Truth. All missiles ms and AOI must be identified to calculate measure. The track reconstruction of all missiles ms must be completed. Missiles that are unengaged are not included in this measure. Missiles that are not tracked are not included in measure. Targets engaged prior to entering AOI will not be included in measure.

4.3.1.4.4 Assessment Methodology

In order to calculate the measure, $t_{MS,D}$ for each missile ms at each unit is identified using local CMS digital data. The variable $t_{ms,D}$ is determined by using the initial time when the local CMS load a Doctrine on missile ms and variable t_{ms} is determined by using the truth time of when the missile ms entered the AOI. $T_{MS,D}$ is calculated using variables $t_{ms,D}$ and t_{ms}.

4.3.2 Duration

The duration of an action or a state is another widely used MOP. It is measured using the digital capture of time the operator or the combat team took to perform a specific task, such as respond to a threats. The duration is obtained through the CMS330 events log from which system events time stamp are extracted.

For this study, a single duration-based measure will be used: Duration of dual (multiple) designation. This measure was taken from the MTMD Forum Master Test Plan (Therrien, 2016).

4.3.2.1 Duration of Dual (Multiple) Designations

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.7.2. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.2.1.1 Description

This measure quantifies the duration of time a dual (multiple) track (s) is/are held for an object. Dual tracks occur when more than one track, assigned to the same object, is displayed to some or all of the participating units. This measure is comprised of two equations:

$$D_{j,m} = \frac{\sum_{1}^{NT}(t_{j,d,m} - t_{j,i,m})}{NT} \qquad (9)$$

Where,

> $D_{j,m}$ = mean duration of dual track when participant m first held dual track and the time dual track was dropped,
>
> $t_{j,i,m}$ = time in which the dual track was initiated at object j (if $t_{j,i,m}$ does not have a value, then $D_{j,m}$ is not computed for the object j),
>
> $t_{j,d,m}$ = time dual track ceased to exist on object j, and
>
> NT = total number of dual tracks that were designated per object j (dual tracks less than 30 seconds will not be considered for NT).

$$\boldsymbol{D_{m=}} \frac{\sum_{J=1}^{J} D_{j,m}}{ND_{j,m}} \qquad (10)$$

Where,

> D_m = unit level measure of time duration of dual tracks for participant m for all objects J (un weighted mean of $D_{j,m}$), and
>
> $ND_{j,m}$ = total number of $D_{j,m}$ samples for participant m.

4.3.2.1.2 Dimension of the Measure

The measure is a time difference expressed in seconds.

The measure $D_{j,m}$ and D_m, values must be greater than zero.

4.3.2.1.3 Conditions of the Measure Assessment

This measure requires digital data from CMS and truth. In addition, the OOI must be known; track reconstruction of OOI must be completed. The measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined.

The measures $D_{j,m}$ and D_m must be determined for each participating unit using CMS digital data.

4.3.3 Accuracy

Accuracy is often used as an MOP in combination with response time as there is a speed/accuracy trade off. Therefore, both should be measured and plotted. Accuracy measures are largely manual, post-event analysis expressed in the form of percentage or proportions of error. Data regarding accuracy is obtained through the CMS330 events.

For this study, 15 accuracy-based measures will be used. They are presented in Sections 4.3.3.2 to 4.3.3.15. Those measures were taken from the *MTMD Forum Master Test Plan* (Therrien, 2016).

4.3.3.1 Percentage of Objects Detected and Tracked

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.4.2. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.1.1 Description

This measure quantifies the percentage of time an object was detected and tracked within an Area of Interest (AOI) by each unit locally at the CMS. This is a time dependent measure. The measure is comprised of two equations-instantaneous and unit level:

$$\boldsymbol{P_{T,m}(t)} = \left(\frac{\boldsymbol{Jt_m(t)}}{\boldsymbol{J(t)}}\right) \times \mathbf{100\%} \qquad (11)$$

Where,

$P_{T,m}$ (t) = instantaneous mean time detected and tracked at time t for participant m in defined AOI at the CMS, and
Jt_m(t) = number of objects with at least one assigned local track held by participant m at time t in defined AOI.
J(t) = number of objects at time t in a defined AOI.

$$\boldsymbol{P_{T,m}} = \sum_{\boldsymbol{t_{start}}}^{\boldsymbol{t_{end}}} \boldsymbol{P_{T,m}(t)} \quad (12)$$

Where,

$P_{T,m}$ = unit level mean time tracked for participant m (object-count weighted mean across time).

4.3.3.1.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measures PT,m(t) and PT,m can range in value from zero to 100%.

4.3.3.1.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The Object of Interest (OOI) must be known; track reconstruction of OOI must be completed. Measure calculation is based on local tracks at participating units' CMS. Evaluation period and AOI must be defined. A one second sampling rate is recommended for measure.

4.3.3.1.4 Assessment Methodology

In order to calculate measure PT,m, PT,m(t) must be determined for each time t during evaluation period for participant m. The variable PT,m(T) is determined by using participating unit's local CMS digital data. The variable J(t) is determined by using the truth data for each OOI. The measure PT,m is the summation of the PT,m(t) during evaluation period

4.3.3.2 Percentage of False Tracks

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.5.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.2.1 Description

This measure quantifies the percentage of false tracks present in a track picture. A track is false when it is not assigned to any object. When collecting data in a live exercise, it is not always possible to determine if false tracks were generated by real objects or not. In these cases, other factors need to be considered that reflect engineering knowledge of the reporting units' sensitivity to environmental and atmospheric clutter or sensor alignment. This calculation is comprised of two equations-instantaneous and unit level:

$$S_m(t) = \left(\frac{N_m(t) - NA_m(t)}{N_m(t)}\right) \times 100\% \qquad (13)$$

Where,

$S_m(t)$ = instantaneous measure of the percentage of tracks that are false at participant m at time t,

$N_m(t)$ = the number of tracks held by participant m at time t, and

$NA_m(t)$ = the number of tracks assigned to an object on participant m at time t.

$$S_m = \sum_{t_{start}}^{t_{end}} S_m(t) \qquad (14)$$

Where,

S_m = unit level measure of the percentage of tracks that are false at participant m (track weighted mean across time).

The force level measure of false tracks (S) is the un-weighted mean of the unit level measure across participants:

4.3.3.2.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measures $S_m(t)$ and S_m values can range from zero to 100%.

4.3.3.2.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined. A one second sampling rate is recommended for measure.

4.3.3.2.4 Assessment Methodology

The measures $S_m(t)$ and S_m must be determined for each participating unit.

4.3.3.3 Positional Accuracy

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.5.2. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.3.1 Description

This measure quantifies the position accuracy of each assigned track compared to the object it is associated to. All calculations will be in terms of how track data at participant m compares to truth data. This measure is comprised of two equations-instantaneous and unit level:

$$PA_{j,n,m}(\mathbf{t}) = \sqrt{\left[X_{j,n,m}(t)\right]^2 + \left[Y_{j,n,m}(t)\right]^2 + w\left[Z_{j,n,m}(t)\right]^2} \quad (15)$$

Where,

$PA_{j,n,m}(t)$ = position accuracy at time t from track n held by participant m associated with object j at time t,

$X_{j,n,m}(t)$, $Y_{j,n,m}(t)$, $Z_{j,n,m}(t)$ = Cartesian position coordinates for object j in the local east-north-up coordinate frame of track n held by participant m (track-centered moving coordinate frame), and

w = weighting factor used to reduce the significance of the vertical position and velocity errors on 2-D sensors. Currently it is anticipated that the weights used will be either one or zero.

$$PA_m = \frac{\sum_{t_{start}}^{t_{end}} \sum_{j=1}^{JT_m(t)} \sum_{n=1}^{NA_{j,m}(t)} PA_{j,n,m}(t)}{\sum_{t_{start}}^{t_{end}} \sum_{j=1}^{JT_m(t)} NA_{j,m}(t)} \quad (16)$$

Where,

PA_m = unit level measure of position accuracy is an mean across assigned tracks and across scoring times t of the instantaneous measure, and

$NA_{j,m}(t)$ = the number of different tracks held by participant m that are assigned to object j at time t.

4.3.3.3.2 Dimension of the Measure

This is a position measure expressed in kilometers (km).

The measures PA_m and PA values must be greater than or equal to 0.

4.3.3.3.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined.

4.3.3.3.4 Assessment Methodology

The measures $PA_{j,n,m}(t)$ and PA_m must be determined for each participating unit.

4.3.3.4 Track Completeness

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.5.3. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.4.1 Description

This measure quantifies the percentage of time an object was tracked and reported for the coalition force. This measure is comprised of two equations-instantaneous and unit level:

$$C_m(t) = \left(\frac{JT_m(t)}{J(t)}\right) \times 100\% \qquad (17)$$

Where,

$C_m(t)$ = instantaneous track completeness at participant m at time t, and

$JT_m(t)$ = number of objects with at least one assigned track held by participant m at time t.

