2013-09

Performance assessment of military teams in simulator and live exercises

Mjelde, Frode V.

Monterey, California: Naval Postgraduate School

http://hdl.handle.net/10945/37677

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

PERFORMANCE ASSESSMENT OF MILITARY TEAMS IN SIMULATOR AND LIVE EXERCISES

by

Frode V. Mjelde

September 2013

Thesis Advisor: Christian (Kip) Smith
Second Reader: Michael McCauley

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704–0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2013	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE PERFORMANCE ASSESSMENT OF MILITARY TEAMS IN SIMULATOR AND LIVE EXERCISES			**5. FUNDING NUMBERS**
6. AUTHOR(S) Frode V. Mjelde			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943–5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) The Royal Norwegian Naval Academy PO Box 1, Haakonsvern 5886 BERGEN, NORWAY			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT

The purpose of this paper is to present and evaluate a tool designed to assess the performance of military teams participating in complex military training exercises and to investigate the effectiveness of simulator training and live training from the matching of inherent stressors. Specifically, this study evaluates a tool that has been used by Norwegian military subject matter experts (SMEs) to assess the performance of eight cadet teams at the Royal Norwegian Naval Academy (RNoNA) during two separate 4-hour simulator exercises and a 48-hour live exercise. The resulting positive Spearman rank correlation coefficients between team performance assessments in the simulator exercises and the live exercise were strongest when the simulator scenario emphasized the stressors inherent in the live exercise and weakest when the simulator scenario did not facilitate the task demands in the live exercise. The study showed that (1) team performance measured in simulator training exercises can predict performance in a subsequent live training exercise, and (2) that scenario-based simulator training can realistically and effectively represent training demands for live operations. Our findings show the RNoNA tool can be easily applied to team training exercises and provide a meaningful evaluation of a team's future performance.

14. SUBJECT TERMS Human Systems Integration, Manpower, Personnel, Training, Human Factors Engineering, Military teams, Team training, Team performance, Team performance assessment, Teamwork, Taskwork, Norwegian Naval Academy, Simulator systems, Virtual environment, Live environment, Reduced cost, Improved schedule, Improved performance			**15. NUMBER OF PAGES** 129
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01–280–5500 Standard Form 298 (Rev. 2–89) Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

PERFORMANCE ASSESSMENT OF MILITARY TEAMS IN SIMULATOR AND LIVE EXERCISES

Frode V. Mjelde
Lieutenant Commander, The Royal Norwegian Navy
B.S., The Royal Norwegian Naval Academy, 1995

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN HUMAN SYSTEMS INTEGRATION

from the

NAVAL POSTGRADUATE SCHOOL
September 2013

Author: Frode V. Mjelde

Approved by: Christian (Kip) Smith
Thesis Advisor

Michael McCauley
Second Reader

Robert F. Dell
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The purpose of this paper is to present and evaluate a tool designed to assess the performance of military teams participating in complex military training exercises and to investigate the effectiveness of simulator training and live training from the matching of inherent stressors. Specifically, this study evaluates a tool that has been used by Norwegian military subject matter experts (SMEs) to assess the performance of eight cadet teams at the Royal Norwegian Naval Academy (RNoNA) during two separate 4-hour simulator exercises and a 48-hour live exercise. The resulting positive Spearman rank correlation coefficients between team performance assessments in the simulator exercises and the live exercise were strongest when the simulator scenario emphasized the stressors inherent in the live exercise and weakest when the simulator scenario did not facilitate the task demands in the live exercise. The study showed that (1) team performance measured in simulator training exercises can predict performance in a subsequent live training exercise, and (2) that scenario-based simulator training can realistically and effectively represent training demands for live operations. Our findings show the RNoNA tool can be easily applied to team training exercises and provide a meaningful evaluation of a team's future performance.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ANOVA	Analysis Of Variance
AOO	Area Of Operation
AOR	Area Of Responsibility
DoD	Department of Defense
EUNAVFOR	European Union Naval Forces
GUI	Graphical User Interface
HFE	Human Factors Engineering
HFES	Human Factors Ergonomics Society
HNoMS	His Norwegian Majesty Ships
HSI	Human Systems Integration
KSAs	Knowledge, Skill, and Abilities
M&SCO	Modeling & Simulation Coordination Office (U.S.)
MATT	Methods, Approaches, Tools and Techniques
MEC	Mission Essential Competencies
MTB	Motor Torpedo Boat
NAVSIM	RNoNA Navigation Simulator
RNoN	Royal Norwegian Navy
RNoNA	Royal Norwegian Naval Academy
ROE	Rules Of Engagement
RPD	Recognition Primed Decision-making
SME	Subject Matter Expert
SMM	Shared mental models
SOP	Standard Operating Procedure
WFP	World Food Programme

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The objective of the thesis is to expand the understanding of how simulators and live training can be used to assess performance and ultimately improve military team effectiveness. Military team training is normally done in a field setting, involving several resources. Such exercises are time consuming to plan and execute. They can be expensive, and the training outcome is often difficult to predict and assess. Properly constructed scenario-based simulator exercises together with an effective performance assessment tool can present a cost-effective solution for military team performance assessment.

The thesis is supported by the Royal Norwegian Naval Academy (RNoNA). The RNoNA has an interest in enhancing the ability of cadet teams to efficiently and effectively achieve mission objectives in complex military environments. This interest has led the Academy to support the effort to develop a tool designed to assess the performance of military teams participating in complex military training exercises.

The RNoNA tool includes twelve metrics to assess teamwork (eight) and taskwork (four) characteristics. The teamwork processes include interactions team members must develop and perform to function effectively as a team: team orientation, backup behavior, mutual performance monitoring (includes mutual trust), closed-loop communication, team leadership, shared mental models, adaptability and agility. The taskwork processes, creative action, speed, thoroughness and success are evaluated from the outcome of individual and team tasks and actions. The taskwork characteristics refer to resilient behaviors related to the operational activities the cadet teams must perform in a complex and stressful environment.

Norwegian military subject matter experts (SMEs) used the RNoNA tool to rate the performance of eight cadet teams at the RNoNA during two separate 4-hour simulator exercises and a 48-hour live exercise. Each cadet team had eight to nine team members, both male and female, with one to four years of prior service in the Norwegian military. The simulator training consisted of realistic and demanding scenarios, representing military operations that required demonstration of teamwork and taskwork competencies according to RNoNA training objectives. The live exercise was representative of actual military operations, while being performed in a controlled training environment. The ratings were analyzed to determine (1) the extent to which team performance assessment in a series of simulator exercises can predict performance in a live exercise, and (2) whether training objectives for a live training exercise can be realistically and effectively achieved through scenario-based simulator training exercises.

All eight teams performed the same exercises in a repeated measures design. The ratings data are ordinal, not ratio. Accordingly, the appropriate statistical analyses are non-parametric. The ordinal ratings data were analyzed using the non-parametric Spearman rank-order correlation coefficient ρ. A Kruskal-Wallis analysis of variance (ANOVA) accounted for tied ranks to compare ranks across teams, and post-hoc Steel-Dwass analysis and pairwise Wilcoxon analysis were used to compare differences in ranks across teams. Positive Spearman rank correlation coefficients between team performance assessments in the simulator exercises and the live exercise were strongest ($\rho = .73$) when the simulator exercise emphasized the stressors inherent in the live exercise, and weakest ($\rho = .05$) when the simulator exercise did not contain the task demands in the live exercise.

The study found support for all hypotheses and showed that the RNoNA tool can (1) measure team performance in simulator training exercises, and predict which team will perform better (or worse) in a subsequent live training exercise, and (2) that scenario-based simulator training can realistically represent demands for live training exercises when there is a match between stressors and resilient behavior in both training domains.

The assessment tool was shown to be effective in predicting team rankings assessed through teamwork and taskwork behavior in both training and live environments. However, inspection of the observed ratings revealed range restriction among SMEs; the majority of raters did not use the available 7-point scale to its full extent. Range restriction and SME feedback have led to improvements in the RNoNA tool and procedures. Some metrics have been changed, and both the tool and the procedures to implement it have been improved. The findings also prompted plans for future research programs in which the RNoNA will use the revised assessment tool in a longitudinal study to assess RNoNA cadet team performance in simulator and live training exercises in 2014 and 2017.

The demands for operational effectiveness and competitive advantage on the battlefield create a need for effective team training exercises and team assessment tools. The RNoNA tool enabled the SMEs to assess RNoNA cadet teams in simulator training exercises and in a live training exercise. The RNoNA tool was shown to be easily applied, within a short timeframe, and to provide a meaningful assessment of a team's future performance.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First and foremost I wish to thank my family for their continuous love and support. Special warmth, gratitude and love goes to my wonderful wife, Eva. You are the anchor that keeps me grounded, and at the same time the source that lifts me up and brings me inspiration. My children, Lene, Dina and Mali; you keep everything interesting and bring me joy. I wish to thank my parents and family for their encouragement always.

I would like to express my deep gratitude to Professor Christian (Kip) Smith, my advisor, for his patient guidance, enthusiastic encouragement and useful critiques of this research work. I would also like to express my great appreciation for my second reader, Professor Michael McCauley, for his constructive suggestions during the development of this research, and to Professor Susan Hutchins for her advice and assistance in the early stages of my work.

I would like to express my gratitude to Col. (retired) Lawrence Shattuck for his valuable and constructive guidance during the graduate studies at the Naval Postgraduate School. His willingness to give his time so generously is remarkable and has been very much appreciated. My grateful thanks are also extended to Dr. Nita Shattuck for her genuine interest and caring. A heartfelt thank-you goes to my fellow HSI cohort members.

This work was sponsored by The Royal Norwegian Naval Academy. Special thanks go to Cdr Sg. Roar Espevik for his contributions to the development of the assessment tool and to Cdr Petter Lunde for coordinating the data collection during the RNoNA exercises. I would also like to extend my thanks to Captain Thomas Wedervang and Cdr Sg. Morten Jansen for their efforts in making the NPS master study a reality.

Finally, I wish to thank the RNoNA staff and cadet teams of 2012. Without you; no data. Without data; no study. Thank you.

Honor, Commitment and Courage.

For King, Country and the Honor of the Flag.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

Teams are a fact of life. From boy scouts to commercial airline pilots to medical teams, from business management to military operations to the 100-meter relay in the 2012 Olympic Games in London, teams are essential for producing results that cannot be achieved by the individual alone. Despite the dependence society has on teams, there is still much to be learned about the processes that occur within a team that yield high levels of performance and successful outcomes (Brannick, Prince, Prince, & Salas, 1995).

The demand for operational effectiveness and competitive advantages on the battlefield requires effective team training interventions and team assessment methods that can be applied easily, within a short timeframe, and that provide a meaningful predictive (valid) evaluation of a team's future performance. The focus of this thesis is the initial evaluation of a tool designed to assess the performance of military teams participating in complex military training exercises.

B. BACKGROUND

The thesis is supported by the Royal Norwegian Naval Academy (RNoNA). The RNoNA has an interest in enhancing the ability of cadet teams to efficiently and effectively achieve mission objectives in complex military environments. This interest has led the Academy to support the effort to develop a team performance assessment tool that can be used to determine (1) the extent to which team performance assessment in simulator exercises can predict performance in a live exercise, and (2) whether training objectives for a live training exercise can be realistically and effectively achieved through scenario-based simulator training exercises.

The data analyzed in this thesis are quasi-experimental measures of teamwork and taskwork collected as part of RNoNA cadet team training. This training takes place in a series of realistic and dynamic exercises throughout their 3-year education. Key objectives in their education are elements of teamwork and taskwork: team cognition, team decision making, adaptability, agility, action and continuous feedback are critical training objectives (Royal Norwegian Naval Academy, 2009). Each exercise is specifically created for the purpose of training cadet teams to perform complex and stressful tasks that may be cognitive, behavioral or attitudinal in nature. The simulator training consists of realistic and demanding exercises, representing military operations that require demonstration of teamwork and taskwork competencies according to RNoNA training objectives (Mjelde, 2013). The live exercises are representative of actual military operations, while being performed in a controlled training environment (Royal Norwegian Naval Academy, 2010). The exercises take place in observable environments, which is necessary to obtain measurements of performance (Salas, Rosen, Burke, Nicholson, & Howse, 2007). The simulator and live exercises are well suited to the collection of quasi-experimental data on the teamwork and taskwork performance of RNoNA cadet teams (Royal Norwegian Naval Academy, 2009).

The performance assessment is made through observation of task execution and scored using the RNoNA Team performance assessment tool. The tool includes twelve markers to assess teamwork (eight) and taskwork (four) characteristics. The teamwork processes include interactions team members must develop and perform to function effectively as a team: team orientation, backup behavior, mutual performance monitoring (includes mutual trust), closed-loop communication, team leadership, shared mental models, adaptability and agility (Alberts, 2007; Brannick, Salas, & Prince, 1997; Cannon-Bowers & Salas, 1998; Entin & Serfaty, 1999; Espevik, Johnsen, & Eid, 2011; Salas, Sims, & Burke, 2005; Zaccaro, Rittman, & Marks, 2001). The four taskwork processes, creative action, speed, thoroughness and success are evaluated from the outcome of individual and team tasks and actions. The taskwork characteristics refer to resilient behaviors related to the operational activities the cadet teams must perform in a complex and stressful environment (Bandura, 1977; Boyd, 2005; Brehmer, 2005; Dalton, 2004; Hollnagel, Woods, & Leveson, 2006). The twelve characteristics are further explained in Chapter II, and the RNoNA Team performance assessment tool is described in Chapter III.

C. OBJECTIVE

The objective of the study is to expand the understanding of how simulator and live training methods can be used to teach skills that will be used for assessment and improvement of military team effectiveness.

Early assessment of team performance can lead to early corrections of key parameters to improve performance. An effective assessment tool can provide not only early adjustment, but also the correct adjustment. Reliable and valid measures of team performance could also prove useful for selection of team members (Brannick, Salas, & Prince, 1997).

D. PROBLEM STATEMENT

When military teams are assigned to tasks and missions in modern warfare, they must employ effective communication, coordination and cooperation strategies to be successful (Cannon-Bowers & Salas, 1998; Entin & Serfaty, 1999). The process of selecting team members, training and evaluating the team is typically time consuming and expensive. Ideally, resources would be unconstrained with ample opportunities to put together teams, to test their performance and then evaluate them against the demands imposed by the task and mission. In practice, the process is constrained by three distinct factors: cost, schedule and performance. The request from the stakeholders is to deliver high performance, at the lowest achievable cost, in the shortest amount of time.

Military team training is normally done in a field setting, involving a lot of resources. Such exercises are time consuming to plan and execute; they can be expensive and the training outcome is often difficult to predict and assess. Properly constructed scenario-based simulator exercises together with an effective performance assessment tool can present a cost-effective solution for military team performance assessment.

E. RESEARCH QUESTIONS & HYPOTHESES

This thesis addresses two broad research topics the Assessment Tool and Stressors in the context of identifying, measuring, supporting and enhancing team performance levels in RNoNA cadet teams.

1. Assessment Tool

The first three research questions address the RNoNA team performance assessment tool. They ask whether the tool enables the SMEs to make assessments of RNoNA cadet teams in simulator training exercises that can predict cadet team performance in a live training exercise.

- H1: When used in a training simulator, the average score of the twelve selected measures in the RNoNA team performance assessment tool can be used to predict performance in a live training exercise.

The tool includes twelve measures to assess team performance: eight teamwork and four taskwork measures. H1 is about an average score of all twelve measures, while H2 isolates the eight teamwork scores, and H3 isolates the four taskwork scores to investigate predictability.

- H2: When used in a training simulator, the average score of the eight selected teamwork measures in the RNoNA team performance assessment tool can be used to predict the average of the eight measures in a live training exercise.
- H3: When used in a training simulator, the average score of the four selected taskwork measures in the RNoNA team performance assessment tool can be used to predict the average of the four measures in a live training exercise.