J(t) = number of objects at time t in a defined AOI.

$$C_m = \sum_{t_{start}}^{t_{end}} C_m(t) \qquad (18)$$

Where,

C_m = unit level track completeness at participant m (object count weighted mean across time).

4.3.3.4.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measures $C_m(t)$ and C_m can range in value from zero to 100%.

4.3.3.4.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined. A one second sampling rate is recommended for measure.

4.3.3.4.4 Assessment Methodology

The measures $C_m(t)$ and C_m must be determined for each participating unit.

4.3.3.5 ID Completeness

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.6.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.5.1 Description

The ID is complete when all tracked objects are labeled in an identified state (i.e., a state other than "not identified"). The measure quantifies the percentage of time an object is identified. This measure is comprised of two equations-instantaneous and unit level:

$$C_{ID,m}(t) = \left(\frac{JT_m(t) - JU_m(t)}{JT_m(t)}\right) \times 100\% \qquad (19)$$

Where,

$C_{ID,m}(t)$ = instantaneous ID completeness at participant m at time t, and

$JU_m(t)$ = number of tracked in an "not identified" ID state held by participant m at time t (This will include any specific ID assignments, or combinations thereof, that are defined to be "not identified" assessments in the scenario).

$JT_m(t)$ = number of objects with at least one assigned track held by participant m at time t.

$$C_{ID,m} = \sum_{t_{start}}^{t_{end}} C_{ID,m}(t) \quad (20)$$

Where,

$C_{ID,m}$ = unit level ID completeness at participant m (object weighted mean over time).

4.3.3.5.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measures $C_{ID,m}(t)$ and $C_{ID,m}$ can range in value from zero to 100%.

4.3.3.5.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI must be known; track reconstruction of OOI must be completed. The OOI correct and incorrect ID assignments must be known in order to determine what the “not identified” state is. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units’ CMS. Evaluation period must be defined. A one second sampling rate is recommended for measure.

4.3.3.5.4 Assessment Methodology

The measures $C_{ID,m}(t)$ and $C_{ID,m}$ must be determined for each participating unit.

4.3.3.6 Percentage of Time Correct ID is Applied to an Object

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.6.5. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.6.1 Description

This measure quantifies the percentage of time all tracked objects are labeled correctly. This measure is comprised of two equations-instantaneous and unit level:

$$IDC_m(t) = \left(\frac{JC_m(t)}{JT_m(t)}\right) \times 100\% \quad (21)$$

Where,

$IDC_m(t)$ = instantaneous ID correctness at participant m at time t, and

$JC_m(t)$ = number of tracked objects with all ID labels correct on all tracks representing each object (as defined in the scenario) held by participant m at time t.

$JT_m(t)$ = number of objects with at least one assigned track held by participant m at time t.

$$IDC_m = \sum_{t_{start}}^{t_{end}} IDC_m(t) \quad (22)$$

Where,

IDC_m = unit level ID correctness at participant m (object-weighted mean over time).

4.3.3.6.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measures $IDC_m(t)$ and IDC_m can range in value from zero to 100%.

4.3.3.6.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI and respective ID assignments must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined. A one second sampling rate is recommended for measure.

4.3.3.6.4 Assessment Methodology

The measures $IDC_m(t)$ and IDC_m must be determined for each participating unit.

4.3.3.7 Percentage of Contacts Initially Identified Correctly which were Later Misidentified

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.6.6. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.7.1 Description

The measure quantifies the percentage of contacts initially identified correctly that which were later misidentified. The measure is comprised of one equation-unit level:

$$\boldsymbol{IDI_m = \frac{\sum_{j=1}^{J_{T_{j,m,ID}}} j_{IDI}}{J_{T_{j,m,ID}}} \times 100\%} \quad (23)$$

Where,

IDI_m = quantifies the percentage of objects j that were initially identified correctly and later misidentified at participant m,

$j_{IDI,n}$ = object j that was initially identified correctly (and has a $T_{j,m,ID}$ value) and later misidentified (only objects with $t_{j,m,ID}$ values are assessed),

$J_{Tj,m,ID}$ = number of object j with that have a $T_{j,m,ID}$ value [see Equation (5)].

4.3.3.7.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measure IDI_m ranges in value from zero to 100%.

4.3.3.7.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI and respective ID assignments must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined.

4.3.3.7.4 Assessment Methodology

The measure IDI_m must be determined for each participating unit.

4.3.3.8 Percentage of Time Contacts were Depicted with Track ID Conflicts

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.6.7. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.8.1 Description

The measure quantifies the percentage of time when a tracked object is labeled with a conflicting ID state. This measure is comprised of two equations-instantaneous and unit level:

$$IDX_m(t) = \frac{JX_m(t)}{JT_m(t)} 100\% \qquad (24)$$

Where,

$IDX_m(t)$ = instantaneous ID conflict at participant m at time t, and

$JX_m(t)$ = number of objects tracked with ambiguous ID labels held by participant m at time t.

$JT_m(t)$ = number of objects with at least one assigned track held by participant m at time t.

$$IDX_m = \sum_{t_{start}}^{t_{end}} IDX_m(t) \qquad (25)$$

Where,

IDX_m = unit level measure of ID conflict at participant m (object weighted mean over time).

4.3.3.8.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measures $IDX_m(t)$ and IDX_m can range in value from zero to 100%.

4.3.3.8.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI and respective ID assignments must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined. A one second sampling rate is recommended for measure. An object must be in a dual track state in order for this measure to be calculated. If the dual track state was labeled with a correct and incorrect ID, then the object is in an ID conflict state.

4.3.3.8.4 Assessment Methodology

The measures $IDX_m(t)$ and IDX_m must be determined for each participating unit.

4.3.3.9 Percentage of Dual Designations

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 1.7.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.9.1 Description

A track picture that contains dual designations is considered ambiguous when more than one track is assigned to the same object. This measure quantifies the percentage of tracks that were dual designations.

The measure is assessed by calculating the percentage of time there were dual tracks (P_D) and determining the number of tracks per object, or Ambiguity (A). The CMS measures are comprised of four equations: instantaneous and unit level:

$$\boldsymbol{P_{D,m}(t) = \left(\frac{JTD(t)}{JT_m(t)}\right) \times 100\%} \quad (26)$$

Where,

$P_{D,m}(t)$ = instantaneous system dual designations at participant m at time t, and

JTD(t) = the number of tracked objects at time t with two or more assigned tracks held by participant m at time t (object count weighted at time t).

$JT_m(t)$ = number of objects with at least one assigned track held by participant m at time t.

$$P_{D,m} = \left(\frac{\sum_{t_{start}}^{t_{end}} JTD(t)}{\sum_{t_{start}}^{t_{end}} JT_m(t)}\right) \times 100\% \quad (27)$$

Where,

$P_{D,m}$ = unit level measure of dual designations at participant m (object count weighted mean over time).

$$A_m(t) = \frac{NA_m(t)}{JT_m(t)} \quad (28)$$

Where,

$A_m(t)$ = instantaneous track picture ambiguity at participant m at time t, and

$NA_m(t)$ = the number of assigned tracks held by participant m at time t.

$JT_m(t)$ = number of objects with at least one assigned track held by participant m at time t.

$$A_m = \frac{\sum_{t_{start}}^{t_{end}} NA_m(t)}{\sum_{t_{start}}^{t_{end}} JT_m(t)} \quad (29)$$

Where,

A_m = unit level measure of track picture ambiguity at participant m (object count weighted mean over time).

4.3.3.9.2 Dimension of the Measure

The measure is a ratio expressed as a percentage.

The measures $P_{D,m}(t)$ and $P_{D,m}$ values ranges are 0% to 100%.

The measures $A_m(t)$ and A_m values are greater than or equal to 1.

4.3.3.9.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined. A one second sampling rate is recommended for measure.

4.3.3.9.4 Assessment Methodology

The measures $P_{D,m}(t)$ and $P_{D,m}$ must be determined for each participating unit.

4.3.3.10 Mean Range of Target(s) from Defended Area or Asset or Ownship by Initial Engagement Ordered

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 2.1.3. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.10.1 Description

This measure quantifies the mean horizontal range between threat(s) and a defended area, asset or ownship by initial engagement time (see Figure 4-3). The term *initial engagement time* refers to the time when the first engagement order was issued at CMS to the weapons system for weapons engagement against threat. Practically, for this trial, the ownship will be used as a reference point.

The measure is will be calculated as:

$$R_{K,x} = \frac{\sum_{k=1}^{K} r_{k,x}(t_{k,EO})}{K_x} \qquad (30)$$

Where,

$R_{K,x}$ = the mean horizontal range between each threat k and point x,

x = is defined as defended area, asset location, or ownship location,

$r_{k,x}$ = the horizontal range between threat k and point x at time $t_{k,EO}$, and

K_x = total threats for which horizontal range was calculated.