The associated null hypotheses state that the tool does not predict team performance in a live exercise based on assessments made in simulator exercises.

2. Stressors

The second group of research questions addresses the impact of stressors, e.g., uncertainty, fatigue, time pressure, etc. These and other stressors are not equally represented in all three exercises, which may affect the cadet teams' behavior and performance. Accordingly, these questions ask whether the match (or difference) between the stressors built into a simulator exercise and a live exercise has an impact on the tool's prediction of team performance, and whether there is a differential impact on measures of teamwork and taskwork.

- H4: There is a difference in assessment predictability depending on the match of stressors built into the training exercise and the stressors in the live exercise.
- H5: There is a difference in teamwork assessment predictability depending on the match of stressors built into the training exercise and the stressors in the live exercise.
- H6: There is a difference in taskwork assessment predictability depending on the match of stressors built into the training exercise and the stressors in the live exercise.

The associated null hypotheses state that a difference in stressors does not influence the tool's prediction of team performance.

F. SCOPE AND LIMITATIONS

The thesis examined the concept of predicting team performance in a live military exercise using performance assessments from two previous simulator exercises. The RNoNA assessment procedure is based on the work by other researchers who concluded that important team processes are identifiable (Boyd, 2005; Brehmer, 2005; Cannon-Bowers & Salas, 1998; Salas, Sims, & Burke, 2005) and can be rated validly (Brannick, Salas, & Prince, 1997; Dickinson & McIntyre, 1997; Entin & Serfaty, 1999; McIntyre & Salas, 1995; Salas, Cooke, & Rosen, 2008). That work found that realistic, ratable, unobtrusive and real-time multiple observations are necessary to assess characteristics of individual teams with any accuracy. The observational study in this thesis includes two exercises in the RNoNA simulator, and a live exercise with several sub-scenarios allowing for multiple performance assessments.

The military teams in the research are Norwegian Navy cadets. The results from the study may increase the understanding of military team performance in general, but may or may not generalize to similar teams from other countries and cultures.

G. HUMAN SYSTEMS INTEGRATION (HSI)

Human Systems Integration (HSI) is a multidisciplinary field of study composed of the integration of the domains of manpower, personnel, training, human factors engineering, system safety, health hazards, habitability and survivability. HSI emphasizes human considerations as the top priority in military systems design to reduce life cycle costs and optimize human and system performance (Naval Postgraduate School, 2013). The following HSI domains apply to this research.

1. Human Factors Engineering

The goal of Human Factors Engineering (HFE) is to maximize the users' ability to perform at required levels for operation, maintenance and support by considering human capabilities and limitations and eliminating design-induced errors (U.S. Army, 2005).

One of the tools used in this thesis is the ship-handling simulator (NAVSIM) at the Royal Norwegian Naval Academy, a simulator system originally procured to facilitate training of navigational skills to cadets at the Academy and other Navy crews. Traditionally, simulators and trainers for military training are used to train a single skill or to provide training within a specific domain. In this study, the NAVSIM is used non-traditionally. Instead of running a typical navigation-training exercise, the simulator system facilitates a scenario-based training event to train team processes on a tactical level in a complex military setting.

Simulator training is heavily used in the Department of Defense (DoD). The mission statement of the U.S. Modeling & Simulation Coordination Office (M&SCO) is to perform key corporate-level coordination functions necessary to encourage cooperation, synergism and cost-effectiveness among the M&S activities of the DoD Components. Among these are interoperability, reuse and affordability to provide improved capabilities for DoD operations.

The non-traditional training intervention discussed here suggests how an existing simulator system with specific functions and task environment can be effectively reused in a cost-effective approach to avoid unnecessarily duplicating tools and as a means to improve capabilities for military operations.

2. Manpower

The Manpower domain addresses the total number of people needed for operation, maintenance and support in order to achieve required operational capabilities. It considers the requirements for peacetime, low intensity operations and conflicts, with current and projected operating strength (available people) for the organization (U.S. Army, 2005).

Assessing processes and outcomes of teamwork and taskwork in a live exercise require extensive resources to ensure reliability and validity of the metrics (Dickinson & McIntyre, 1997). With an effective and reliable team performance assessment tool, measures can be collected in a simulator environment that is less demanding of coordination activities and manpower resources than a live environment is.

3. Personnel

The Personnel domain is closely related to the Manpower domain. While the Manpower domain looks at how many people are needed, the Personnel domain addresses which capabilities they need to perform at required levels for operation, maintenance and support. Personnel capabilities are normally described as a combination of Knowledge, Skills and Abilities (KSAs) and other characteristics (e.g., cognitive, physical, hardiness and sensory) (U.S. Army, 2005). The availability of personnel and their KSAs should be identified to establish performance thresholds and personnel selection criteria.

A successful team is more than the sum of its individual skills, abilities and knowledge; it requires an active team process to make the team greater than its parts. The use of a realistic scenario-based simulator exercise combined with the team performance assessment tool provides a decision maker with performance thresholds to identify, evaluate and choose the best mix of team members prior to mission execution. Documentation of performance thresholds and standardization of scenario designs will benefit operators when they are matched to jobs for which they are well suited. In addition, the use of systematic assessment tools will reduce the degree of error in making selection decisions and often results in more favorable operator reactions to the selection process (U. S. Office of Personnel Management, 2007).

4. Training

Training is defined as the instruction, education and learning process by which personnel individually or collectively acquire or enhance essential job knowledge, skills, abilities and attitudes necessary to effectively operate, deploy/employ, maintain and support the system (U.S. Army, 2005).

> Key considerations include developing an affordable, effective and efficient training strategy; determining the resources required to implement it in support of fielding and the most efficient method for dissemination; and evaluating the effectiveness of the training. (U.S. Army, 2005, chapter 1.1.4)

The process of establishing military teams and their specialized team training is time consuming and expensive. With constrained resources, there is a need to establish effective team training methods that can be applied easily, and within a short timeframe. By comparing team assessments in simulator exercises and a field exercise, this thesis will address factors influencing team effectiveness and provide insight to team training.

An adaptive simulator environment can allow teams to explore consequences of different options to test intuitive predictions against doctrine to establish best-practice models. The scenario-based approach to training can offer specialized knowledge of resilience training by visualizing critical change factors, facilitating solution alternatives and developing methods, approaches, tools and techniques (MATTs) to address them.

H. SUMMARY

One of the training principles used to meet team performance objectives is to balance the processes of teamwork and taskwork (McIntyre & Salas, 1995). To assess this balance, it is important to measure both teamwork and taskwork. Assessing team performance in a dynamic military environment requires the evaluator, or Subject Matter

Expert (SME), to be well equipped for the task. Matthews et al. (2011) expressed concern that performance evaluations are challenging in a laboratory and even more daunting in the field. To this end, this research developed a tool designed to address a pair of issues that appear to have received relatively little attention in the teamwork literature: reliable measures of (1) team performance in both virtual and live military exercises and (2) the match between simulator and live exercises for military training (Ross, Phillips, Klein, & Cohn, 2005; Salas, Cooke, & Rosen, 2008).

I. THESIS ORGANIZATION

This first chapter has presented the background, the research questions and different aspects of framing the subject from an HSI point of view. Chapter II is the literature review on teamwork and taskwork. It discusses essential terms related to scenario-based team training. Chapter III is part 1 of the methodology section. The chapter starts with an explanation of the design and construction of the RNoNA team performance assessment tool and ends with a discussion of resilience and validity. Chapter IV is part 2 of the methodology section. It covers the participants, design, apparatus, and procedures of the observational research. Chapter V presents the collected data and the analysis. The findings are then summarized and discussed in Chapter VI, which resulted in changes in the design of the RNoNA team assessment tool and its delivery process. Suggestions for future research are also included in Chapter VI.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. OVERVIEW

This study necessarily incorporated research on training principles used to meet team performance objectives and on team performance assessment. Most of the existing literature that was used for this study focused on the challenges of assessing the balance of teamwork and taskwork (McIntyre & Salas, 1995), particularly in a dynamic military environment. Matthews et al. (2011) expressed concern that performance evaluations are challenging in a laboratory and even more daunting in the field. To this end, this research developed a tool designed to address a pair of issues that appear to have received relatively little attention in the teamwork literature: reliable measures of (1) team performance in both virtual and live military exercises and (2) the match between simulator and live exercises for military training (Ross, Phillips, Klein, & Cohn, 2005; Salas, Cooke, & Rosen, 2008).

The literature review explains the stressors associated with team training at the RNoNA, and how overall team performance as the result of teamwork and taskwork performance. Scenario-based team training for military teams is described as an effective, adjustable and controllable method that provides useful opportunities in simulator and live training for team performance training and assessment.

B. STRESSORS

While complex simulations and field exercises cannot fully replicate the actual combat environment, the RNoNA team training exercises expose cadet teams to a wide range of psychological and physical stressors representative of those found in military operations. These stressors include sleep deprivation, food deprivation, fatigue, time pressure, unambiguous information, uncertainty, and mismatches between expectations, perception and the unfolding of actual events. Such physical and psychological stressors are included in the RNoNA team training (Royal Norwegian Naval Academy, 2009). In the live exercise, the weather conditions introduce additional environmental stressors like cold, heat, noise, etc.

The primary stressors found in modern military operations are isolation, powerlessness, ambiguity, boredom, danger and workload (Bartone, 2006). Strategies and coping mechanisms found to increase resiliency or resistance to such stressors at the team levels include backup behavior, trust, team leadership, adaptability, agility and thoroughness (Adams & Webb, 2002; Bandura, 1977; Entin & Serfaty, 1999; Jarvenpaa & Leidner, 1999; Wilson, Salas, Priest, & Andrews, 2007). Different stressors will affect different people to varying extents (Civil Aviation Authority, 2006). Stressors affect cognitive processing and teamwork behavior and can limit the scope of team cognition and team decision making (Salas, Rosen, Burke, Nicholson, & Howse, 2007).

One of the intents of the RNoNA team training exercises is to enable cadet teams to automate behavioral processes, which will make the behavior more resistant to the effects of stressors (Paris, Salas, & Cannon-Bowers, 2000). It is possible to build training exercises to emphasize different combinations of stressors. If the stressors differ in exercises X and Y, that difference may constrain how well measures of performance in X can predict performance in Y. Hence hypotheses H4, H5 and H6 investigate whether the match (or difference) between the stressors built into a simulator exercise and a live exercise has an impact on the tool's prediction of team performance.

C. TEAM COGNITION AND DECISION MAKING

Team cognition and team decision making are aspects and processes that impacts team performance.

1. Team cognition

Cognition literally means "*to know.*" Knowledge can be thought of as memories formed from the manipulation and assimilation of sensory input, perceived via our senses. Using knowledge to direct and adapt action towards goals is the foundation of the cognitive process. Past experiences and trends inform our sense of what the future might hold and help us to act accordingly. Figure 1 illustrates a three-stage model of information processing psychology (Wickens & Hollands, 2000). The model depicts the hierarchy of cognitive functions involved to explain how we process available information, and how attention and memory determine the outcome.

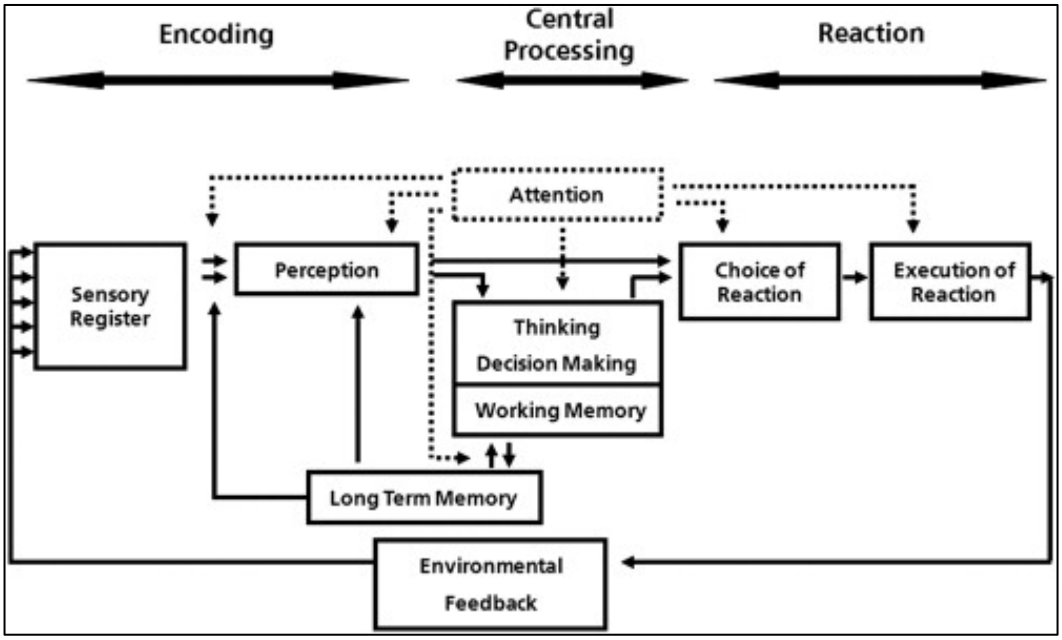

Figure 1. Information processing model (From Wickens & Hollands, 2000).

The three stages are: the encoding stage, the central processing stage and the reaction stage. The central processing stage represents the cognitive process, where observed information is evaluated using structures in both working and long-term memory (Wickens & Hollands, 2000). The process takes place internally for each team member.

Team cognition can be viewed as a process, or a product of group interaction (Salas & Fiore, 2004). Team cognition relies on the individual team member's cognitive abilities, where each member processes the information available and communicates his results back to the team as inputs to the process of establishing a collective product. The coordinated team cognition process will support team performance, task performance and team decision making to achieve mission objectives (Salas, Cooke, & Rosen, 2008).

Team cognition also involves the understanding of how knowledge important to team performance is mentally held and distributed within the team (DeChurch & Mesmer-Magnus, 2010). Team cognition provides a foundation for team members to coordinate their actions jointly. The individual's *knowledge* about teamwork and taskwork processes, their *skills* to perform those processes, their *abilities* to aid the team in them, and their *attitudes* towards learning and understanding the benefits of them are important for establishing a truly collective team cognition.

Assessment of team cognition is challenging because not every piece of data necessary for a full understanding can be observed or elicited (Cooke, Salas, Kiekel, & Bell, 2004). A team's total knowledge can be aggregated from individual cognition elicited through conceptual methods, observations, interviews, surveys, process tracing, etc. Such information is usually elicited from long-term knowledge or episodic memories (events), since the assessment is often performed after the experience. The assessment activities based on the collective knowledge elicited can measure how well teams meet a criterion, whether it is process or outcome, and provide explanations as to why specific actions fall short of their respective criteria or why they exceed them (Kiekel & Cooke, 2011).

Team cognition assessment will not necessarily capture the dynamic fleeting knowledge that occurs during an event, and it is difficult to know if the behaviors observed are representative of what individual team members experience (Cooke, Salas, Kiekel, & Bell, 2004). This challenge will require measures of team cognition that can be administered and scored in real time as the task is performed and events unfold (Kiekel & Cooke, 2011).

There is also the question of whether the measures are to be quantitative, qualitative or a combination of both. Quantitative measures can be performed automatically (e.g., counts), but qualitative measures are better performed through observations or after-action reviews (Kiekel & Cooke, 2011). To understand team cognition in military teams, the assessment tool should include both quantitative and qualitative approaches.