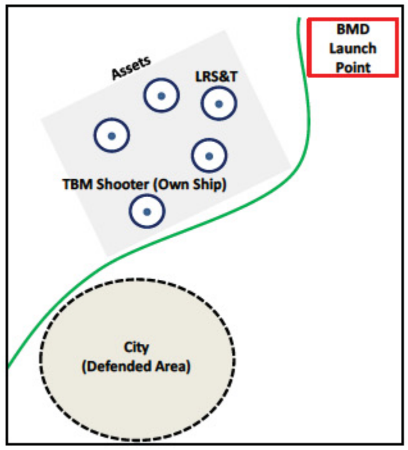


Figure 4-3: Diagram of Defended Area, Asset, and Ownship

4.3.3.10.2 Dimension of the Measure

This is a distance measure expressed in km.

The measure $R_{K,x}$ value is greater than or equal to zero.

4.3.3.10.3 Conditions of the Measure Assessment

This measure is calculated using digital data from the CMS and Truth. Defended area, asset, or ownship position are referred to as point x in the equation; point x is defined in a test objective for a specific test event. All threats k and point x must be identified and Truth data must be available to calculate this measure; however, if x is a stationary point and there is no truth data available then the latitude and longitude data is needed for x in order to calculate measure. CMS digital data is needed to determine the initial engagement order time and ownship position. Threats that are unengaged are not included in this measure. Threats that are not tracked are not included in measure.

4.3.3.10.4 Assessment Methodology

If point x is defined as ownship position, then this measure is calculated at Level 1 assessment.

If point x is defined as defended area or asset positions, then the Level 1 assessment is the input to the Level 2 assessment. This measure is calculated at the Level 2 assessment.

4.3.3.10.4.1 Level 1 Assessment Methodology

To calculate the measure, $t_{k,EO}$ is determined for each threat k for each unit using the local CMS digital data by using the initial time when the local CMS ordered weapons engagement on threat k.

If calculating $R_{K,x}$ between ownship and threat position, variable $r_{x,k}$ is calculated using the truth digital data position at time $t_{k,EO}$ for each threat k.

4.3.3.10.4.2 Level 2 Assessment Methodology

The time $t_{k,EO}$ for each threat k across the units are used to indicate when in time to calculate $R_{K,x}$ between point x equal to defended area or asset position and each threat k truth digital data position. If a threat k has multiple $t_{k,EO}$ the earliest time will be used for the measure and the additional times will be noted in the assessment.

4.3.3.11 Mean Number of Track Number Changes per Object

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 3.1.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.11.1 Description

This measure quantifies the mean number of track changes per object. Air picture is continuous (consistent with no change) if the Track Number (TN) assigned to an object does not change. There two different ways to assess the Mean Number of TN Changes and Mean Number of Unique TN Changes.

4.3.3.11.1.1 Mean Number of TN Changes

This measure is comprised of two equations-TN changes per object and unit level:

$$\boldsymbol{RTN_{j,m}} = \frac{\boldsymbol{NL_{j,m}}}{\boldsymbol{Tt_{j,m}}} \quad (31)$$

Where,

$RTN_{j,m}$ = rate of TN changes on object j from the perspective of participant m,

$NL_{j,m}$= is the number of TN segments assigned to object j which cover the period during which object j is tracked by participant m and

$Tt_{j,m}$ = is the total time object j is tracked by participant m.

The variable, $NL_{j,m}$, is determined by counting the total number of TN segments on an object regardless if the TN is repeated. For example, if a TN tracks off of object j onto a different object and in a later time the TN tracks back on to object j this would equate to two separate TN segments.

$$\boldsymbol{RTN_m} = \frac{\sum_{j=1}^{J} \boldsymbol{NL_{j,m}}}{\sum_{j=1}^{J} \boldsymbol{Tt_{j,m}}} \quad (32)$$

Where,

RTN_m = unit level rate of TN changes for participant m (object weighted mean).

4.3.3.11.1.2 Mean Number of Unique Track Number (TN) Changes

This measure is comprised of two equations-TN changes per object and unit level:

$$R_{j,m} = \frac{Nu_{j,m} - 1}{Tt_{j,m}} \quad (33)$$

Where,

$R_{j,m}$ = rate of TN changes (based on minimum number of TNs) on object j from the perspective of participant m,

$Nu_{j,m}$= is the minimum number of TNs assigned to object j which cover the period during which object j is tracked by participant m (see Figure 4-4), and

$$R_m = \frac{\sum_{j=1}^{J} Nu_{j,m} - 1}{\sum_{j=1}^{J} Tt_{j,m}} \quad (34)$$

Where,

R_m = unit level rate of TN changes (based on minimum number of TNs) for participant m (object weighted mean).

Figure 4-4 describes the measure in further detail. The total time the object is tracked ($Tt_{j,m}$) is the sum of the Green line. TN 101, TN 202, TN 302 and TN 303 are the minimum number of tracks needed to cover the period of the object tracked. The boxed areas indicate the times only one track is assigned to the object. The minimum number of tracks is used to determine the rate of TN changes.

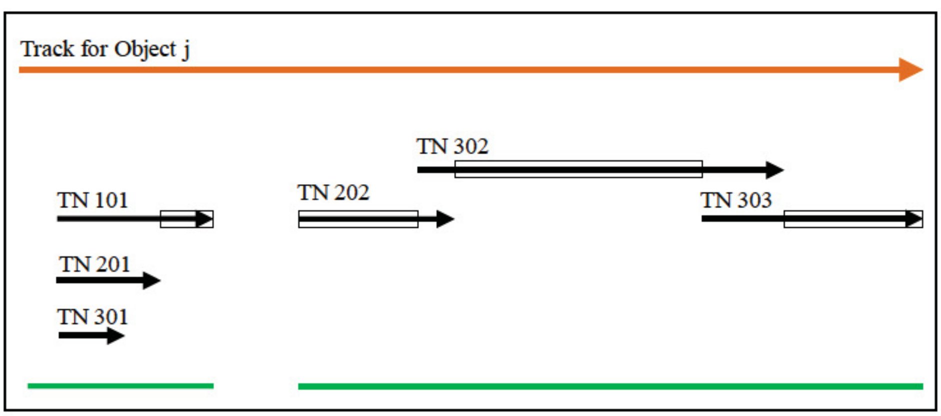


Figure 4-4: Mean Number of TN Changes per Object Example #1

The above definition is not affected by a track assignment that alternates between two possible objects. In Figure 4-5, the green line represents the time the object is tracked. The tracks that cover the period when the object is tracked are TN 101 and TN 201. In this example, the minimum number of tracks would be two.

For the Mean Number of Unique TN Changes, the number of tracks assigned and the rate of TN changes will remain the same no matter how quickly the track assignment swaps between the two closely spaced objects.

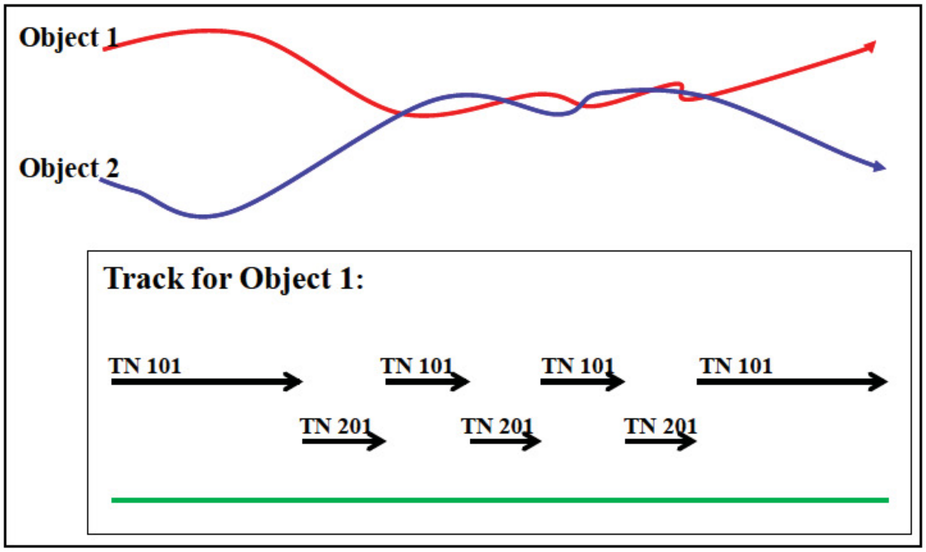


Figure 4-5: Mean Number of TN Changes per Object Example #2

4.3.3.11.2 Dimension of the Measure

The measure is rate of change that is calculated on a per hour bases (e.g., 3 TN changes/hour).

The measures $RTN_{j,m}$, RTN_m, $R_{j,m}$, and R_m, values can be greater than or equal to zero.

4.3.3.11.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined.

Mean Number of TN Changes and Mean Number of Unique TN Changes assessment must be specified.

4.3.3.11.4 Assessment Methodology

The measures $RTN_{j,m}$ and RTN_m or $R_{j,m}$ and R_m must be determined for each participating unit.

4.3.3.12 Number of Threats Engaged by More than One Firing Unit

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 3.2.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.12.1 Description

This measure quantifies the number of threats engaged by more than one firing unit. The term *threats engaged* refers to a threat being fired on, or weapons launch sequence initialized against threat k. This measure will be calculated as:

$$\boldsymbol{K_{EM}} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{k_{EM}} \qquad (35)$$

Where,

K_{EM} = number of threats engaged by more than one firing unit, and

k_{EM} = a threat k that was engaged by more than one firing unit.

4.3.3.12.2 Dimension of the Measure

This measure value is a count expressed as an integer.

The variable k_{EM} and the measure K_{EM} values can range from zero to total number of threats K.

4.3.3.12.3 Conditions of the Measure Assessment

This measure is calculated using digital data from the CMS. The threats k must be identified in order to determine K_{EM}.

4.3.3.12.4 Assessment Methodology

In order to calculate the measure, CMS digital data will be used to determine which threats where engaged k_E by each unit.

4.3.3.13 Number of Engagements Ordered Against Each Threat

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 3.2.2. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.13.1.1 Description

This measure quantifies the number of engagements ordered against each threat. The term *engagements ordered* refers to the orders issued at CMS to the weapons system for weapons engagement against a threat k. This measure is the sum of engagements ordered per threat (k_{EO}):

$$k_{EO} = \sum_{k=1}^{K} EO_k \quad (36)$$

Where,

EO_k = engagement ordered against threat k.

4.3.3.13.2 Dimension of the Measure

This measure value is a count expressed as an integer.

The measure k_{EO} value is greater than or equal to 0.

4.3.3.13.3 Conditions of the Measure Assessment

This measure is calculated using digital data from the CMS. All threats k must be identified in order to determine if engagement orders were issued against each.

The engagement order issued by voice or chat cannot be assessed using this measurement because it is not supported by digital data. An assessment of these types of engagement orders would require a tactical/operational evaluator.

4.3.3.13.4 Assessment Methodology

In order to calculate the measure, the variable EO_k must be determined for each unit by counting the engagement orders for threat k using each unit's local CMS digital data.

4.3.3.14 Percentage of Successful Engagements

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 3.3.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.14.1 Description

This measure quantifies the percentage of threats that were successfully engaged versus the total number of threats K. The term *successfully engaged* refers to a threat k being killed. This measure will be calculated as:

$$P_{SE} = \left(\frac{\sum_{k=1}^{K} k_{SE}}{K}\right) \times 100\% \quad (37)$$

Where,

P_{SE} = Percentage of successful engagements, and

k_{SE} = threat k that was successfully engaged.

4.3.3.14.2 Dimension of the Measure

This measure is a ratio expressed in a percentage.

The measure P_{SE} can range in value from zero to 100%.

4.3.3.14.3 Conditions of Measure Assessment

This measure is calculated using digital data from the CMS using a controlled test conduct. The term *controlled test conduct* refers to a firing exercise in which the following is known: firing units, number of missiles, and number of threats. The threats k must be identified in order to determine K. The first successful engagement per threat k will be used to calculate success for each target.

4.3.3.14.4 Assessment Methodology

In order to calculate the measure, variable kSE must be determined for each threat k using each unit's local CMS digital data.

4.3.3.15 Ratio of Successful Penetrations to Total Number of Threats

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 3.4.1. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.3.15.1 Description

This measure quantifies the ratio of threats that penetrated a specified area without being killed (PL). This measure will be calculated as:

$$\boldsymbol{P_L = \frac{\sum_{k=1}^{K} k_L}{K}} \qquad (38)$$

Where,

k_L = threat k that penetrated a specified area without being killed.

4.3.3.15.2 Dimension of the Measure

This measure is a ratio expressed as a real number.
The variable k_L value can range from zero to K. The measure P_L value can range from zero to one.

4.3.3.15.3 Conditions of the Measure Assessment

This measure is calculated using digital data from the CMS. The threats k must be identified in order to determine K. The specified area must be identified.

4.3.3.15.4 Assessment Methodology

The variable k_L is identified using each unit's local CMS data for each threat k. If k_L is determined for a threat k, determine if an engagement order was issued from local CMS to weapons system against threat.

4.3.4 Errors

The number or proportion of errors committed by an operator is used as an indicator of poor performance. In this study, only two error-based measures of performance are proposed: 1) number of invalid assignment of weapons (Section 4.3.4.2); and 2) number of invalid selection of Doctrines (Section 4.3.4.3).

4.3.4.1 Percentage of Time Incorrect ID is Applied to an Object

This section was inspired from the *MTMD Forum Master Test Plan* (Therrien, 2016).

4.3.4.1.1 Description

This measure quantifies the percentage of time all tracked objects are labeled incorrectly. This measure is comprised of two equations-instantaneous and unit level:

$$\boldsymbol{IDIC_m(t)} = \left(\frac{\boldsymbol{JIC_m(t)}}{\boldsymbol{JT_m(t)}}\right) \times \mathbf{100\%} \qquad (39)$$

Where,

$IDIC_m(t)$ = instantaneous ID incorrectness at participant m at time t, and

$JIC_m(t)$ = number of tracked objects with some ID labels incorrect on all tracks representing each object (as defined in the scenario) held by participant m at time t.

$JT_m(t)$ = number of objects with at least one assigned track held by participant m at time t.

$$IDIC_m = \sum_{t_{start}}^{t_{end}} IDIC_m(t) \qquad (40)$$

Where,

$IDIC_m$ = unit level ID incorrectness at participant m (object-weighted mean over time).

4.3.4.1.2 Dimension of the Measure

This measure is a ratio expressed as a percentage.

The measures $IDIC_m(t)$ and $IDIC_m$ can range in value from zero to 100%.

4.3.4.1.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI and respective ID assignments must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS. Evaluation period must be defined. A one second sampling rate is recommended for measure.

4.3.4.1.4 Assessment Methodology

The measures $IDIC_m(t)$ and $IDIC_m$ must be determined for each participating unit.

4.3.4.2 Number of Invalid Assignments

This section was taken from the *MTMD Forum Master Test Plan* (Therrien, 2016), Section 3.2.3. Force Level measures has been removed to reflect only the measures that will be used in this study.

4.3.4.2.1 Description

This measure quantifies the number of engagement orders that were assigned to non-threats. The term *engagement orders* refers to when the CMS ordered weapons engagement on a track. This measure is the sum of engagements ordered against non-threats (IA):

$$\boldsymbol{IA} = \sum_{\boldsymbol{nk=1}}^{\infty} \boldsymbol{EO_{nk}} \qquad (41)$$

Where,

EO_{nk} = engagement ordered against non-threats (nk).

4.3.4.2.2 Dimensions of the Measure

This measure value is a count expressed as an integer.

The measure IA is greater than or equal to 0.

4.3.4.2.3 Condition of Measure Assessment

This measure is calculated using digital data from the CMS. In order to determine which objects are non-threats nk all threats k must be identified.

The engagement order issued by voice or chat cannot be assessed using this measure because it is not supported by digital data. An assessment of these types of engagement orders would require a tactical/operational evaluator.

4.3.4.2.4 Assessment Methodology

In order to calculate the measure, the variable EO_{nk} must be determined for each unit by counting the engagement orders on non-threat nk using each unit's local CMS digital data.

4.3.4.3 Number of Invalid Doctrine

This section was inspired from the *MTMD Forum Master Test Plan* (Therrien, 2016).

4.3.4.3.1 Description

This measure quantifies the number of Doctrine wrongly selected[6] (DS) against each missile. This measure is the sum of Doctrine selected per missile (ms_{DS}):

$$\boldsymbol{ms_{DS}} = \sum_{\boldsymbol{ms=1}}^{\boldsymbol{MS}} \boldsymbol{DS_{ms}} - \mathbf{1} \quad (42)$$

Where,

DS_{ms} = Doctrine selected against missile ms.

4.3.4.3.2 Dimension of the Measure

This measure value is a count expressed as an integer.

The measure ms_{DS} value is greater than or equal to 0.

4.3.4.3.3 Conditions of the Measure Assessment

This measure is calculated using digital data from the CMS. All missiles ms must be identified in order to determine if more than one Doctrine was selected against each.

4.3.4.3.4 Assessment Methodology

In order to calculate the measure, the variable DS_{ms} must be determined for each unit by counting the Doctrines selected for missile ms using each unit's local CMS digital data.

[6] There are three ways Doctrine might be considered 'selected'. The operator first selects the Doctrine drop-down menu, then selects a Doctrine by left clicking on it, and finally approves the Doctrine by clicking on 'Approve'. A Doctrine may generally be considered as 'selected' at all of these three steps. For this specific measure however, a Doctrine will be considered 'selected' when the operator left clicks on a specific Doctrine.

4.3.5 Detection Rate

Detection rate is the proportion of stimuli that are detected by an operator during a pre-defined period of time. In this study, the detection-based metric is used to evaluate the ability of the combat team to detect the contacts present in a AOI. They do not take into account the identities of objects.