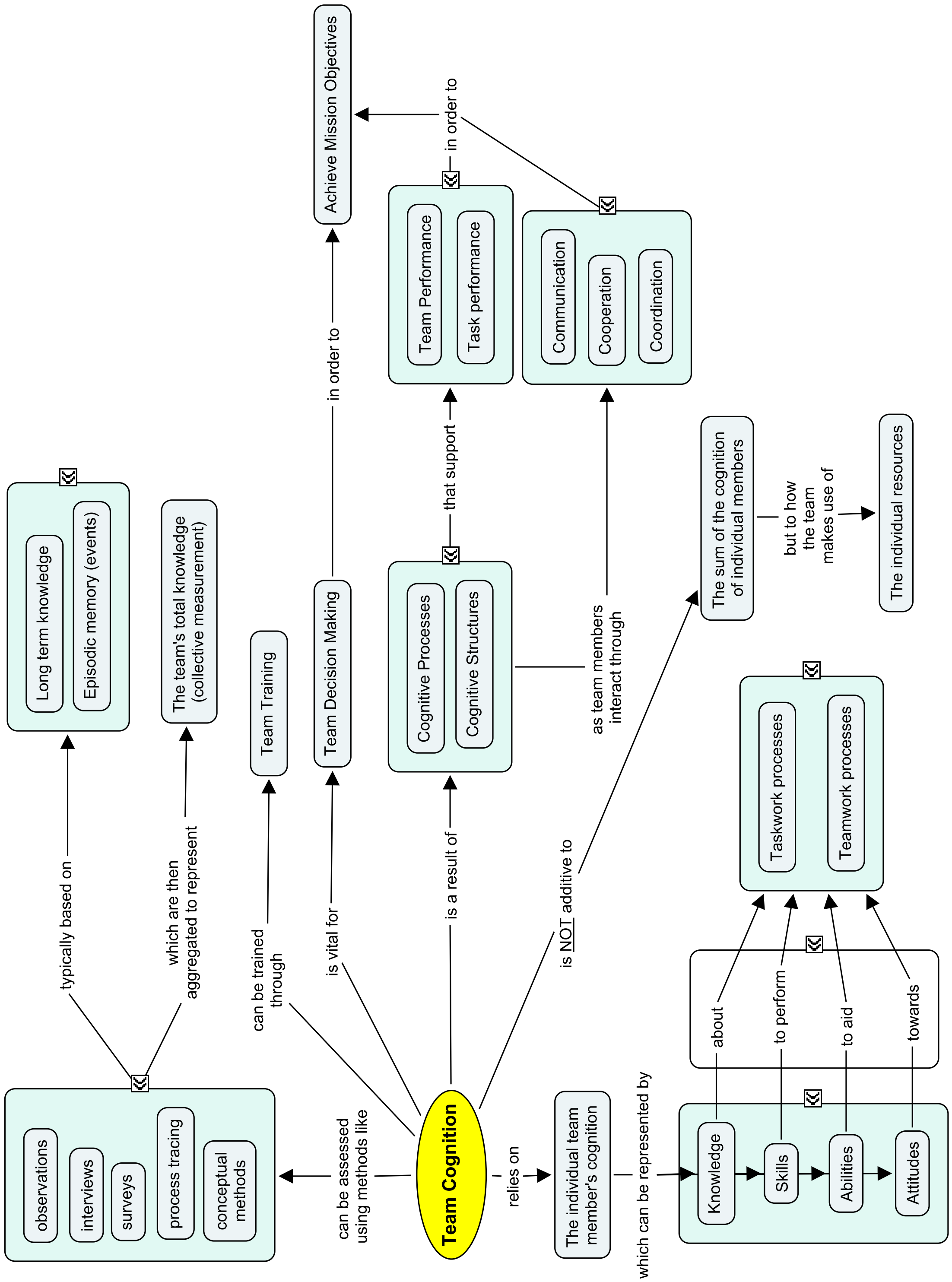

Figure 2. Team cognition concept map.

2. Team decision making

One of the key aspects of teamwork is team decision making (Figure 3), which refers to the process of reaching a decision undertaken by interdependent individuals to achieve a common goal [Orasanu and Salas, 1993 in (Flin, O'Connor, & Crichton, 2008)]. Team decision making relies on individual team members communicating critical information to the rest of the team. The outcome of an individual team member’s cognitive processes is communicated back to the team as inputs to establishing team cognition that supports effective team processes.

Organizations rely on teams to perform tasks that are complex, demanding or require a coordinated effort. Some advantages that teams offer include the capacity to pool resources, exchange information, coordinate actions and even share the responsibility for team decisions. However, team members do not always operate efficiently as a team, and the lack of teamwork or failure to function collectively has had serious consequences (Driskell, Salas, & Hughes, 2010). Examples of high-profile accidents include the Pan Am and KLM collision at Tenerife (1977), the Three Mile Island accident in the U.S. (1979), and the USS Vincennes incident (1988). Reviews of these tragedies identified three main teamwork problems: roles not clearly defined, lack of explicit coordination and communication problems (Flin, O'Connor, & Crichton, 2008).

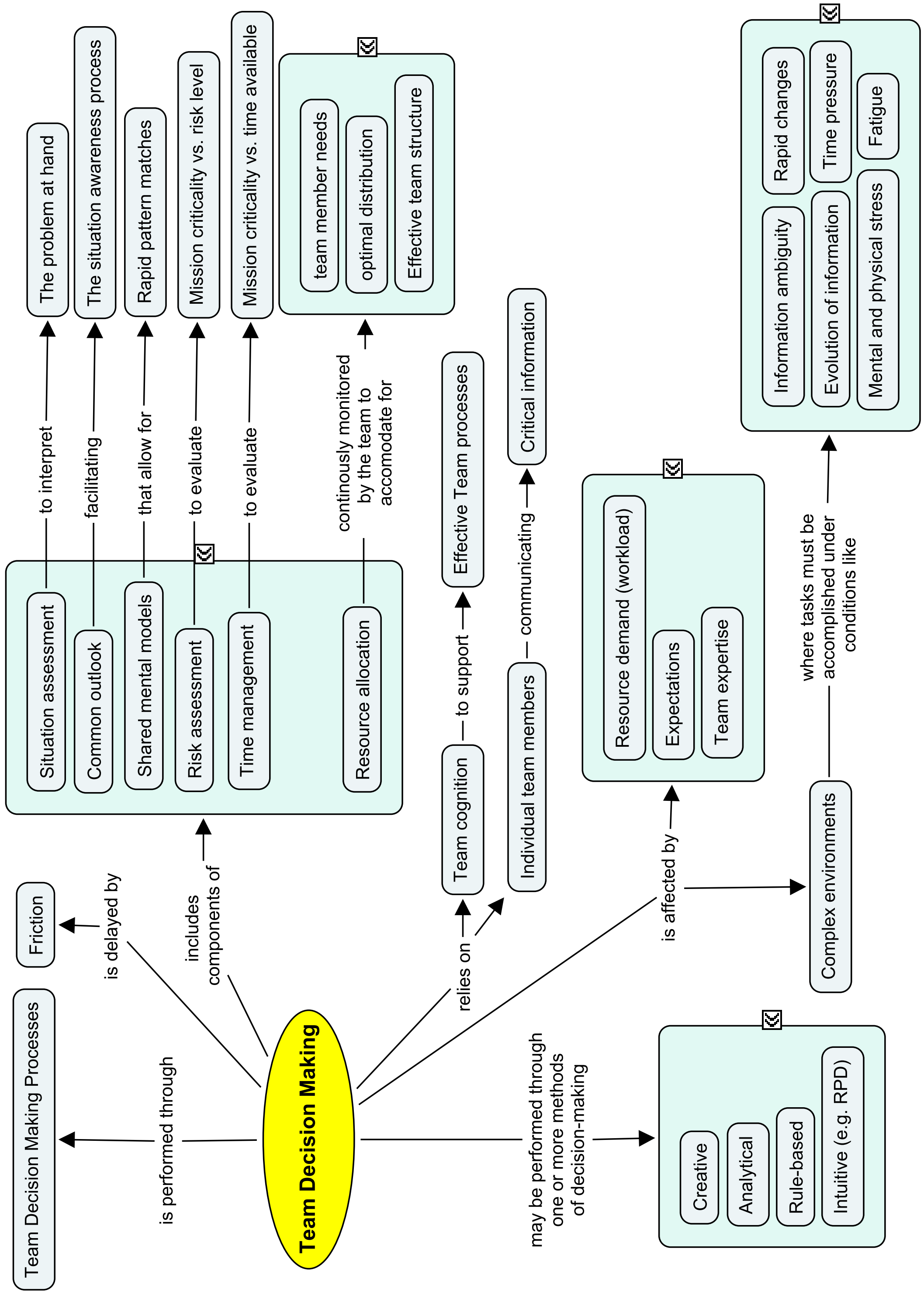

Figure 3. Team decision making concept map.

Team decision making is affected by the complexity of the operational environment (Cannon-Bowers & Salas, 1998), where tasks must be accomplished under stressors like time pressure, performance pressure, information ambiguity, rapidly evolving situations, resource demand, fatigue, and excessive mental or physical task demand. Team decision making in a dynamic environment is highly complex and requires an investigation of process variables to understand and manage components affecting team performance (Urban, Weaver, Bowers, & Rhodenizer, 1996). Components of team decision making include: situation assessment to interpret the problem at hand, the establishment of a common outlook to facilitate the situation awareness process, shared mental models that allows for rapid pattern matches, risk assessment to evaluate mission criticality vs. risk level and time management to evaluate mission criticality vs. time available. Resource allocation is a part of the cognitive and practical structure employed by the team to monitor and accommodate individual team members' needs and optimize distribution of resources effective for the task/mission and environment.

Team decision making is a continuous process where decisions and actions have both reasons and consequences. Situation assessment defines the objective, and that knowledge decides team cognition processes and defines the nature and importance of the means to be employed to achieve the mission objective. The *objective* was the answer to the question that faced Verdy du Vernois when he reached the field of battle at Nachod, Czech Republic (Foch, 1918). When he realized the difficulties to be overcome he searched in vain through his memory for a method, tool, technique or approach for what to do. After not finding an answer, he said "*To the devil with history and principles! After all,* ***what is my objective***?" The narrative should indicate that du Vernois realized he needed to know *what to do* before he could plan on *how to do it*. He then assessed the situation from the perspective of the mission objective to face what was before him. Similarly in modern warfare, a tactical move or action is achieved through situation assessment and decision making based on the perspective of mission objective.

Time management is critical for situation assessment, and how much time that is estimated to be available for the decision-making process will affect risk assessment as well as choosing a decision-making method. Judith Orasanu's Decision Process Model in Figure 4 illustrates the influence of time pressure and risk on the choice of decision strategies (Orasanu, 1995).

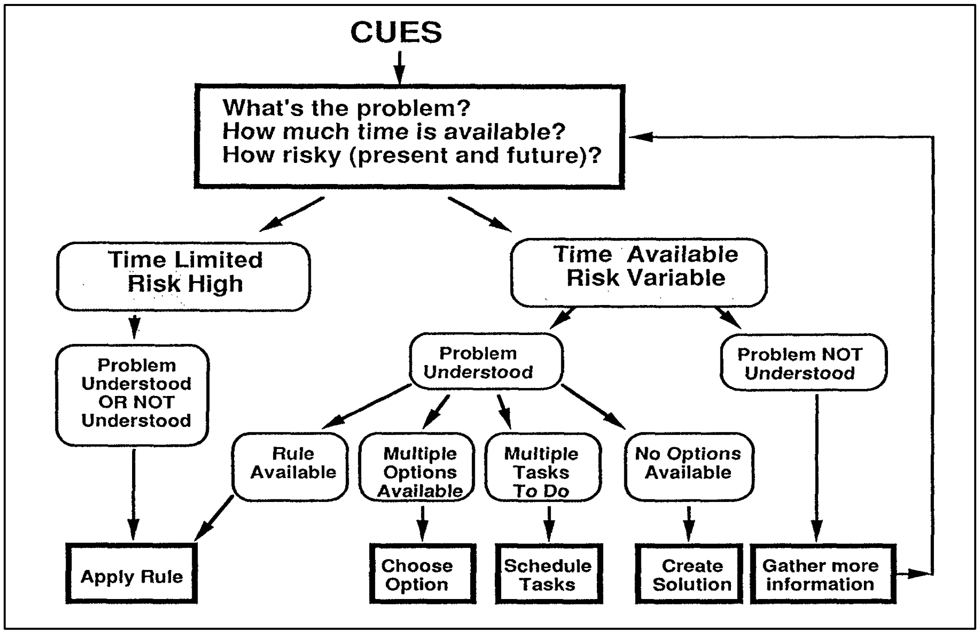

Figure 4. Decision Process Model (From Orasanu, 1995).

Orasanu showed that the estimation of available time and level of risk during the situation assessment phase, determined the type of decision method adopted. Team decision making may then be performed through one or more methods, such as: (a) creative, (b) analytical (choice), (c) rule-based and (d) intuitive (e.g., recognition primed).

a. Creative

The creative method is rarely used when time is limited, as it requires considerable cognitive effort to devise a novel course of action for an unfamiliar situation (Flin, O'Connor, & Crichton, 2008). Creativity is important, however, for innovation when other interventions fail to be effective. Creative decision making seems to be best used when the situation is unfamiliar and/or there is ample time available to facilitate a process for sharing diverse perspectives and points of view through creative thinking (Osinga, 2005).

b. Analytical

Analytical, or choice decisions, is the method of comparing options. The team generates a number of possible courses of action that are compared to determine which one is best suited for the needs of the situation. Ideally, for this method, all the relevant information and features of the options should be identified and then weighed to determine their match to the requirements (Flin, O'Connor, & Crichton, 2008). A disadvantage of the analytical method is that it requires time and cognitive effort, and it can be affected by stress.

c. Rule-based

The rule-based, or procedure-based, decision-making method involves identifying the situation encountered and remembering or looking up in a manual the rule or procedure that applies (Flin, O'Connor, & Crichton, 2008). Rule-based decision making is extensively used for novice teams to learn standard operating procedures (SOPs) and provides a course of action already determined by domain experts. Procedures also are useful for expert teams, especially to avoid increased cognitive strain during stressful events.

d. Recognition-primed

Recognition-primed decision making (RPD) relies on remembering the responses to previous situations of the same type. Situational cues can be matched with past experience, resulting in a satisfactory and workable course of action. The team can simulate implementing the recalled solution in the current situation, and if no problems are visualized, the solution can be implemented. Alternatively, if the visualization indicates a problem, the solution can be modified. The feedback the team receives from implementing the plan serves as input to the next decision that must be made (Klein, 1993). Advantages to RPD are that it is very fast, requires little conscious thought, can provide satisfactory solutions and is reasonably resistant to stress, but it also requires that the user be experienced in the domain (Flin, O'Connor, & Crichton, 2008).

The team decision-making process can be delayed by internal and external friction. Internal friction may occur when team members do not have the same goals, motivation or do have poor teamwork KSAs. Team members may also feel pressured to conform and be susceptible to groupthink (Flin, O'Connor, & Crichton, 2008). External friction within the process usually can be found in the environment as delays between the decisions and actions, delays between actions and effects, and between effects, results and consequences of that act (Brehmer, 2005). John Boyd said about friction that

> (1) the atmosphere of war is friction; (2) friction is generated and magnified by menace, ambiguity, deception, rapidity, uncertainty, mistrust, etc.; (3) friction is diminished by implicit understanding, trust, cooperation, simplicity, focus, etc.; and (4) in this sense, variety and rapidity tend to magnify friction, while harmony and initiative tend to diminish friction. (Osinga, 2005, p. 235)

Friction can be diminished through the application of effective methods and tools to the team decision-making processes resulting in enhanced team performance.

D. TEAM PERFORMANCE

Team performance relies on individual team members to schedule and perform individual and team tasks and communicate critical information to maximize the collective performance. As team members interact through communication and coordination, their individual work, results and responses will come together in a multilevel process to support mission objectives (Figure 5).

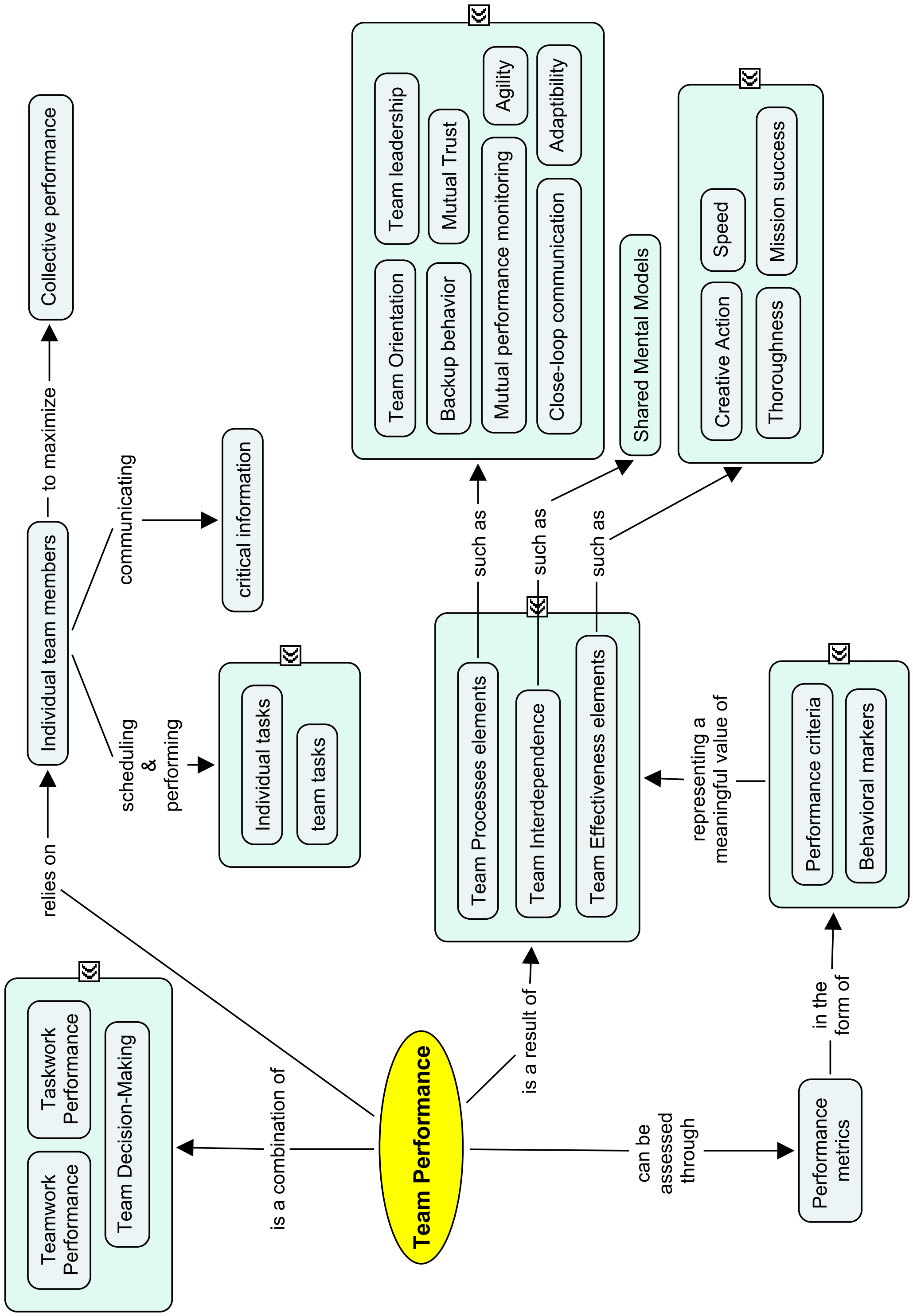

Figure 5. Team performance concept map.

Decisions and actions are achieved through a combination of taskwork, teamwork and team decision making (Flin, O'Connor, & Crichton, 2008). The level of performance depends on team processes, team interdependence and team effectiveness, identified as vital elements for understanding team performance (Salas, Cooke, & Rosen, 2008). Team process elements include the behavioral teamwork interactions that team members must develop and perform to function effectively as a team, such as team orientation, backup behavior, mutual performance monitoring, mutual trust, communication, team leadership, adaptability, agility and orientation (Salas, Sims, & Burke, 2005; Entin & Serfaty, 1999; Alberts, 2007). Team interdependence requires coordination and synchronization among members and integration of their contributions to achieve team goals (Zaccaro, Rittman, & Marks, 2001). Teams who understand their interdependence and the benefits of working together are more likely to establish shared mental models.

A team's effectiveness is often interpreted as the result of the taskwork, and it is gained through the performance of individual and team tasks. Taskwork refers to behaviors related to the operations activities the team must perform (Flin, O'Connor, & Crichton, 2008). Taskwork effectiveness criteria can be based on effects that are quantitative, qualitative or a combination of both (Kiekel & Cooke, 2011). The tasks performed will depend on the team structure (e.g., dispersed, local, established, scheduled) and the nature of the team's task or work and the stressors in the environment in which they operate. Team effectiveness elements can refer to resilient behaviors necessary for success in certain operational activities (Hollnagel, Woods, & Leveson, 2006) and can be measured by taskwork characteristics like creative action, speed to complete assignments, thoroughness and successful accomplishment of mission objectives (Mjelde, 2013).

Understanding and analyzing team performance should begin with an understanding of the tasks to be performed (Cannon-Bowers & Bowers, 2011). The RNoNA team performance assessment tool has been applied to assess performance metrics in the form of eight behavioral markers that represent a meaningful value for teamwork and four performance markers for taskwork effectiveness.

E. TEAMWORK

Teamwork is behavior related to interactivity and attitudes the team must develop before it can function effectively in a stressful environment (Flin, O'Connor, & Crichton, 2008). Elements of teamwork relate to how members of a team, independent of role and task within the team, enhance team effectiveness. The eight teamwork components included in the RNoNA assessment tool are (1) team orientation, (2) backup behavior, (3) mutual performance monitoring and mutual trust, (4) closed-loop communication, (5) team leadership, (6) shared mental models and interdependence, (7) adaptability and (8) agility.

The eight teamwork components are discussed in turn. The definition quoted at the beginning of each discussion is the assessment statement exactly as it is printed the 2012 version of the RNoNA assessment tool. The data collection presented in Chapter V is based on the 2012 version of the tool. In Chapter VI several of these definitions were updated.

1. Team orientation

> The team showed a high degree of involvement (team members monitored and paid attention to other team members, not many "free riders" in the teamwork process). (RNoNA assessment tool, 2012)

Team orientation is more an attitude than a behavior. It addresses how team members orient towards working with others while sharing the desire and effort to enhance the outcome of the team tasks (Salas, Sims, & Burke, 2005). Salas et al. (2005) indicate that team orientation is based on previous experience in teams, on the anticipated ability to complete the task (task efficacy) and on expected positive outcomes. Members of a successful team value each other's perspectives. Team goals are set before individual goals, and task inputs provided by team members are appreciated as collective involvement to team performance (Wilson, Salas, Priest, & Andrews, 2007). Team orientation has been found to facilitate decision making, cooperation and coordination within teams (Salas, Sims, & Burke, 2005), and it is considered to be a mediating factor for mutual performance monitoring and backup behavior.

2. Backup behavior

> The team showed a high degree of backup behavior (team members helped/assisted without being asked, push of information). (RNoNA assessment tool, 2012)

Backup behavior means that team members are providing and requesting assistance when needed. It involves assisting other team members in completion of tasks or correcting mistakes (Brannick, Salas, & Prince, 1997), and shifting workload to optimize the distribution of resources (Marks, Mathieu, & Zaccaro, 2000). Team members who actively assist one another are knowledgeable of each other's roles and responsibilities (Wilson, Salas, Priest, & Andrews, 2007) and can assess the performance of teammates through mutual performance monitoring (Salas, Sims, & Burke, 2005). Empirical research has shown that backup behavior improves performance and reduces the negative effect of stressors (Wilson, Salas, Priest, & Andrews, 2007) and is a mediating factor to team adaptability in dynamic environments (Salas, Sims, & Burke, 2005).

A challenge for assessment of backup behavior is that not all team members will engage in it since it depends on team tasks, team needs and team structure. Different team tasks give team members different opportunities for engagement. For example, a team member whose task is to drive a vehicle is probably unable to assist a team member who needs help with troubleshooting a radio and will therefore not offer assistance. If the radio is considered more important than mobility, then the team need may dictate that the driver stops the vehicle to assist with the radio. The team structure and composition (KSAs) may be such that only some team members can assist with the troubleshooting, and not everyone will display backup behavior for that specific task.

Backup behavior is more than simply helping out based on a request; it includes shared mental models, trust and mutual performance monitoring that facilitate the understanding of implicit and explicit need from team members in order to be effective (Salas, Sims, & Burke, 2005).

3. Mutual performance monitoring and mutual trust

> The team adjusted and reinforced each other (feedback when wrong or right was accepted and implemented by team members). (RNoNA assessment tool, 2012)

Mutual trust is a reciprocal process between team members based on risk acceptance and information sharing to support cooperation in interdependent teams (Salas, Sims, & Burke, 2005). Risk and interdependence are two conditions considered necessary for trust, where the source of risk can be described as uncertainty about the intentions of others and interdependence as the reliance team members have upon one another for a successful outcome (Rosseau, Sitkin, Burt, & Camerer, 1998). The development of trust in small military teams is conditional to the situational factors of risk, vulnerability and uncertainty (Adams & Webb, 2002). Trusting teams allow team members to confront each other in an effective manner in order to achieve team goals without fearing reprisals.

> Feedback shall be considered a gift. Being a part of this team, you shall assume that your peers have your best interest at heart when giving you the gift. As with all gifts, it is yours to use as you wish, you can use it straight away, you can store it for later use, and you can even decide not to use it, it is your choice. The only thing I ask of you is that you accept the risk of trusting that the senders mean you well. (Myran, 2008)

The quotation is from a cadet team training session at the RNoNA. Commander (CDR) Myran's intention is for the cadets to experience that feedback is more easily accepted when the receiver expects good intentions from the sender.

Research has shown that teams with a high level of trust appear to be more capable of managing stressors like uncertainty and complexity than those with low trust levels (Jarvenpaa & Leidner, 1999). Trusting teams will be more willing to appear uniform in their behaviors and actions, encouraging backup behaviors and mutual performance monitoring and willing to allow the team leader to effectively manage the team (Salas, Sims, & Burke, 2005).

Mutual performance monitoring is the awareness and observation of behavior, actions and level of performance of other team members (Dickinson & McIntyre, 1997).

Successful monitoring requires that team members have an understanding of the team tasks and each other's roles and responsibilities (i.e., a shared mental model) in order to provide effective feedback (Wilson, Salas, Priest, & Andrews, 2007). Fellow team members observe each other's performance while conducting their own tasks, a behavior that becomes more important when stress increases. Overloaded team members are more likely to make mistakes, and team member feedback can reduce failures and increase performance (Cannon-Bowers & Salas, 1998). Mutual monitoring increases the teams' ability to realize effective and ineffective team performance and facilitates self-correcting adjustments through timely and accurate feedback to achieve team goals (Flin, O'Connor, & Crichton, 2008).

4. Closed-loop communication

> The team exchanged information and coordinated actions through feedback and response. (RNoNA assessment tool, 2012)

Communication is the exchange of information, feedback and response, and is a key activity in coordination among team members to achieve team goals. Communication problems are known to contribute to many incidents and accidents. Failure to exchange information can be the difference between good and poor team performance (Flin, O'Connor, & Crichton, 2008).

Closed-loop communication is a three-step sequence in which the information is transmitted by the sender, received and acknowledged by the recipient and clarified by the sender that the message was correctly interpreted (Salas, Sims, & Burke, 2005). Employing closed-loop communication techniques ensures that information is clearly and concisely transmitted, received, and correctly understood (Wilson, Salas, Priest, & Andrews, 2007).

5. Team leadership

> The leader was effective to solve team problems (roles and responsibilities were distributed in the team). (RNoNA assessment tool, 2012)

Team leadership includes the ability to provide direction and facilitate team problem solving, clarify team member's roles and responsibilities, coordinate activities, convey team performance expectations and assess its performance (Salas, Sims, & Burke, 2005). Team leadership involves a continuous monitoring and evaluation of the mission environment to reach a judgment or choose an option to meet the needs of a specific situation. Not all situations or team structures require a single individual that acts as a team leader; other team members can apply team leadership skills as well (Brannick, Salas, & Prince, 1997) by using authority and assertiveness, maintaining standards, allocating resources and managing workload (Flin, O'Connor, & Crichton, 2008).

The team leader can positively affect mutual performance monitoring and backup behavior through support, motivation and feedback. Team leadership supports team adaptability (Salas, Sims, & Burke, 2005) that ensures the team adjusts strategies based on information gathered from the environment. Effective team leaders enable teamwork and taskwork processes by creating shared mental models of the mission objective(s), the team tasks and goals, the team structure and recourses available to the team (Salas, Sims, & Burke, 2005).

6. Shared mental models and interdependence

> The team showed the ability to create a common outlook (all team members are kept updated on the objectives, situation and priorities, both for teamwork and taskwork objectives, "what if"-processes). (RNoNA assessment tool, 2012)

Shared mental models (SMM) is a shared understanding among team members of the team's goals, its tasks, and how the team members will interact and adapt to the evolving demands of the task and the needs of team members (Salas, Sims, & Burke, 2005). Interdependence among team members will depend on role assignment and role definitions in the team structure.

Experienced team members have a store of mental models for the domain they work in. An emerging situation does not need to be an exact match to a previous situation to be recognized. However, it does need to present enough features for a categorization into type of model to be made (Flin, O'Connor, & Crichton, 2008).

> Teams who develop a high level of congruence between their mental models, both situational and mutual, are able to make use of these models to anticipate the way the situation will evolve as well as the needs of the other team members. (Serfaty, Entin, & Johnston, 1998)

This congruence and mutual anticipation contributes significantly to the team's performance under stress. Team members who use available time to update the common understanding of the evolving tasks and team needs will consistently perform better under a wide range of tactical conditions (Orasanu, 1990). Developing SMMs among team members allow teams to communicate and coordinate implicitly rather than explicitly during high-workload conditions, making cognitive resources available for other tasks (Stout, Cannon-Bowers, Salas, & Milanovich, 1999).

Two types of SMMs can be distinguished as a team's model of its teamwork and of its taskwork. Shared teamwork-based and taskwork-based models have been found to relate positively to team processes and performance (Mathieu, Goodwin, Heffner, Salas, & Cannon-Bowers, 2000). Shared teamwork models include team orientation, team cooperation and expected behaviors. Shared taskwork models include the understanding of each other's tasks, responsibilities and KSAs as requirements for mutual performance monitoring, team coordination, team leadership and adaptability (Salas, Sims, & Burke, 2005). The research by Mathieu et al. (2000) suggests that team and task mental models have differential influences on team processes and performance. For assessment of a specific team, one needs to understand the interdependence required for the task presented in the training environment and to facilitate team member interactions that are both expected and required for task performance (Paris, Salas, & Cannon-Bowers, 2000).

7. Adaptability

> The team showed the ability to adjust strategies (dynamic co-ordination to meet shifting internal and external needs). (RNoNA assessment tool, 2012)

Adaptability is considered a critical team skill in the dynamic and complex environment experienced by military teams (Johnston, Serfaty, & Freeman, 2003). It enables the team to recognize mismatches and appropriately adjust actions (Salas, Sims, & Burke, 2005). The level of adaptability is highly dependent on the SMM of team goals and mission objectives (Kozlowski, Gully, Salas, & Cannon-Bowers, 1996). Adaptation to stressors can take the form of changes to processes, new communication styles and/or coordination strategies, or structural changes, reallocation of resources or restructuring of team roles and responsibilities (Serfaty, Entin, & Johnston, 1998). The adaptation model shown in Figure 6 (Entin & Serfaty, 1999) shows a framework for how well-trained teams maintain performance in the presence of environmental friction. By adapting their decision-making strategy, coordination strategy and organizational structure to meet the demands of the situation, they are able to keep stress below an acceptable threshold (Serfaty, Entin, & Johnston, 1998).