4.3.5.1 Percentage of Tracks Detected

This section was inspired from the *MTMD Forum Master Test Plan* (Therrien, 2016).

4.3.5.1.1 Description

This measure quantifies the percentage of tracks that were detected and tracked within an AOI by each unit locally at the CMS. The CMS measure is comprised of one equation:

$$P_T = \left(\frac{Jt}{J}\right) \times 100\% \quad (43)$$

Where,

> $P_{T,}$ = percentage of tracks that were detected and tracked within an AOI; and
> Jt= number of objects with at least one assigned local track

4.3.5.1.2 Dimension of the Measure

The measure is a ratio expressed as a percentage.

The measure values ranges are 0% to 100%.

4.3.5.1.3 Conditions of the Measure Assessment

This measure is calculated using digital data from CMS and truth. The OOI must be known; track reconstruction of OOI must be completed. Measure calculation is based on the entire track picture (local, remote, and mutual) at participating units' CMS.

4.3.5.1.4 Assessment Methodology

The measures Jt and P_T must be determined at the unit level.

4.4 Subjective Measures

In addition to the objective measures described above, a set of subjective measures is required to complement the measurement of performance of the operators. The subjective measures chosen aim at assessing slightly different aspects of the work than the objective measures. The subjective measures are described below.

4.4.1 Subject Matter Expert Ratings

SME ratings of performance will be used where objective measures are not available, or as a complement to objective measures. As SME ratings are qualitative in nature, they are more suited for the assessment of soft skills such as communication, collaboration, decision making, and so on.

Since scorers are reluctant to use extremes of a rating scale, Behaviorally Anchored Rating Scales (BARS), where each point on the scale is assigned a verbal description, are recommended for ratings and will be used in this study.

For the current project, SME ratings will be used to assess the quality of engagement plan, alterations of the engagement plan, preparation done ahead for possible engagement, prioritization of tracks, information flow, and errors in Doctrine selection. *SME ratings will be limited to consideration of performance effects that are mediated by the crew configuration*. The coding will occur both online, during the scenario, and afterward (if necessary), through video recording of the simulation. In due course, specific categories will be developed, based on a thorough analysis of the validation scenarios.

4.4.2 Workload

Workload refers to the 'effort' (physical and mental) a person needs to expend to maintain a level of performance on a task. Workload may impact human performance through both overload and underload of work. In this study, workload will be measured using the well-known NASA TLX (see Section 4.4.2.1) and the ISA (see Section 4.4.2.2).

4.4.2.1 NASA TLX

The NASA TLX (Hart & Staveland, 1988) is a multidimensional retrospective subjective workload assessment scale designed to obtain workload estimates from one or more operators during pauses in task performance or immediately afterwards. It provides a weighted mean of 6 workload subscales:

- **Mental Demand** – how much mental demand and perceptual activity was required? Was the task easy or demanding, simple or complex, exacting or forgiving?

- **Physical Demand** – how much physical activity was required? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

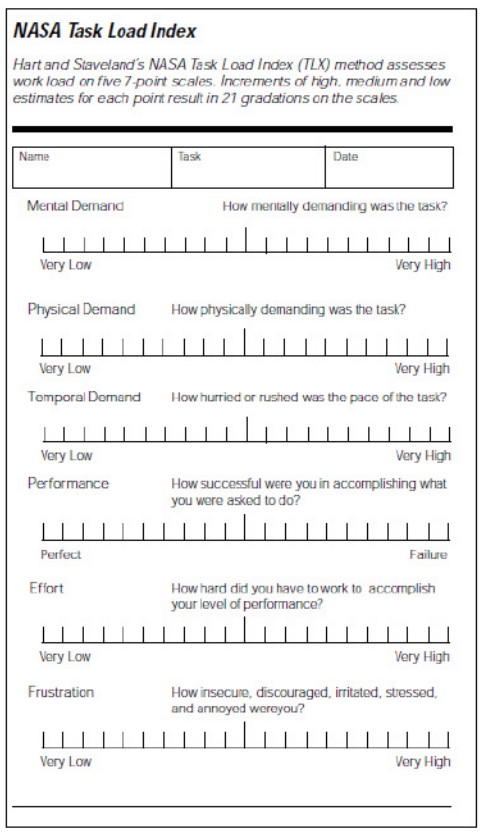
- **Temporal Demand** – how much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

- **Effort** – how hard did you have to work (mentally and physically) to accomplish your level of performance?

- **Performance** – how successful do you think you were in accomplishing the goals of the task set by the analyst (or yourself)? How satisfied were you with your performance in accomplishing these goals?

- **Frustration level** – how insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

For each of the subscales, the participants are asked to rate their workload for the preceding period of time based upon an interval scale divided into 20 intervals, ranging from low (1) to high (20). The NASA TLX is applied both during and after the trial period (see Figure 4-6). The second part of the NASA TLX results in an individual weighting for each subscale by asking subjects to select the subscale with the greatest impact on workload in a systematic pairwise consideration of the subscales (i.e. 15 pairwise comparisons). This procedure accounts for differences in the sources of workload between tasks and differences in workload definition between raters.


NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?
Very Low — Very High

Physical Demand — How physically demanding was the task?
Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?
Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?
Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?
Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?
Very Low — Very High

Figure 4-6: NASA TLX (NASA, 1986)

4.4.2.2 Instantaneous Self-Assessment

ISA is a unidimensional measure of workload, i.e., it attempts to elicit a measure of overall workload. In ISA, the operator rates their perceived workload during real-time simulated or actual tasks. Operator gives a rating of perceived workload, or “busyness”, on a scale of 1 - very low to 5 - very high (see Figure 4-7). ISA is a simple and rapid measure of workload that is useful for a rough indication of workload in a dynamic environment.

Level	Workload	Spare Capacity	Description
1	Under-utilised	Very much	Little or nothing to do. Rather boring
2	Relaxed	Ample	More time than necessary to complete the tasks. Time passes slowly.
3	Comfortable	Some	The controller has enough work to keep him/her stimulated. All tasks are under control.
4	High	Very little	Certain non-essential tasks are postponed. Could not work at this level very long. Controller is working ‘at the limit’. Time passes quickly.
5	Excessive	None	Some tasks are not completed. The controller is overloaded and does not feel in control.

Figure 4-7: ISA (Brennan, 1992)

ISA ratings are capture via discrete button presses or a ‘slider’ to permit the operator to leave the slider in position to indicate unchanged workload from the last rating, or to make changes to the slider position of a magnitude that represents their perceived change in their workload over the latest interval. The operator is cued to make a workload rating by a flashing Light Emitting Diode (LED) or an audible reminder. Portable data capture units have been developed for use in helicopters and could be used on the X-Ship, placed discretely in the periphery of the operator’s field of view.

ISA permits analysis to correlate workload with the scenario and performance at a greater level of sensitivity than is possible using the TLX. This is primarily because, with familiarity, ISA is a realtime measure of workload that is unobtrusive to operators.

4.4.3 Post-Measurement Period Interview and Debrief

A short interview with each operator at the conclusion of a measurement period will be used to capture their opinions on the crew configurations they experienced and how well they suited the level of demand imposed by the scenario. The goal is to obtain feedback on the perceived validity of the crewing level and whether they think adjusting the crewing level with workload is both feasible and desirable. The list of exact questions will be developed in due course based on an analysis of the validation scenarios.

4.5 Summary of MOPs

Table 4-4 presents a summary of the MOPs that will be used during this trial. Each MOP is described and categorised. The data sources required to develop the MOP and the KPEs of interest are listed, followed by the system events that will generate the KPE (this is still to be developed based on detailed analysis of the scenario). Finally, the hypotheses that will be answered, in part, by the MOP are listed.

4.6 Level of Performance Required

Level of performance with respect to time and accuracy required for D2E in different mission situations will be developed through review of CFCD106. This information will be classified and will be developed in a subsequent contract.