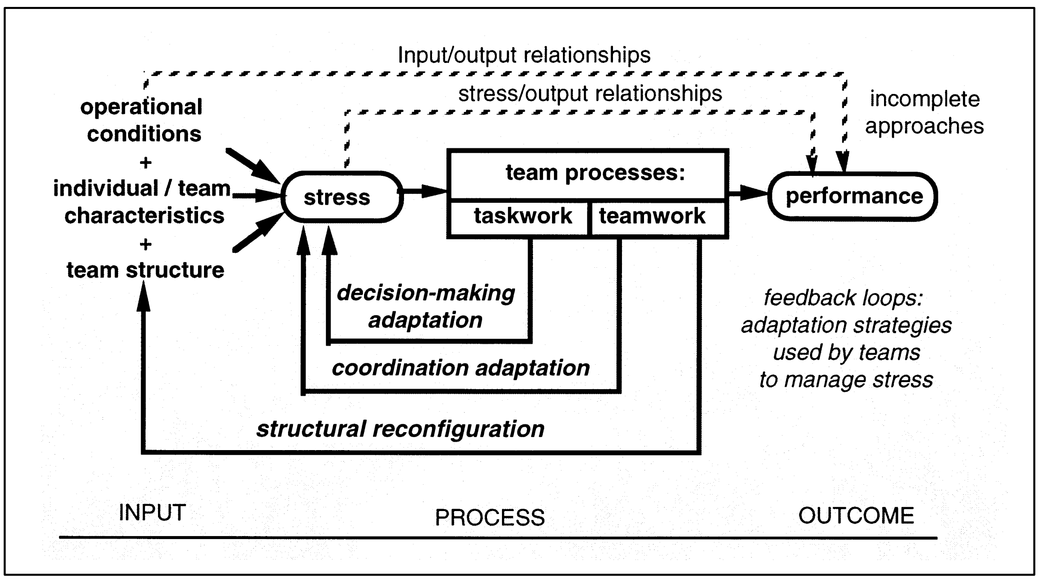

Figure 6. Theoretical framework for team adaptation (From Entin & Serfaty, 1999).

Adaptability can be manifested in many different ways depending on the team goals and objectives. It requires SMM, mutual performance monitoring and backup behavior to be effective (Salas, Sims, & Burke, 2005). Team adaptability includes the need for behavioral and operational change when the situation demands it (Wilson, Salas, Priest, & Andrews, 2007). In order to perform swift changes to behavior or structure, the team must both anticipate and accept that changes are necessary during the mission and must be sufficiently aware and heedful of changes in the environment that suggest changes. When teams accept that changes are inevitable in military missions, they gain the awareness necessary to employ action to overcome the friction.

8. Agility

> The team showed the ability to rapidly change their orientation in response to what is happening (monitor, detect and respond to resource allocation needs, *e.g., alert and ready to move*). (RNoNA assessment tool, 2012)

The Oxford definition of agility is "the ability to move quickly and easily" (Oxford Dictionaries, 2013), which can also relate to a quick and alert mind. A more scientific definition is presented by David S. Alberts as a "synergistic combination of robustness, resilience, responsiveness, flexibility, innovation, and adaptation" *(Alberts, 2007)*. John Boyd referred to agility as "the ability to rapidly change one's orientation in response to what is happening in the world" (Boyd, 2005), actions that require team members to focus attention on internal and external changes, and compare those changes to existing SMMs.

An agile behavior style requires team members to interact with the environment actively and not isolate themselves from it. A constant monitoring and assessment of the evolving mission is critical to determine the appropriate response needed to approach different circumstances (Alberts, 2007).

> The ability to gain victory by changing and adapting according to the opponent is called genius. (Sun Tzu, 500 BC)

Mental agility has been described as the visualization of team goals and objectives from different perspectives to match important cues in the environment. This mental readiness speeds up the decision-making process, and immediate actions can be taken (Osinga, 2005). Agile team action is revealed by rapid decisions and actions (Richards,

2005) and by innovative thinking that results in creative solutions (and speed) to outmaneuver an opponent (Boyd, 2005).

F. TASKWORK

Taskwork is behavior related to operational activities the team members must perform. Typical taskwork behavior includes understanding the mission demands and objectives, SOPs and information about the mission (Flin, O'Connor, & Crichton, 2008). The four taskwork components included in the RNoNA tool are creative action, speed, thoroughness and success.

1. Creative action

> The team was creative in their actions (taking action to generate and exploit advantage over the situation/opponent to achieve their objectives, *e.g., cause friction to opponent, "command both sides"*). (RNoNA assessment tool, 2012)

Creative actions are novel approaches to a situation (Ford, 1996) where teams avoid anticipated rules and/or actions characteristic of that domain. Creative action generates unexpected changes in the mission environment, allowing agile teams to take advantage of the new opportunities presented (Boyd, 2005; Richards, 2005).

Creative action can be proactive (focus on agility), passive (by convenience) or reactive (by necessity). Proactive teams actively use creative action to shift friction to the opponent, away from own team goals and mission objectives (Brehmer, 2005). Creativity includes adaptations of habitual practices necessary to fit a specific situation (Dalton, 2004). Creative actions in passive teams are therefore only likely to emerge if the team expects the actions to produce outcomes relatively more desirable than the outcomes of habitual actions (Ford, 1996). Reactive creativity can occur when familiar procedures and actions have failed, and the only way out is a novel course of action.

A team's self-efficacy (Bandura, 1977) is influenced by teamwork processes and can motivate the use of creative actions over habitual actions to achieve team goals and objectives.

2. Speed

> The team was effective to complete assignments (short time, appropriate method and strategy). (RNoNA assessment tool, 2012)

Time is a dominant concern in warfare. The gap between planning and starting an action is considered a time delay in military operations (Brehmer, 2005). Speed includes the correct and timely coordination of actions that contribute to the completion of tasks. The gap can be made smaller when team members are able to sequence, synchronize, integrate and complete tasks without wasting valuable time and resources (Wilson, Salas, Priest, & Andrews, 2007). High performing teams conduct situation assessment, planning and execution faster than low performing teams (Salas, Rosen, Burke, Nicholson, & Howse, 2007). Speed also has been listed as an important factor for achieving mission objectivesby acting faster than your opponent (Boyd, 2005). The execution of an order must be monitored and quickly assessed to avoid time delays (Brehmer, 2005).

Team coordinating activities are essential to reach conclusions and make decisions rapidly (Wilson, Salas, Priest, & Andrews, 2007). Adaptability and agility are key characteristics for converting the established ideas into action (Richards, 2005). Actions must be delivered at the right time, in the right number and at the right level of intensity to accomplish the goals (Royal Norwegian Naval Academy, 2009).

3. Thoroughness

> The team was thorough in their assignments (solutions and actions that fit with the stated plan). (RNoNA assessment tool, 2012)

Thoroughness is the ability to maintain commitment and determination (Bandura, 1977). Commitment means to continuously evaluate mission objectives (Boyd, 2005). Determination means to challenge the situation if objectives are not sufficiently achieved (Brehmer, 2005). Thorough teams bounce back from adversity, showing resilient behavior (Hollnagel, Woods, & Leveson, 2006; Reivich, Seligman, & McBride, 2011). Thoroughness is also the capacity to be sufficiently (resource dependent) detailed in planning and to employ solutions that fit with the stated plan. Thorough teams monitor the execution of actions by appropriate means of feedback (Brehmer, 2005) and assess critical change factors that require alterations to the plan. Thoroughness includes cognitive hardiness, a mechanism found to mediate stress in military units (Bartone, 2006) and contains the elements of commitment, control and challenge (Maddi, Matthews, Kelly, Villarreal, & White, 2012):

- Commitment contains the ability for involvement.
- Control implies the belief that it is possible to influence incidents that occur.
- Challenge has to do with a belief that change in itself is exciting and positive.

4. Success

> The team successfully accomplished the task/mission (based on the current training objectives/requirements). (RNoNA assessment tool, 2012)

Successful achievement of task and mission objectives has occurred when certain criteria have been met or a distinct advantage has been established (Richards, 2005). In a training situation this is illustrated when the team accomplishes training objectives. Military victory, on the other hand, is determined by the status of your adversaries and is not finally accomplished until the opponent has surrendered (Boyd, 2005).

In a training environment it is especially important to assess how success was accomplished. A team can achieve a successful outcome without good team performance, either by luck or by mistakes made by an opponent, and vice versa can display high performance but not achieve the objectives. Effectiveness criteria in a given training environment are dependent on the current training objectives, the mission requirements and on the teams' expected level of proficiency at the time of assessment (Ross, Phillips, Klein, & Cohn, 2005).

G. SCENARIO-BASED TEAM TRAINING

Successful training programs support active participation and learning experiences, and consistently incorporate components that reflect and meet the demands of the mission environment (Robertson, 2002). RNoNA scenario-based team training (Figure 7) aims at designing stressors found in a complex military task environment into realistic simulated exercises for team training. The RNoNA NAVSIM represents an affordable, adjustable and controllable technological setting that actively provides useful opportunities for improvements to train RNoNA cadet teams.

Training that involves the use of technology solutions is often described by which of three types of environment the solution employs: live, virtual or constructive training solutions (USN Office of Naval Research, 2011). Live training solutions (e.g., field exercises) involve real users, operating real systems and operating in the real world. The environment is closely similar to the operational environment, but the scenarios, systems and platforms operated are adjustable and controllable (Testa, Aldinger, Wilson, & Caruana, 2006). Virtual training solutions involve real users, operating real systems, where a synthetic environment replicates the operational environment (e.g., The RNoNA ship-handling simulator) (Shufelt, 2006). Constructive training solutions involve simulated forces that respond to real user inputs and carry out actions in a synthetic environment where all entities and activities are simulated (e.g., war-gaming models) (Testa, Aldinger, Wilson, & Caruana, 2006). For this thesis, *virtual* will normally refer to RNoNA simulator systems, and *live* will refer to RNoNA field exercises.

Delivery methods of scenario-based team training can range from small exercises specially designed to teach a single skill, to an event-based approach that assesses stressors within an event (Cannon-Bowers & Salas, 1998), all the way to large-scale scenarios and holistic elicitation methods (Cooke, Salas, Cannon-Bowers, & Stout, 2000). The small, single-skill sessions are presumably true to the training objective, and performance is relatively easy to evaluate. However, they tend to be static and may not fully engage the trainees. In contrast, scenario-based training is more dynamic and can contain a number of embedded training objectives. The downside of scenario-based training is that it is more difficult to isolate and trigger events, and it requires more time to observe behavior to record and diagnose performance. In scenario-based training *the scenario itself is the curriculum (Cannon-Bowers & Salas, 1998)*. This means that the scenario must be crafted and executed in a manner that accomplishes the desired training objectives.

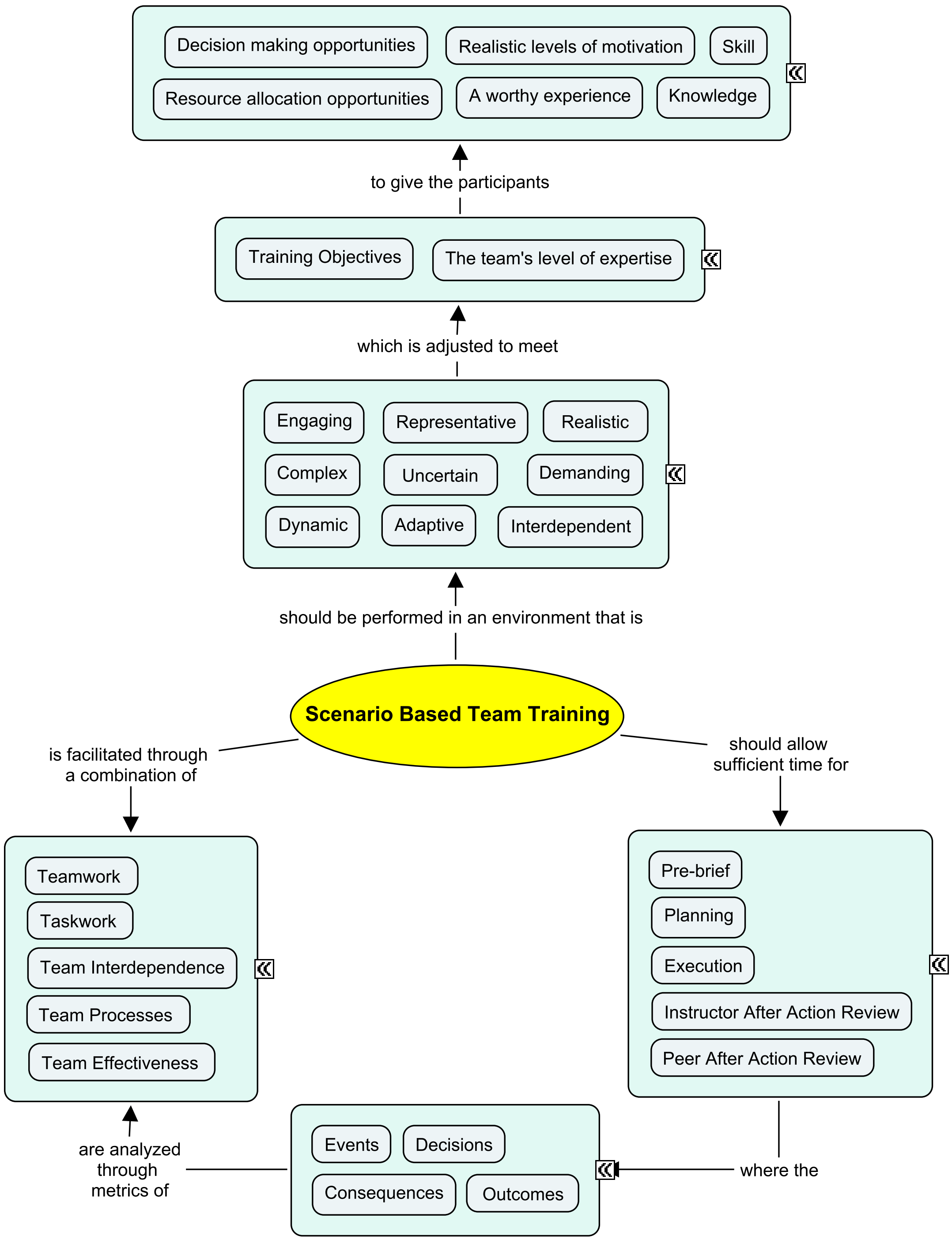

Figure 7. Scenario-based team training concept map.

Research suggests that adaptive training can reduce complexity and friction through comprehensive use of flexible simulator systems and realistic scenario-based training interventions (Entin & Serfaty, 1999). Teams can be trained to employ strategies to adapt to stressful (high-workload) situations. However, the introduction of technology does not guarantee effective team training by itself (Salas, Cooke, & Rosen, 2008). Without transfer of learning from the training intervention to the working environment, even a gold-plated simulator system is of little use. For a specific team, one needs to understand how much shared knowledge is required for the task presented in the training environment, and facilitate team member interactions that are both expected and required for task performance (Paris, Salas, & Cannon-Bowers, 2000).

Team training should be engaging and representative of the training objectives. The level of difficulty must be adjusted to meet the training objectives and the team's level of expertise. Those two factors decide how complex, demanding and dynamic the task must be to foster task interdependence and team adaptability. A well-balanced training session should give the participants decision-making opportunities, resource allocation opportunities and realistic levels of motivation to increase (or maintain) team competencies (Flin, O'Connor, & Crichton, 2008). A simulator session that represents a worthy experience to the participants increases the likelihood of reflection and learning.

H. SUSPENSION OF DISBELIEF

The worthy experience is achieved when both facilitator and trainee have a feeling of ownership and commit to support the training objectives. Active involvement by the organization, decision makers, training facilitators and the trainees assures the motivation required to develop knowledge, skills and abilities. In *Safety at the Sharp End*, Flin, O'Connor and Crichton (2008) describe the "A" in KSA as *attitude* rather than *ability,* an important distinction in a simulator environment where the notion of attitude is paramount for the motivation to train. To achieve learning, trainees must reflect upon the experience they just had in the simulation. It is easy to deflect blame and excuse one's inabilities or shortcomings by saying that "a simulator is not real life," and that one's "actions would naturally have been different if the mission had been for real," and so on.

A training program that lacks active participation will have a high resistance to learning behavior (Robertson, 2002).

Samuel Taylor Coleridge, a writer of poetry and lyrical ballads, coined the term "willing suspension of disbelief for the moment" in a poem in his work *Biographia Literaria (Coleridge, 1817)*. He argued that a willing suspension of disbelief would make fictional characters more real and therefore grant the viewer/reader a richer experience (Jacobsen, 1982). This idea also seems to have been important to William Shakespeare when he designed the famous Globe Theatre in London in 1599. He wanted the Globe's audience to explore and experience interactions with the players when attending a play (Wikipedia, 2012).

A properly designed and facilitated training intervention, whether in simulator or field setting, will motivate the trainee to believe in its value. The purpose is not to ask or coerce the trainees to step aside from reality, but to accept the secondary reality as a means to explore and develop new knowledge, skills, abilities and attitudes. The outcome and value of a training experience is highly dependent on the trainee's acceptance of the training curriculum and a willing suspension of disbelief for the moment.

I. SUMMARY

This discussion has reviewed the literature that suggests that team performance can be measured. The work by Salas and others gives meaning to the use of the twelve measures chosen in the RNoNA assessment tool to assess team performance in simulator and field settings.

Based on those findings, this thesis discusses a test of the hypotheses that the RNoNA team performance assessment tool of Figure 2 (a) provides useful measures of team performance in a pair of simulator exercises, (b) predicts team performance in a subsequent live exercise, and (c) uncovers whether the stressors designed into the two simulator exercises match the stressors of the live exercise.

III. METHODOLOGY PART 1 – ASSESSMENT TOOL

A. DESIGN AND CONSTRUCTION

At the RNoNA, team performance assessment is performed through observation of task execution and team member interactions and scored using the RNoNA Team performance assessment tool. Behavioral markers and performance markers are designed to assess performance in teamwork and taskwork by measuring team processes and team effectiveness in RNoNA cadet training exercises.

The 2012 version of the RNoNA assessment tool is shown in Figure 8. The tool includes twelve categories, scored using 7-point Likert scales. The first eight describe teamwork, and the last four categories describe taskwork processes. Each of the twelve categories has a short description to give further meaning to the category and to describe the level of performance associated with the anchors on scales. At the bottom of the tool, there is a comment section where SMEs can fill in additional information on team behavior, special assignments that can explain scores, overheard comments, etc. that can further describe their assessment.

The teamwork processes include the behavioral interactions team members must develop and perform to function effectively as a team, as discussed in Chapter II.E: team orientation, backup behavior, mutual performance monitoring (includes mutual trust), closed-loop communication, team leadership, shared mental models, adaptability and agility (Alberts, 2007; Brannick, Salas, & Prince, 1997; Cannon-Bowers & Salas, 1998; Entin & Serfaty, 1999; Espevik, Johnsen, & Eid, 2011; Salas, Sims, & Burke, 2005; Zaccaro, Rittman, & Marks, 2001).

The last four categories describe taskwork processes, as discussed in Chapter II.F: creative action, speed, thoroughness and success (Bandura, 1977; Boyd, 2005; Brehmer, 2005; Dalton, 2004; Hollnagel, Woods, & Leveson, 2006). These are resilient behaviors related to operational activities that a team must perform. They are evaluated from the outcome of individual and team tasks and actions (Flin, O'Connor, & Crichton, 2008).

RNoNA Team Performance Assessment **Team:** **Rater:**

Teamwork

1. **Team Orientation:**
The team showed a high degree of involvement (team members monitored and paid attention to other team members, not many "free riders" in the teamwork process)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

2. **Backup Behavior:**
The team showed a high degree of backup behavior (team members helped/assisted without being asked, push of information)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

3. **Mutual Performance Monitoring (includes Mutual Trust):**
The team adjusted and reinforced each other (feedback when wrong or right was accepted and implemented by team members)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

4. **Closed-loop Communication:** The team exchanged information and coordinated actions through feedback and response
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

5. **Team Leadership:** The leader was effective to solve team problems (roles and responsibilities were distributed in the team)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

6. **Shared Mental Models / Interdependence:** The team showed the ability to create a common outlook (all team members are kept updated on the objectives, situation and priorities, both for teamwork and taskwork objectives, "what if"-processes)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

7. **Adaptability:** The team showed the ability to adjust strategies (dynamic co-ordination to meet shifting internal and external needs)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

8. **Agility:** The team showed the ability to rapidly change their orientation in response to what is happening (monitor, detect and respond to resource allocation needs, *e.g. alert and ready to move*)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

Taskwork

9. **Creative Action:** The team was creative in their actions (taking action to generate and exploit advantage over the situation/opponent to achieve their objectives, *e.g. cause friction to opponent, "command both sides"*)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

10. **Speed:** The team was effective to complete assignments (short time, appropriate method and strategy)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

11. **Thoroughness:** The team was thorough in their assignments (solutions and actions that fit with the stated plan)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

12. **Success:** The team successfully accomplished the task/mission (based on the current training objectives/requirements)
Strongly Disagree *Partially agree* *Strongly Agree*
1 2 3 4 5 6 7

Comments (fill in additional information, team behaviors, phrases, etc. that can further describe the assessment) :

Figure 8. RNoNA Team Performance Assessment Tool (2012 version).

B. RESILIENCE AND VALIDITY

The array of factors in the RNoNA assessment tool emphasizes the complexity inherent in effective team performance and in its assessment. An assessment tool for military teams can be looked upon as any test or procedure administered to evaluate mission essential competencies (MECs), motivation or fitness for deployment. The accuracy with which assessment scores can be used to forecast performance on the job is the tool's most important characteristic, referred to as predictive validity (Schmidt & Hunter, 1998). The effectiveness assessment criteria can be based on effects that are quantitative, qualitative or a combination of both (Kiekel & Cooke, 2011).

Assessing team performance should therefore begin with an understanding of the tasks to be performed in the projected operational environment (Cannon-Bowers & Bowers, 2011). The RNoNA established the training objectives investigated in this research: training cadet teams for squadron level missions in complex maritime environments. A squadron is built up of several teams, and the command structure relies on teams to perform tasks in coordinated efforts. The complex environment includes uncertainty and stressors and sets a high demand for team and team-member resilience (Royal Norwegian Naval Academy, 2010). The demand for resilience requires the cadet teams to develop *an ability to persist in the face of challenges and bounce back from adversity* (Reivich, Seligman, & McBride, 2011). Among the measures of resilience in teamwork are interdependence, adaptability and agility (Hollnagel, Woods, & Leveson, 2006). Resilient teams must demonstrate the ability to rapidly change orientation in response to what is happening in the real world (Boyd, 2005) and adjust strategies through dynamic coordination to meet shifting internal and external needs (Wilson, Salas, Priest, & Andrews, 2007). To be effective in naval combat environments, team processes must translate into appropriate actions where the team exploits advantages in the environment and shifts friction from themselves to the opponent (Brehmer, 2005). Correct and timely coordination of actions is vital to achieving mission objectives.

These considerations suggest that mission objectives in RNoNA team training exercises can be achieved through resilient behavior, where the team follows through to re-engage if goals are not sufficiently fulfilled, even if failure threatens. The RNoNA tool includes categories of teamwork and taskwork processes and outcomes to assess the resilient behavior in cadet teams.

IV. METHODOLOGY PART 2 – EMPIRICAL TESTS

> The military Officer carries the sole responsibility when it comes to defending society with military force, thus implicating the large personal sacrifice, death as the last resort. This requires Officers with a certain sentiment of duty to the Norwegian society as well as Officers with the ability to operate and lead in complex combat environments. (Wedervang, 2009)
>
> Captain Thomas T. Wedervang, Chief Royal Norwegian Naval Academy

The RNoNA develops important capacities and capabilities for leadership on individual as well as on team levels. Operational leadership is improved, first and foremost, through realistic and good experience and through improvement of behavioral patterns. Good leadership is created through repetition, and through gradual, increasing challenges. Courage to challenge one's limits is encouraged and stimulated through realistic military exercises, developing mature leaders and teams (Royal Norwegian Naval Academy, 2009). These exercises provide an opportunity to conduct quasi-experimental research. The focus of the 2012 exercises was primarily on training and not on this research, although this research benefited from observational studies of the RNoNA exercises.

Counseling and systematic reflection play a key role in developing emotional and cognitive maturity. The cadets are familiar with being observed, evaluated and receiving feedback from their instructors and counselors during their training. They will therefore consider the presence of SMEs to evaluate their performance during the simulator and live exercises normal procedure.

This chapter discusses how RNoNA conducted the exercises and how the SMEs used the RNoNA assessment tool to gather the data analyzed in Chapter V. The exercises cannot be discussed in detail, however, mainly for two reasons: (1) Some details are not considered public information, and (2) too much openness can affect future cadets' expectations for similar exercises and consequently reduce their learning.

A. PARTICIPANTS

The Royal Norwegian Naval Academy combined 72 first year cadets to form eight teams. Each team had both male and female cadets with one to four years of prior service in the Norwegian military. The cadets ranged in age from 20 to 33 (M=25). Prior to the first simulator exercise, they had been training as teams for seven months.

RNoNA Staff functioned as facilitators, SMEs, educators, instructors, etc. throughout the research. An effective evaluation system uses people who are highly skilled and knowledgeable in a particular field (Proctor & Zandt, 2008), also known as subject matter experts. The SMEs participating in the study were all officers in the Norwegian Navy, with military rank ranging from Sub Lieutenant to Commander.

B. DESIGN

All eight teams performed the same exercises in a repeated measures design. Two simulator exercises, *Carey* and *Aden*, were performed in January and April 2012, followed by a live exercise, *Dolphin*, in June the same year. All three exercises were run as "controlled free-play" exercises. *Controlled* means that the exercise has a framework that includes a main mission, sub-missions, orders, intelligence reports, time schedules and a command & control hierarchy. *Free-play* means that the cadets are given extensive leeway to plan and execute missions based upon their own interpretation and assessment of the mission objectives and current situation. Comprehension, decisions and actions emerge from the teams' own processes, greatly influencing the course of the scenario. One implication of controlled free-play is that there is no blueprint for what constitutes a success or a failure.

1. Simulator exercise, Carey

The Carey scenario was based on actual historical events from World War II in the North Sea and represented realistic uncertainty and *fog of war* in a complex maritime environment. Carey was conducted as a covert operation where avoidance of detection was critical. The scenario was set in the 1940s, which limited the level of technology available to the teams.

An operation named “Cartoon” which took place in late 1942 was used as a source of inspiration for the Carey scenario design. The following narrative is a brief summary of operation Cartoon as described on the website Shetlopedia (Shetlopedia, 2011). The information is largely based upon *The Waves Are Free,* a book by James W. Irvine, published in (1988).

The Lerwick-based Norwegian Motor Torpedo Boat (MTB) flotilla operations on the Norwegian coast are an important part of the Norwegian and British naval history during World War II. A picture of a typical MTB, the Fairmile D, is shown in Figure 9. The flotilla of MTBs was a nuisance for the Germans; the Nazis never knew where the next attack would come from. The Germans strengthened defenses along the Norwegian coast to meet the threat, and that required many extra men, who otherwise could have fought in other parts of Europe.

Figure 9. Fairmile D Motor Torpedo Boat (From Wikipedia, 2013).

The December weather in 1942 had been extremely bad. On the 27th, MTB *619* and two other MTBs encountered a severe storm; they had to return to Shetland without fulfilling their mission, but all arrived safely. Some weeks later, in the early hours of the 23rd of January 1943, they set out on one of their most successful and important operations, "*Operation Cartoon.*" The target was the German pyrite mine at Litlabø on the island of Stord. Two Norwegian fishing boats, the *Gullborg* and *Sjølivet*, had for some time blended with the local fishing boats in the area and gathered information about German activity.

Seven MTBs took part in the operation. Fifty commandos of the *12th Command*, many of them Norwegians, under the command of Major Flynn, were on board the seven boats. MTB *626*, under the command of Lt. Bøgeberg, was the leading vessel, with Commander in Chief, Lt. Tamber on board. German airplanes soon spotted the flotilla, and Lt. Tamber ordered a more southerly course to mislead the Germans. The plan was to land most of the men in the little harbor of Sagvåg, on the southwest side of Stord, not far from the pyrite mine. MTB *626* and MTB *627* were to attack the harbor and land the commandos; the other boats would take care of any kind of German interruption. MTBs *618* and *623* were to cover the northern flank, while MTBs *620*, *625* and *631* covered the south and east areas. MTBs *626* and *627* fired torpedoes into the harbor and blew up the pier and a cannon and then followed up the attack with fire from all their weapons. At midnight MTB *626* landed the commandos on what was left of the pier, while MTB *627* found a small pier on the other side of the harbor to land theirs. The Germans had of course returned the fire, so they were under constant fire during the landing; one commando was hit and killed. The two MTBs used all their firearms to give covering fire. Then they just had to wait.

Meanwhile, the other boats did all they could to confuse the Germans, to make them believe that something else was the main target. They attracted a lot of German attention and were fired at from the coastal forts. MTBs *620* and *631* went into the harbor of Leirvik, while MTB *625* laid mines around the east side of the island. MTB *620* also fired on the German ship *"Ilse M. Russ,"* which caught fire and grounded. On land, the commandos had done their job; the mine was destroyed and did not become operational again for more than a year, and the harbor defenses were destroyed.

But the victory had taken its toll. One commando was killed and two wounded, and on the MTB *626*, seven crewmembers were wounded. MTBs *618* and *623*, which were going northward before the attack on the pyrite mine, came under fire from a German coastal battery on Store Karlsøy, both boats were damaged, and three of the crew were wounded. A German Junkers JU88 attacked MTB 625 on the return, but they returned the fire, and the plane went down. Finally all MTBs made it safely back to Lerwick.

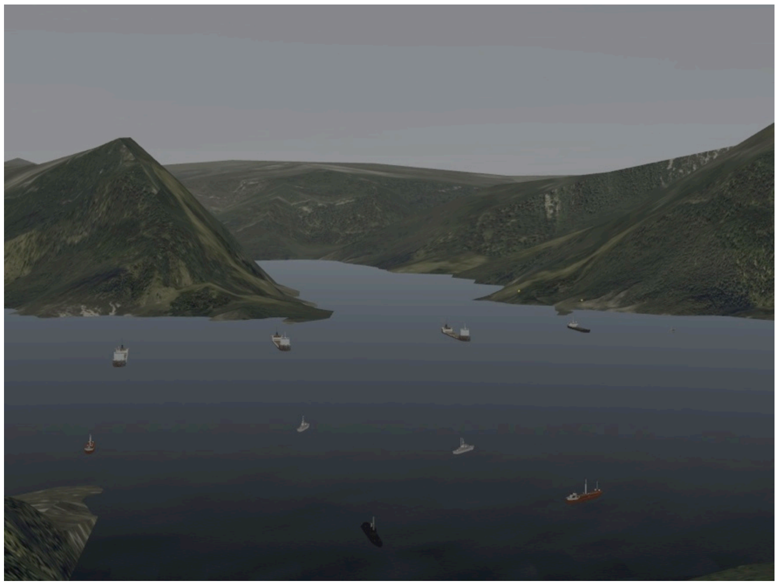

Figure 10. Screenshot from the simulator exercise Carey.

The Carey exercise recreates some of the stressors identified from operation Cartoon, such as uncertainty, vulnerability and information ambiguity. Carey is a complex simulator exercise that places a high demand on resilient behavior in the cadet teams.