Table 4-4: Summary of MOPs

MOP	Description	Type	Data Source	KPEs	System Events	Related Hypothesis
Percentage of Objects Detected and Tracked	The percentage of time an object was detected and tracked within an Area of Interest (AOI) by each unit locally at the CMS	Objective: Accuracy	CMS and truth	Detection of the contact by the system; Detection of the contact by the operator	TBD	H(NESOP)1; H(NCIOP)1; H(I)1; H(I)2; H(I)3; H(T)1; H(T)2; H(T)3
Mean Time from Entry into Battlespace to Initial Detection	Difference in time when an object entered the defined battlespace and when the participant(s) first detected it at the CMS	Objective: Response Time	CMS and truth	Detection of the contact by the system; Detection of the contact by the operator	TBD	H(NESOP)1; H(NCIOP)1; H(I)1; H(I)2; H(I)3; H(T)1; H(T)2; H(T)3
Mean Time between Initial Detection and Correct Identification	The mean time it took from the initial detection of an object to correctly identifying that object at the CMS	Objective: Response Time	CMS and truth	Detection of the contact by the operator; Identification of the contact by the operator	TBD	H(NESOP)1; H(NCIOP)1; H(I)1; H(I)2; H(I)3; H(T)1; H(T)2; H(T)3

MOP	Description	Type	Data Source	KPEs	System Events	Related Hypothesis
Mean Time to Engage Threat	The mean time when a threat enters the Area of Interest (AOI), to when the first engagement is ordered	Objective: Response Time	CMS and Truth	Identification of the contact by the operator; Engagement of the contact by the operator	TBD	H(Dir)1; H(Dir)2; H(NESOP)1; H(NESOP)2; H(C)1; H(C)2; H(C)3; H(T)1; H(T)2; H(T)3
Mean Time to Apply Doctrine	The mean time when a missile enters the AOI, to when the Doctrine is loaded	Objective: Response Time	CMS and Truth	Identification of the contact by the operator; Engagement of the contact by the operator	TBD	H(Dir)1; H(Dir)2; H(NESOP)1; H(NESOP)2; H(C)1; H(C)2; H(C)3; H(T)1; H(T)2; H(T)3
Duration of Dual (Multiple) Designations	Duration of time a dual (multiple) track (s) is/are held for an object	Objective: Duration	CMS and Truth	Detection of the contact by the operator; Classification of the contact by the operator	TBD	H(NESOP)1; H(NCIOP)1; H(I)1; H(I)2; H(I)3;
Percentage of False Tracks	Percentage of false tracks	Objective: Accuracy	CMS and truth	Detection of the contact by the system	TBD	H(NESOP)1; H(NCIOP)1; H(I)1;

MOP	Description	Type	Data Source	KPEs	System Events	Related Hypothesis
	present in a track picture					H(I)2; H(I)3
Positional Accuracy	Position accuracy of each assigned track compared to the object it is associated to	Objective: Accuracy	CMS and truth	Detection of the contact by the system; Detection of the contact by the operator; Classification of the contact by the operator	TBD	H(NESOP)1; H(NCIOP)1; H(I)1; H(I)2; H(I)3
Track Completeness	Percentage of time an object was tracked and reported for the coalition force	Objective: Accuracy	CMS and truth	Identification of the contact by the operator	TBD	H(NCIOP)2; H(I)1; H(I)2; H(I)3
ID Completeness	Percentage of time an object is identified	Objective: Accuracy	CMS and truth	Identification of the contact by the operator	TBD	H(NCIOP)2; H(I)1; H(I)2; H(I)3
Percentage of Time Correct ID is Applied to an Object	Percentage of time all tracked objects are labeled correctly	Objective: Accuracy	CMS and truth	Identification of the contact by the operator	TBD	H(NCIOP)2; H(I)1; H(I)2; H(I)3
Percentage of Contacts Initially Identified Correctly Which	The percentage of air and space vehicles initially identified correctly	Objective: Accuracy	CMS and truth	Identification of the contact by the operator	TBD	H(NCIOP)2; H(I)1;

MOP	Description	Type	Data Source	KPEs	System Events	Related Hypothesis
Were Later Misidentified	that which were later misidentified					
Percentage of Time Contacts were Depicted with Track ID Conflicts	The percentage of time when a tracked object is labeled with a conflicting ID state	Objective: Accuracy	CMS and truth	Identification of the contact by the operator	TBD	H(NCIOP)2; H(I)1;
Percentage of Dual Designations	Percentage of tracks that were dual designations	Objective: Accuracy	CMS and truth	Identification of the contact by the operator	TBD	H(NCIOP)1; H(I)1; H(I)2; H(I)3
Mean Range of Target(s) from Defended Area or Asset or Ownship by Initial Engagement Ordered	Mean horizontal range between threat(s) and a defended area, asset, or ownship by initial engagement time	Objective: Accuracy	CMS and Truth	Detection of the contact by the operator; Engagement of the contact by the operator	TBD	H(NESOP)1; H(NESOP)2
Mean Number of TN Changes per Object	Quantifies the mean number of track changes per object	Objective: Accuracy	CMS and truth			H(NESOP)1; H(NESOP)2
Number of Threats Engaged By More Than One Firing Unit	The number of threats engaged by more than one firing unit	Objective: Accuracy	CMS	Engagement of the contact by the operator	TBD	H(NESOP)1; H(NESOP)2; H(C)1; H(C)2; H(C)3

MOP	Description	Type	Data Source	KPEs	System Events	Related Hypothesis
Number of Engagements Ordered Against Each Threat	The number of engagements ordered against each threat the number of engagements ordered against each threat	Objective: Accuracy	CMS	Engagement of the contact by the operator	TBD	H(NESOP)1; H(NESOP)2; H(C)1; H(C)2; H(C)3
Number of Invalid Assignments	The number of engagement orders that were assigned to non-threats	Objective: Error	CMS	Engagement of the contact by the operator	TBD	H(Error); H(NESOP)1; H(NESOP)2; H(C)1; H(C)2; H(C)3
Number of Invalid Doctrine	The number of Doctrine wrongly selected (DS) against each missile	Objective: Error	CMS	Engagement of the contact by the operator	TBD	H(Error); H(NESOP)1; H(NESOP)2; H(C)1; H(C)2; H(C)3
Percentage of Successful Engagements	The percentage of threats that were successfully engaged versus the total number of threats K	Objective: Accuracy	CMS	Engagement of the contact by the operator; Contact is no longer in the system	TBD	H(NESOP)1; H(NESOP)2; H(C)1; H(C)2; H(C)3
Ratio of Successful Penetrations to	The ratio of threats that penetrated a specified area	Objective: Accuracy	CMS	Engagement of the contact by the operator;	TBD	H(Dir)1; H(NESOP)1; H(NESOP)2; H(C)1; H(C)2;

MOP	Description	Type	Data Source	KPEs	System Events	Related Hypothesis
Total Number of Threats	without being killed (PL)			Contact is no longer in the system		H(C)3
Subject Matter Expert ratings	Quality of engagement plan, alteration of engagement plan, preparation, Prioritization of tracks, Information flow	Subjective: Rating scale	Live observation	NA	NA	H(C)1; H(C)2; H(C)3 H(NCIOP)2; H(NESOP)1; H(NESOP)2; H(Sonar)1
NASA TLX	Multidimensional workload assessment scale	Subjective: Questionnaire	Paper-based	NA	NA	H(WL)1; H(WL)2
Instantaneous Self-Assessment	Unidimensional measure of workload	Subjective: Questionnaire	Paper-based	NA	NA	H(WL)1; H(WL)2
Post-measurement session Interview and Debrief	Capture various impressions of the overall process and to debrief the participants	Subjective: Interviews	Interview	NA	NA	H(C)1; H(C)2; H(C)3 H(NCIOP)2; H(NESOP)1; H(NESOP)2
Percentage of Time Incorrect ID is Applied to an Object	The percentage of time all tracked objects are labeled incorrectly	Objective: Error	CMS	Identification of the contact by the operator	TBD	H(NESOP)1; H(NCIOP)1; H(I)1; H(I)2; H(I)3

MOP	Description	Type	Data Source	KPEs	System Events	Related Hypothesis
Percentage of Tracks Detected	The percentage of tracks that were detected and tracked within an AOI by each unit locally at the CMS	Objective: Detection Rate	CMS and truth	Detection of the contact by the system; Detection of the contact by the operator	TBD	H(NESOP)1; H(NCIOP)1; H(I)1; H(I)2; H(I)3; H(T)1; H(T)2; H(T)3

5 ANALYSIS

To this point, the validation trial plan has dealt with planning, preparation, execution and data collection. The analysis of the data will be particularly daunting. Although this plan lays out how objective and subjective measures will be used to answer hypotheses, and specifically what those objective and subjective MOEs and MOPs are, the practicalities of analysis have not been described.

Based on the scenario, the first analysis to be undertaken will be a ground truth analysis. This analysis will list all contacts, their identity, their location, speed and course. On the basis of this analysis, specific scenarios (or portions of scenarios) will be selected for use during the trial. Any scripts describing what the contacts should do and injects the role players should make will also be considered. The resulting narrative for each Contact of Interest (COI) will comprise the scenario event component of the KPEs. This analysis will be used to compare with the non audio-video data to support the generation of MOPs. This analysis has not been completed to date and is anticipated to be carried out in spring and summer 2017.

Based on the pilot test, the second analysis to be undertaken will be the non audio-video analysis. This will be based on known and noted inputs against known and noted contacts in the scenario. The non audio-video data will be downloaded and reviewed with scenario contacts being correlated with the data log and all actions on those contacts being summarised and characterised. These comprise the system event component of the KPEs. The process by which contacts are identified and their evolution followed in the non audio-video data (i.e. by system events related to specific track numbers) will be documented forming a detailed data preparation procedure. Developing a chronology of all actions performed by operators on a contact will be necessary to compare with ground truth to identify when or if any contacts are misidentified. This analysis has not been completed to date and is anticipated to be carried out in autumn 2017.