2. Simulator exercise, Aden

The Aden exercise is a modern and realistic anti-piracy scenario set in the Gulf of Aden. The Aden scenario is conducted as an overt operation where visibility and presence of force is important. The high levels of communication, coordination, cooperation and the extensive use of technology that one would expect to find in modern allied naval operations are represented in the Aden scenario. Mission descriptions and regulations for Operation ATALANTA (EUNAVFOR) are used to build this simulator exercise to construct a realistic and modern military context and environment.

EU Naval Force (EUNAVFOR) operates in the Southern Red Sea, the Gulf of Aden and a large part of the Indian Ocean, including the Seychelles. Part of their mission is to deter, prevent and repress acts of piracy and armed robbery in the area of operation (AOO) and protect vessels of the United Nations' World Food Programme (WFP) delivering aid to displaced persons in Somalia (EUNAVFOR, 2013).

The Aden exercise performed in 2012 started with the following information to the cadet teams:

Due to the dramatic increase in the number of pirate attacks on civilian vessels in the Gulf of Aden, The Norwegian government has decided to order a Norwegian Fast Patrol Boat Squadron (the cadet teams), detached from a joint operation with the Saudi Coast Guard, to join the task group TG 432.01. UN Security Council Resolutions 1814, 1816, and 1838 are in effect, authorizing the use of all necessary means to fight pirates off Somalia's coast.

Figure 11 shows the initial setup for the exercise with the respective areas of responsibility (AOR) for the Norwegian Navy ships. It also illustrates the graphical user interface (GUI) for the NAVSIM instructor.

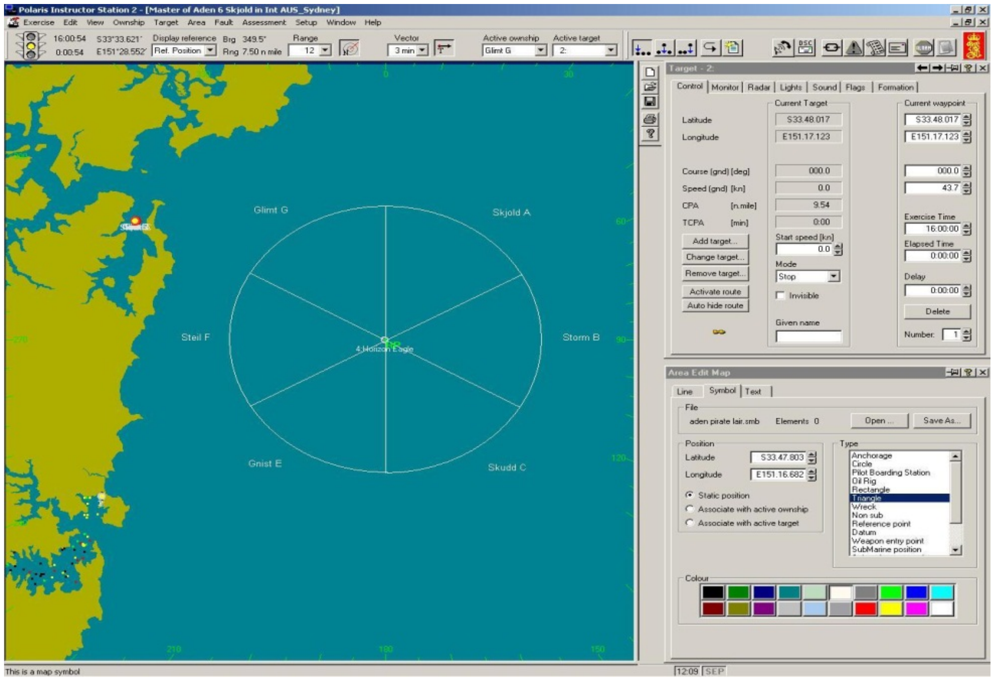

Figure 11. Screenshot from NAVSIM instructor station GUI.

The cadet teams received all necessary information for the AOO: intelligence briefs, rules of engagement (ROEs), OP ATALANTA orders, maritime security information (Figure 12), political situation, and policies for civilian shipping (e.g., EUNAVFOR booklet for Counter Piracy - Advice and Checklist for Masters, 2009), and used the available information to plan and execute the mission.

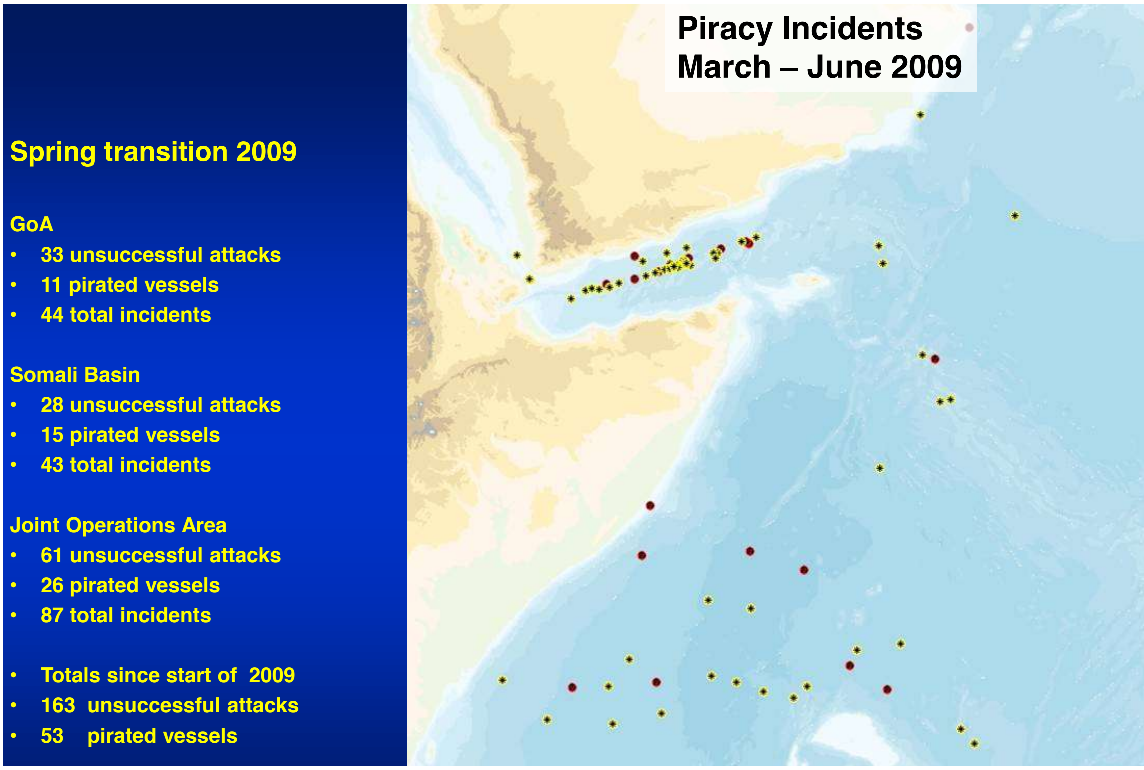

Figure 12. Piracy incidents in the Gulf of Aden (From Foster, 2009).

The Aden exercise recreates stressors identified from modern military operations, like powerlessness, ambiguity, boredom, danger and workload. It is a complex simulator exercise that places a moderate demand on resilient behavior in the cadet teams.

3. Live exercise, Dolphin

Exercise Dolphin was one of several live exercises making up the final training stage for the cadet teams in 2012. The cadet teams experienced a complex maritime military scenario where events varied in intensity and content. The exercise was conducted as a combat survival course presenting operational leadership challenges for the individual, team and squadron levels during periods of high physical and mental stress, combined with sleep- and food-deprivation. Training objectives included letting the cadets experience how teamwork and taskwork performance impacted operational effectiveness (Figure 13), and how physical and mental stress affected resilient behavior. The inherent complexity of the exercise environment challenges a team's ability to maintain shared cognition, and thereby affects team factors like communication, coordination and cooperation (Wilson, Salas, Priest, & Andrews, 2007).

Figure 13. RNoNA, exercise Dolphin 2012.

The cadet teams' operational orders were to take control of a hijacked vessel and bring it safely to a Norwegian port with the intention to stop and deter weapons smuggling.

The Dolphin exercise involved the risk of loss and injury of personnel and materiel. The exercise started with a training session to increase the probability of mission success. The training session was split into individual events designed to train adaptability, agility, creative action, speed and thoroughness to give the teams an advantage in addressing unforeseen events. The teams rotated through separate training stations based on a set schedule. SMEs trained and evaluated the teams at each station using the RNoNA assessment tool.

Once the training session was finished, the teams received updated mission orders. Opposing forces randomly, but consistently, targeted and interacted with the cadet teams. Sleep rhythms were continuously disrupted throughout the duration of the exercise.

The Dolphin exercise recreated stressors identified from a wide range of military operations, like powerlessness, ambiguity, boredom, danger, workload, uncertainty, vulnerability, fear, weather effects, sleep and food deprivation, mental and physical fatigue and time pressure. It was a complex military exercise performed in a maritime setting that placed extremely high demand for resilient behavior in the cadet teams.

C. APPARATUS

1. Simulator system

The RNoNA ship-handling simulator (NAVSIM) was used to run the simulator exercises Carey and Aden (Figure 14).

Figure 14. RNoNA NAVSIM – Bergen, Norway.

The NAVSIM is a high-fidelity simulator system with seven bridge cubicles that can represent different ships to be operated simultaneously in the same scenario. Every cubicle is equipped with all necessary navigation and communication systems and presents realistic "out the window" views of the maritime environment (Figure 15).

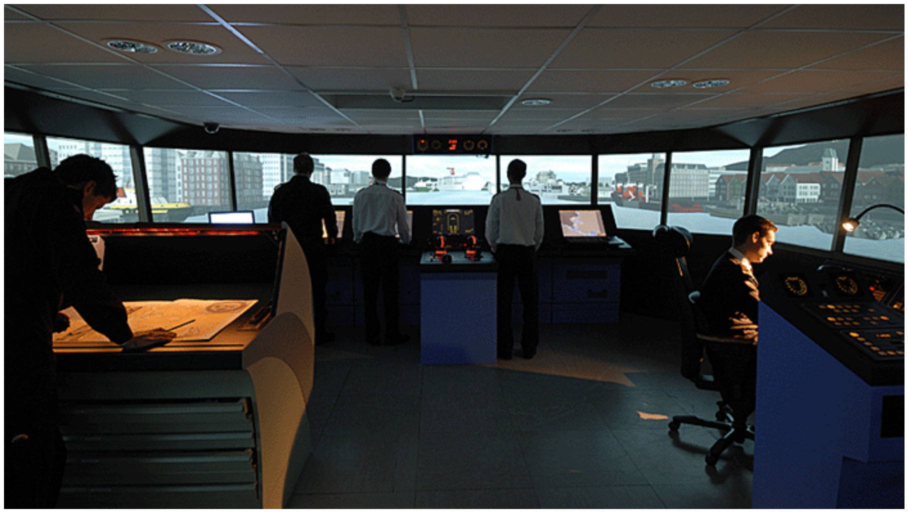

Figure 15. RNoNA NAVSIM – Bridge G.

The control room (Figure 16) is equipped with multiple instructor stations and contains network structures, server systems, computer hardware and communication system needed to run the simulation.

Figure 16. RNoNA NAVSIM – Control room.

The control room has a slave monitor system allowing the instructor to observe team behaviors via cameras inside the cubicles, including digital images replicating the "out of window" scene for each cubicle. The facility also includes an auditorium, equipped with functionality for pre-brief and debrief.

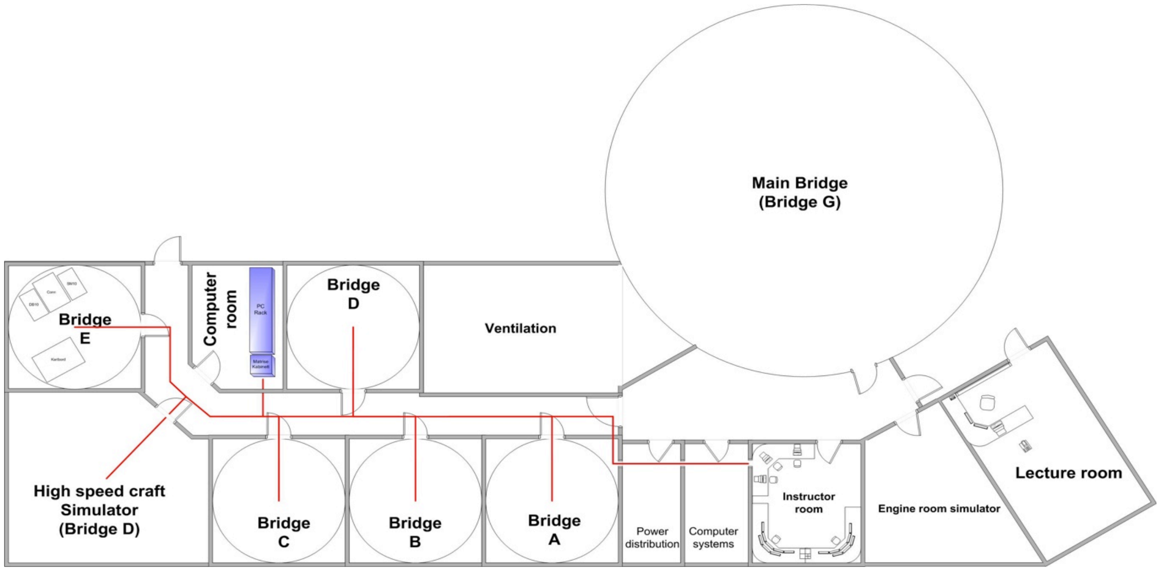

Figure 17. RNoNA NAVSIM – Facility layout.

For the purpose of cadet team training, the NAVSIM was used non-traditionally. Instead of running a typical navigation-training exercise, the simulator system facilitated scenario-based training events to train team processes on a tactical level in a complex military setting.

2. Live environment

The live environment used for exercise Dolphin was the archipelago west of Bergen, Norway (Figure 18).

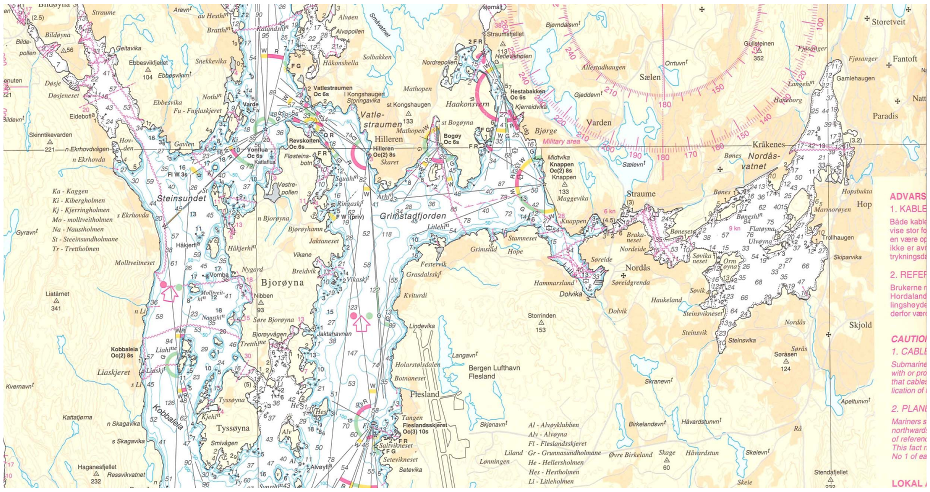

Figure 18. Extract from Sea Chart no. 21 – Bergen SW-area.

These littoral waters are at times both shallow and confined, making them difficult to navigate and imposing challenges for military operations (Figure 19).

Figure 19. HNoMS Skarv navigating typical archipelago west of Bergen (Photo: NOR Navy).