The combination of scenario- and system-based KPEs will be used to develop a detailed analysis plan describing how KPEs will be used to generate the MOPs which will be used to accept or reject the hypotheses and, ultimately, answer the research question. More details regarding the analysis will be provided after ground truth analysis, pilot test and tabletop exercise have occurred.

6 CONCLUSION

This document describes the validation trial plan for the X-Ship to answer the question:

> *Can a reduced operator complement permit the required objective level of performance for the combat department to be attained during all missions and mission phases on the post-HCM HALIFAX Class frigate?*

There remains much detailed work to be completed in order to be fully prepared to run this validation trial. However, this plan represents a roadmap to the activities that need to occur before the trial can be run. A tentative schedule has been created with activities occurring in the second half of 2017 (i.e., pilot test) and the second half of 2019 (i.e., validation trial on X-Ship). Outside of these two significant milestones, effort will be directed to analyses, rehearsal and preparation.

7 REFERENCES

Baker, K., & Banbury, S. (2011). Halifax Class Modernization Program – Human Engineering Design Approach Document – Operator (HEDAD-O). Report for Lockheed-Martin, Ottawa, Document No. 1006-046-02.

Benaskeur, A. R., Rhéaume, F., & Paradis, S. (2007). Target Engageability Improvement through Adaptive Tracking. *J. Adv. Inf. Fusion*, *2*(2), 99-112.

Brennan. S.D. (1992). An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth.

Dubé, G. & Lamoureux, T.M. (2017). HALIFAX Class Modernization Automation Impact on Crew (3) Experimental Plan - SEAWOLF. Report No. 6010-006-01.

Dubé, G. & Lamoureux, T. (2016a). HALIFAX Class Modernization (HCM), Automation Impact on Crew II.

Dubé, G. & Lamoureux, T. (2016b). Impact of HCM Automation (2) – Development of a Research Programme to Support Naval Personnel and Training [CLASSIFIED CONFIDENTIAL]. Report to Defence Research and Development Canada – Toronto, CAE Contract Report 5879-001.

Dubé, G., Lawrynczyk, A., & Lamoureux, T. (2016). Halifax Class Modernization automation impact on crew III: Interim report on experimental planning.

Gauthier, Y., Bourdon, S., Dore, S. & Fong, V. (2004). Defining and selecting metrics for intelligence, surveillance, and reconnaissance (ISR). Directorate of Operational Research (Maritime, Land & Air), DOR(MLA) Research Note RN2004/08.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139-183). North-Holland.

Helleur, C., Mathews, M., Kashyap, N., & Rafuse, J. (2007, July). Track-to-track fusion by a Human Operator for Maritime Domain Awareness. In Information Fusion, 2007 10th International Conference on (pp. 1-8). IEEE.

Ho, G., Chow, R., & Crebolder, J. (2015). "Impact of Automation on Crew Performance - Scoping Study Review", Scientific Letter, DRDC-RDDC-2015-L302.

Kusche, J. (2016). Image IS15-2016-003-046, 31 October 2016. Downloaded from www.combatcamera.forces.gc.ca, 21 September 2018.

Lawrynczyk, A., Lamoureux, T., & Dubé, G. CAE Inc., HALIFAX Class Modernization (HCM), Automation Impact on Crew, DRDC-RDDC-2016-C021, 2016.

Matthews, M.L., Keeble, A.R., Bruyn, L. E., Webb, R.D., & Lamoureux, T.M. (2004). *Development of Measures of Performance for Evaluating the COMDAT Technology Demonstrator: Potential Use of Training Records from the Operations Room Team Trainer (ORTT)* (No. 7703-02). Humansystems INC Guelph (Ontario).

Matthews, M.L. & Webb, R.D.G. (2000). Validation of Cognitive Task Analysis (CTA) for the HALIFAX-Class ORO. Defence R&D Canada, unpublished contract report.

Matthews, M.L., Webb, R.D.G., & Keeble, A.R. (2002). Assessing the Impact of Multi-Sensor Data Fusion on Command and Control Operations in the Halifax Class Frigate: Recommendations for Measures of Performance and Detailed Test Plan. DRDC Toronto CR-2002-018.

Therrien, P. (2016). Master Test Plan: MTMD Forum. Interim Version 3.2, 12 December 2016.

APPENDIX A SUB-HYPOTHESES

NOTE: In Table A-1, many hypotheses end with a bracket containing some combination of “L” and a number. This refers to the step (level) in the incremental plan for increasing the crew complement in the operations room.

Table A-1: List of Sub-Hypotheses

Time-Based Hypotheses
H(T)1: Observed performance times for identified tasks/task sequences will meet or improve upon RCN standards.
H(T)2: The time required to perform tasks, respond to contacts, respond to threats will not increase with any level of intensity or complexity.
H(T)3: The time required to perform tasks, respond to contacts, respond to threats will not increase over the duration of a deployment.
NESOP-Based Hypotheses
H(NESOP)1: Sensors
H(NESOP)1.1: The SWC can meet the objective operational performance requirement for sensor employment (ESM/CIO) with no assistance (L1).
H(NESOP)1.1: The EWS can meet the objective operational performance requirement for sensor employment (ESM/CIO) with no assistance (L2).
H(NESOP)1.2: The EWS can meet the objective operational performance requirement for sensor employment (ESM/CIO) with assistance from ESM (L3).
H(NESOP)1.3: The EWS and ESM can meet the objective operational performance requirement for sensor employment (ESM/CIO) with assistance from FCS (L4).
H(NESOP)1.4: The EWS, ESM and FCS can meet the objective operational performance requirement for sensor employment (ESM/CIO) with assistance from FCO(1) (L5).
H(NESOP)1.5: The EWS, ESM, FCS and FCO(1) can meet the objective operational performance requirement for sensor employment (ESM/CIO) with assistance from CIO (L6).
H(NESOP)1.6: The EWS, ESM, FCS, FCO(1) and CIO can meet the objective operational performance requirement for sensor employment (ESM/CIO) with assistance from FCO(2) (L7).
H(NESOP)2: Counter Measures
H(NESOP)2.1: The SWC can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with no assistance (L1).
H(NESOP)2.1: The EWS can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with no assistance (L2).
H(NESOP)2.2: The EWS can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with assistance from ESM (L3).
H(NESOP)2.3: The EWS and ESM can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with assistance from FCS (L4).

H(NESOP)2.4: The EWS, ESM and FCS can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with assistance from FCO(1) (L5).
H(NESOP)2.5: The EWS, ESM, FCS and FCO(1) can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with assistance from CIO (L6).
H(NESOP)2.6: The EWS, ESM, FCS, FCO(1) and CIO can meet the objective operational performance requirement for countermeasure employment (MASS/RAMSES/57mm/CIWS) with assistance from FCO(2) (L7).
NCIOP-Based Hypotheses
H(NCIOP)1: Sensors
H(NCIOP)1.1: The TS can meet the objective operational performance requirement for radar utilisation with no assistance (L1).
H(NCIOP)1.2: The TS can meet the objective operational performance requirement for radar utilisation with assistance from ARRO (L2).
H(NCIOP)1.3: The TS and ARRO can meet the objective operational performance requirement for radar utilisation with assistance from ASPO (L4).
H(NCIOP)1.4: The TS, ARRO and ASPO can meet the objective operational performance requirement for radar utilisation with assistance from C4I Manager (L5).
H(NCIOP)2: Track Management
H(NCIOP)2.1: The TS can meet the objective operational performance requirement for track management (LINK, GCCS, AIS, IFF) with no assistance (L1).
H(NCIOP)2.2: The TS can meet the objective operational performance requirement for track management (LINK, GCCS, AIS, IFF) with assistance from ARRO (L2).
H(NCIOP)2.3: The TS and ARRO can meet the objective operational performance requirement for track management (LINK, GCCS, AIS, IFF) with assistance from ASPO (L4).
H(NCIOP)2.4: The TS, ARRO and ASPO can meet the objective operational performance requirement for track management (LINK, GCCS, AIS, IFF) with assistance from C4I Manager (L5).
H(NCIOP)3: The SAC can effectively manage the helicopter resources (L1)
H(NCIOP)4: Information Flow
H(NCIOP)4.1: The ORS and IWD can effectively facilitate the timely and accurate flow of information around the Ops Room with no assistance (L1).
H(NCIOP)4.2: The ORS and IWD can effectively facilitate the timely and accurate flow of information around the Ops Room with the assistance of the C4I Manager (L5).
SONAR-Based Hypotheses
H(UW)1:The ASWC can meet the objective operational performance requirement for UWW (sensors and weapons) with no assistance from subordinates (L1).
H(UW)2:The ASWC can meet the objective operational performance requirement for UWW (sensors and weapons) with assistance from the SCS (L2).

H(UW)3:The ASWC can meet the objective operational performance requirement for UWW (sensors and weapons) with assistance from the SCS and ASPO (L4).
Director-Based Hypotheses
H(Dir)1: Planning & Decision Making
H(Dir)1.1: The SWC has the information required to support effective planning and decision making (L1).
H(Dir)1.2: The ASWC has the information required to support effective planning and decision making (L1).
H(Dir)1.3: The ORO has the information required to support effective planning and decision making (L1.
H(Dir)1.4: The IWD has the information required to support effective planning and decision making (L1).
H(Dir)2: Execution
H(Dir)2.1: The SWC has the means to act upon their decision (L1).
H(Dir)2.1: The IWD has the means to act upon their decision (L1).
H(Dir)2.3: The ORO has the means to act upon their decision (L1).
H(Dir)2.2: The ASWC has the means to act upon their decision (L1).
Identification-Based Hypotheses
H(I)1: Air
H(I)1.1: Sensors are used effectively to detect contacts in the area of responsibility.
H(I)1.2: Sensors are used effectively to geolocate contacts in the Area of Responsibility.
H(I)1.3: Sensors are used effectively to derive kinematic information about contacts in the Area of Responsibility.
H(I)1.4: Other sources of information are used effectively to amplify the identity of contacts in the Area of Responsibility.
H(I)2: Surface
H(I)2.1: Sensors are used effectively to detect contacts in the area of responsibility.
H(I)2.2: Sensors are used effectively to geolocate contacts in the Area of Responsibility.
H(I)2.3: Sensors are used effectively to derive kinematic information about contacts in the Area of Responsibility.
H(I)2.4: Other sources of information are used effectively to amplify the identity of contacts in the Area of Responsibility.
H(I)3: Subsurface
H(I)3.1: Sensors are used effectively to detect contacts in the area of responsibility.
H(I)3.2: Sensors are used effectively to geolocate contacts in the Area of Responsibility.

H(I)3.3: Sensors are used effectively to derive kinematic information about contacts in the Area of Responsibility.
H(I)3.4: Other sources of information are used effectively to amplify the identity of contacts in the Area of Responsibility.
Countermeasure-Based Hypotheses
H(C)1: Countermeasures are used in a timely, effective and accurate manner to defeat an air threat.
H(C)2: Countermeasures are used in a timely, effective and accurate manner to defeat a surface threat.
H(C)3: Countermeasures are used in a timely, effective and accurate manner to defeat a sub-surface threat.
Countermeasures includes mostly weapons, but could also include fire control radar, underwater telephone, etc.
Defeat means not only destroy, but also dissuade from a course of action or scare off.
Dividing into domains means that overlapping combos of countermeasures could be used in different domains. We still need to identify measures to assess countermeasure employment, including the construction of scenarios wherein we expect Directors to REMOVE countermeasures from an engagement to protect against another potential threat.
We should be on particular lookout for situations where the Director calls upon one of the outstations to do something with the weapons. For the SWC it would be the FCO/FCS/EWS. For the ASWC it would most likely be a call to someone physically in the torpedo locker.
Another thing to look out for is the instance when one of the directors actually splits their warfare domain(s) and gives responsibility to the ORO.
Workload-Based Hypotheses
H(WL)1: With the crew configuration that permitted the objective level of performance to be attained, workload was not judged to be excessive.
H(WL)2: The level of workload associated with the objective level of performance can be maintained for the full duration of a deployment Potential hypothesis, if specific hypotheses for DF, TEWA and Doctrine are developed: After doctrine is applied and the threat is subsequently finished, workload for all front row operators increases as they reset sensor settings. OR they just reapply MSD (and then start making their incremental changes). OR they do not make any changes to sensor settings.
Error-Based Hypotheses
H(E)1: the observed error rate does not increase with any level of intensity or complexity.
H(E)2: the observed error rate does not increase over the duration of a deployment.

APPENDIX B CREWING LEVELS

			SENSORS											COUNTERMEASURES				WEAPONS					C2/COMMS	
			AW	AW	EW	UW	EW	EW	UW	UW	UW	EW	EW	SWC	SWC	SWC	ASWC	SWC	SWC	SWC	SWC	ASWC	All	AW
		ROLE	3D Volume Search Radar	2D Redundancy Radar	CEROS 200 Fire Control Radar	Navigation Radar - NWS 36S and X Band Radars	SEOSS	SIRIUS Long Range IR Search and Track System	AN/SQS-510 SONAR	AN/SQR-501 CANTASS SONAR	Sonobuoys	Elisra EW System	AN/SRD-504	MK 137 Mod 7 EW System	Multi Ammunition Softkill System (MASS) EW Launchers	AN/SLQ-503 (RAMSES)	AN/SLQ-25 Decoy System	CIWS 1B Close In Weapons Systems	57mm Gun	Sea Sparrow - ESSM	Harpoon	MK-46 Torpedo	Tactical Digital Information Link	GCCS-M
1		ORO																						
1		IWD																						
1	Radar	TS (air/surface)	X	X		X																	X	
2		ARRO	Y																				Y	
1		ORS																						X
4		ASPO				A																	A	
5		C4I Mgr																					B	B
1		SAC	X																					
1	Electronic	SWC			X		X	X				X	X	X	X	X		X	X	X	X			
2		EWS			Y		Y	Y				Y	Y											
3		ESM										Z												
4		FCS			A		A	A																
5		FCO (1)			B		D																	
6		CIO											C											
7		FCO (2)			D			D																

			SENSORS											COUNTERMEASURES				WEAPONS					C2/COMMS	
			AW	AW	EW	UW	EW	EW	UW	UW	UW	EW	EW	SWC	SWC	SWC	ASWC	SWC	SWC	SWC	SWC	ASWC	All	AW
		ROLE	3D Volume Search Radar	2D Redundancy Radar	CEROS 200 Fire Control Radar	Navigation Radar - NWS 36S and X Band Radars	SEOSS	SIRIUS Long Range IR Search and Track System	AN/SQS-510 SONAR	AN/SQR-501 CANTASS SONAR	Sonobuoys	Elisra EW System	AN/SRD-504	MK 137 Mod 7 EW System	Multi Ammunition Softkill System (MASS) EW Launchers	AN/SLQ-503 (RAMSES)	AN/SLQ-25 Decoy System	CIWS 1B Close In Weapons Systems	57mm Gun	Sea Sparrow - ESSM	Harpoon	MK-46 Torpedo	Tactical Digital Information Link	GCCS-M
1	Sonar	ASWC															X					X		
2		SCS				Y																		
1		HMS							X															
1		TAS (1)								X														
2		TAS (2) + TAS Sup								Y														
1		SPS									X													

APPENDIX C DETECT-TO-ENGAGE MODEL

The D2E model (e.g., Matthews & Webb, 2000; Matthews, Keeble, Bruyn, Webb & Lamoureux, 2004; Benaskeur, Rhéaume & Paradis, 2007) describes the sequential process through which an operator progresses in a warfighting scenario. Typically, D2E involves Detection, Classification, Recognition, Identification and Action, although the precise terminology and stages described will vary from service to service and from nation to nation. The D2E model accords with the "input-processing-output" (IPO), Human Information Processing (HIP), observe-orient-decide-act (OODA) and Situation Awareness (SA) models in that it refers to an input stage (Detection), a processing stage (Classification, Recognition, Identification) and an output stage (Action). Unlike HIP and OODA, however, D2E does not make explicit the selection of an appropriate output, although operators will understand that this is encompassed within the D2E sequence. During the voyage on HMCS Charlottetown operators were explicitly asked whether they understood the term "Detect-to-Engage" and all expressed their familiarity with it.

As with the other models that informed the approach, the D2E assisted in the development of questions as well as the interpretation of responses and the formulation of research questions. The D2E model will continue to be relevant as the research programme, with associated conditions and metrics, is developed to support naval personnel and training interests.

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.) CAE Inc., 1135 Innovation Drive, Ottawa, ON, K2K 3G7, 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.)

Halifax Class Modernization Automation Impact on Crew (3) Validation Trial Plan—HMCS Montreal

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

Lawrynczyk, A.

5. DATE OF PUBLICATION (Month and year of publication of document.) November 2018	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.) 94	6b. NO. OF REFS (Total references cited.) 18

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

Contract 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
Toronto, Ontario M3K 2C9
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.) Impact of Automation on Crew Performance	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.) W7707-145734
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.) DRDC-RDDC-2018-C187	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

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 (Approval for further dissemination of the document. Security classification must also be considered.)

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

Automation, Naval Crewing, Data Fusion, TEWA, HCM

13. ABSTRACT/RÉSUMÉ (When available in the document, the French version of the abstract must be included here.)

The report describes the initial plan for an upcoming validation trial using the HALIFAX Class Frigate Her Majesty's Canadian Ship (HMCS) MONTREAL (MON) to investigate the possibility of augmenting the Operations Room (Ops room) crew size.

....

Le rapport décrit le plan initial d'un prochain essai de validation utilisant la frégate de classe HALIFAX du Navire canadien de Sa Majesté (NCSM) MONTREAL (MON) pour étudier la possibilité d'augmenter la taille de l'équipe de la salle des opérations.