The physical environment, time of day, temperature, weather, wind, sea state, etc. contributed to the complexity and affected cadet team performance (Figure 20) in the live exercise.

Figure 20. RNoNA Cadet team during exercise Dolphin.

The live environment was used to create a complex military maritime scenario, performed in real-time, with multiple military threats, ambiguous information and technological challenges.

D. PROCEDURE

Each cadet team was assessed on the 12 teamwork and taskwork constructs 10 times, once after each simulator exercise, and eight times for specific events during the live exercise. SMEs made the assessments using the RNoNA team assessment tool either during or immediately after each event/exercise, depending on the situation.

The high level of realism in the simulator scenarios and the nature of the field exercises constrained the number of observers/SMEs. Each event had room for only a single SME to conduct the ratings. These limitations preclude measures of inter-rater reliability and associated metrics of internal and construct validity. However, other research has documented the validity of the constructs in the tool (Brehmer, 2005; Hollnagel, Woods, & Leveson, 2006; Reivich, Seligman, & McBride, 2011; Salas, Sims, & Burke, 2005).

The RNoNA would have conducted performance assessments using a similar assessment tool regardless of this study. The SMEs were not aware of the study's purpose. All SMEs used the same format for evaluating the teams' performance.

1. Simulator exercise, Carey

Each cadet team was followed by one SME. The SMEs used the RNoNA assessment tool to score the teams and did not interfere with the mission. The assessment forms were completed during or after the simulator exercise and then submitted to other RNoNA staff for logging of scores. Only one form was collected for each team for the Carey exercise.

2. Simulator exercise, Aden

Each cadet team was followed by one SME. The SMEs used the RNoNA assessment tool to score the teams and did not interfere with the mission. The assessment forms were completed during or after the simulator exercise and then submitted to other RNoNA staff for logging of scores. Only one form was collected for each team for the Aden exercise.

3. Live exercise, Dolphin

Each cadet team was followed by one or more SMEs. The SMEs used the RNoNA assessment tool to score the teams and did not interfere with the mission. The two SMEs took turns so that they were rested and observant during the observations. One SME was always present with the cadet team. The assessment forms were completed during or after each training sequence and then handed in to other RNoNA staff for logging of scores. A total of eight forms were collected for each cadet team during the Dolphin exercise.

E. MEASURES AND ANALYSES

The RNoNA assessment tool was used to rate eight cadet teams during exercise Carey, Aden and Dolphin. The exercises themselves are quasi-independent variables.

Dependent variables are the SME evaluations reported using the team performance assessments tool.

The ratings data are ordinal, not ratio. Accordingly, the appropriate statistical analyses are non-parametric. The ordinal ratings data were analyzed using the non-parametric Spearman rank-order correlation coefficient ρ (Siegel & Castellan, 1988). A Kruskal-Wallis ANOVA accounted for tied ranks to compare ranks across teams, and post-hoc Steel-Dwass analysis and pairwise Wilcoxon analysis were used to compare differences in ranks across teams.

The first set of analyses assesses the RNoNA assessment tool predictability. This is done by comparing the scores of (1) overall team performance, (2) teamwork and (3) taskwork for the eight cadet teams in each of the simulator exercise and in the live exercise.

The second set of analyses assesses how different stressors designed into the exercises impact the tool's ability to predict team ranks across exercises. This is done by comparing team ranks in (1) Carey vs. Dolphin and (2) Aden vs. Dolphin.

The third set of analyses assesses the individual metrics in the RNoNA tool. Spearman rank correlations in (1) Carey vs. Dolphin and (2) Aden vs. Dolphin are analyzed to compare the individual metric with task demands found in the exercises. The discussion includes a comparison with previous team performance research.

V. RESULTS

A. EXERCISE SCORES

1. Simulator exercise Carey

a. Raters and rating scales

The RNoNA assessment tool was used to rate eight cadet teams during exercise Carey. Four SMEs, Raters 1 through 4, rated two teams each. The scoring was entered during and/or immediately after the simulator exercise (Table 1).

Measurement			Team score Carey							
			Team 1	Team 2	Team 3	Team 4	Team 5	Team 6	Team 7	Team 8
Teamwork	1	Team Orientation	6	6	6	4	6	7	4	6
Teamwork	2	Backup Behavior	5	6	5	5	5	6	3	5
Teamwork	3	Mutual Performance Monitoring/Mutual Trust	5	7	5	4	4	4	4	4
Teamwork	4	Closed-loop Communication	6	6	6	3	3	3	5	4
Teamwork	5	Team Leadership	6	6	5	4	6	4	4	3
Teamwork	6	Shared Mental Models/Interdependence	5	5	5	3	4	6	2	4
Teamwork	7	Adaptability	6	6	5	4	5	6	6	2
Teamwork	8	Agility	5	5	5	3	5	6	6	3
Taskwork	9	Creative Action	6	6	6	3	5	5	4	2
Taskwork	10	Speed	5	6	5	4	4	4	3	7
Taskwork	11	Thoroughness	5	6	5	3	4	5	4	4
Taskwork	12	Success	6	6	5	3	6	6	6	2
		Average score: Overall	5.50	5.92	5.25	3.58	4.75	5.17	4.25	3.83
		Average score: Teamwork	5.50	5.88	5.25	3.75	4.75	5.25	4.25	3.88
		Average score: Taskwork	5.50	6.00	5.25	3.25	4.75	5.00	4.25	3.75
		Range of scores	1	2	1	2	3	4	4	5
		Rater #	1		2		3		4	

Table 1. Assessment scores, exercise Carey.

As can be seen at the bottom of Table 1, the range of scores for Raters 1 and 2 was 1 or 2; Rater 3 had a range of 3 to 4 and Rater 4 had a range of 4 to 5. Thus there were three categories of raters. Raters 1 and 2 show restrictions of scale use; Rater 3 shows moderate restriction and Rater 4 shows little restriction when scoring the cadet teams. Overall, the average range restriction in scale use in the Carey exercise was high. The low variability results in less information about the N=12 metrics, suggesting that it may be beneficial to encourage greater use of range in assessments. This encouragement may be done through better instructions to raters and changing anchors in the RNoNA assessment tool.

b. Team Performance

The boxplot in Figure 21 reveals differences in overall team performance across the cadet teams exemplified by the contrast between the low average scores for Teams 4 and 8 and the high scores for Teams 1, 2, 3 and 6.

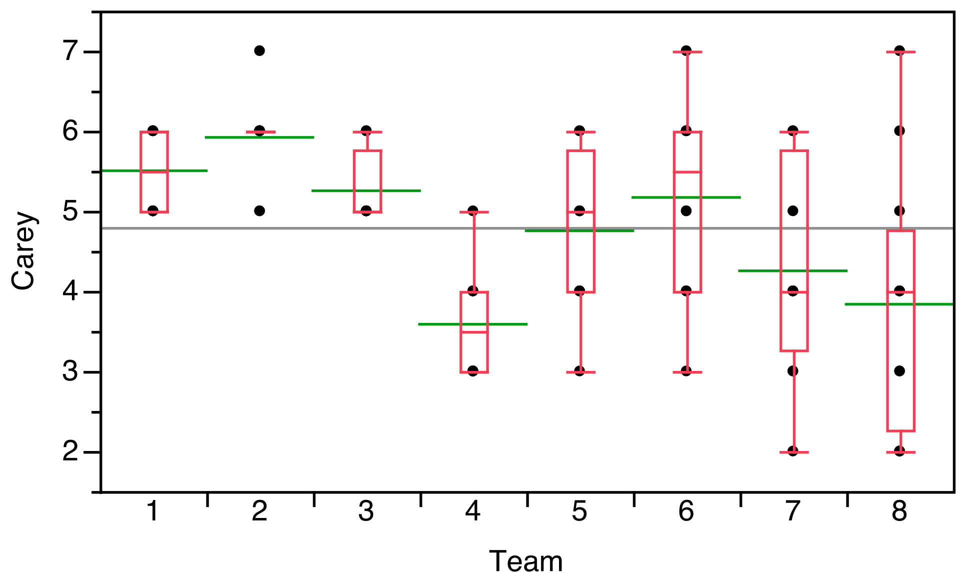

Figure 21. Boxplot Team performance, exercise Carey.

The ratings data are ordinal, not ratio. Accordingly, the appropriate statistical analyses are non-parametric. These tests require the raw ratings to be converted to ranks. The Kruskal-Wallis ANOVA accounts for tied ranks to compare ranks across teams. The test showed a significant difference across teams, Chi-square (7) =39.04, $p < .0001$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found six pairs of teams to be significantly different. The rankings for Team 4 were significantly lower than those for Teams 1, 2, 3 and 6, and the rankings for Team 2 were significantly greater than those for Teams 7 and 8.

The difference in overall ranks across teams is data used to discuss the tool's ability to predict team performance ranking across teams, Chapter V.B.1.

c. *Teamwork*

The boxplot in Figure 22 reveal differences in teamwork performance across the cadet teams exemplified by the contrast between the low average scores for Teams 4 and 8 and the high scores for Teams 1, 2, 3 and 6.

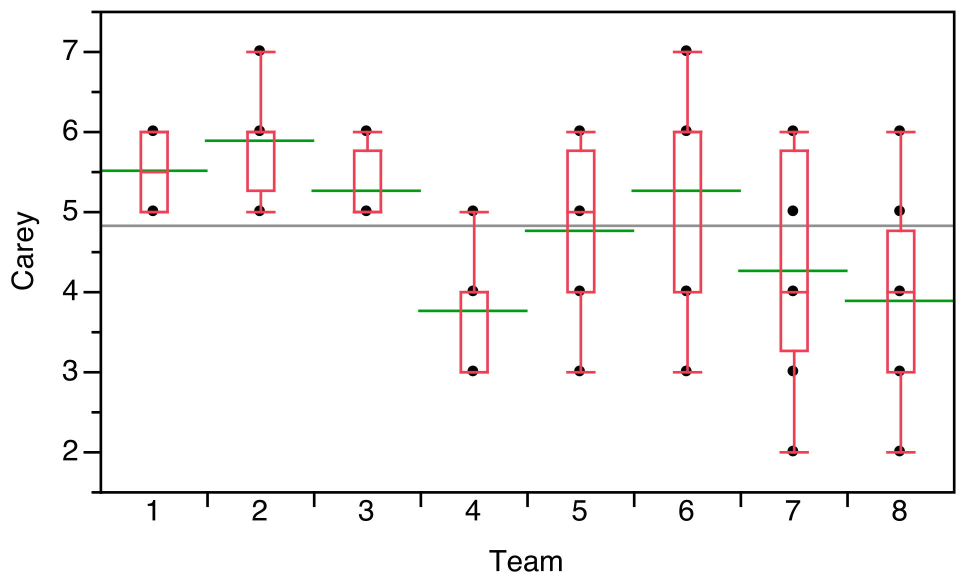

Figure 22. Boxplot Teamwork, exercise Carey.

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) =24.68, $p < .0009$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found three pairs of teams to be significantly different. The rankings for Team 4 were significantly lower than the rankings of Teams 1, 2, and 3.

The difference in teamwork average scores across teams is data used to discuss the tool's ability to predict teamwork ranking across teams, Chapter V.B.2.

d. *Taskwork*

The boxplot in Figure 23 reveal differences in taskwork performance across the cadet teams, exemplified by the contrast between the low average scores for Team 4 and the very high score for Team 2.

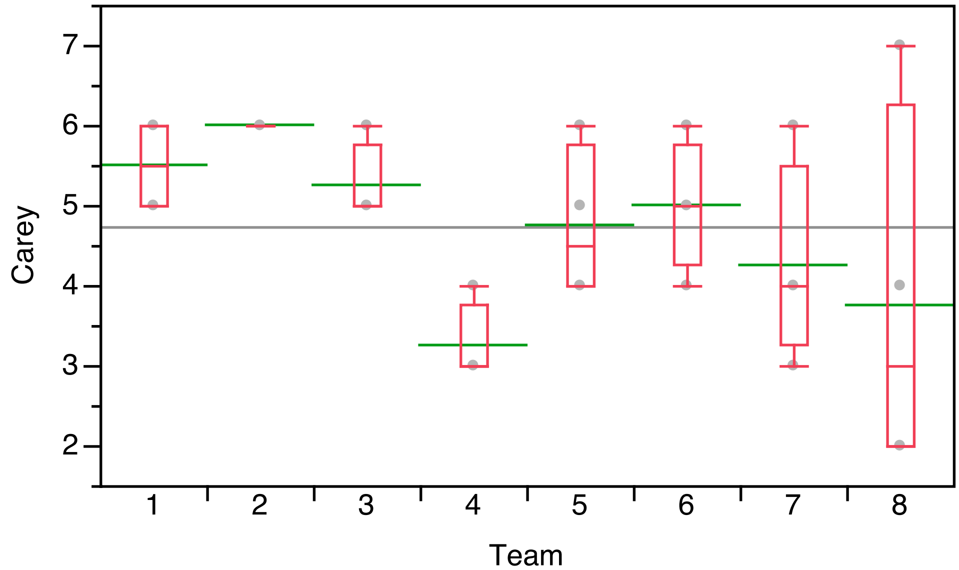

Figure 23. Boxplot Taskwork, exercise Carey

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) =14.35, $p < .0453$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found no pairs of teams to be significantly different. The less stringent pairwise Wilcoxon ($\alpha = 0.05$) comparisons, without the Bonferonni correction, found five pairs of teams to be significantly different. The Wilcoxon test showed that the rankings for Team 4 were significantly lower than ranks for Teams 1, 2, 3, 4 and 5.

Team 2 shows absolutely no variability in taskwork scores, which amplifies the earlier statement about the need to encourage greater variability in scale use.

The difference in taskwork average scores across teams is data used to discuss the tool's ability to predict teamwork scores ranking across teams, Chapter V.B.3.

2. Simulator exercise Aden

a. Raters and rating scales

The RNoNA assessment tool was used to rate eight cadet teams during exercise Aden. Four SMEs, Rater 1 through 4, rated two teams each. Each SME rated the same teams as they did in exercise Carey. The scoring was entered during and/or immediately after the simulator exercise. The results are displayed in Table 2.

		#		Team score Aden							
				Team 1	Team 2	Team 3	Team 4	Team 5	Team 6	Team 7	Team 8
Measurement	Teamwork	1	Team Orientation	5	6	5	4	7	6	5	6
		2	Backup Behavior	4	6	5	4	7	6	4	4
		3	Mutual Performance Monitoring/Mutual Trust	5	5	4	5	6	4	5	5
		4	Closed-loop Communication	5	6	4	4	6	4	4	4
		5	Team Leadership	4	6	5	4	6	5	5	4
		6	Shared Mental Models/Interdependence	5	5	4	5	6	5	4	6
		7	Adaptability	6	5	5	4	7	5	3	5
		8	Agility	4	6	5	4	5	5	5	5
	Taskwork	9	Creative Action	5	4	5	4	5	6	4	6
		10	Speed	4	5	4	4	6	5	5	5
		11	Thoroughness	4	4	4	4	6	4	6	5
		12	Success	5	5	4	3	6	4	4	5
			Average score: Overall	4.67	5.25	4.50	4.08	6.08	4.92	4.50	5.00
			Average score: Teamwork	4.75	5.63	4.63	4.25	6.25	5.00	4.38	4.88
			Average score: Taskwork	4.50	4.50	4.25	3.75	5.75	4.75	4.75	5.25
			Range of scores	2	2	1	2	2	2	3	2
			Rater #	1		2		3		4	

Table 2. Assessment scores, exercise Aden.

As can be seen at the bottom of Table 2, Raters 1, 2 and 3 show high restriction of range use, and Rater 4 shows moderate range restriction when scoring the cadet teams. As for exercise Carey, the low variability in SME scale use in Aden results in less information about the N=12 metrics.

b. Team Performance

The boxplot in Figure 24 reveals differences in overall team performance across the cadet teams, exemplified by the contrast between the low average scores for Team 4 to the very high score for Team 5.

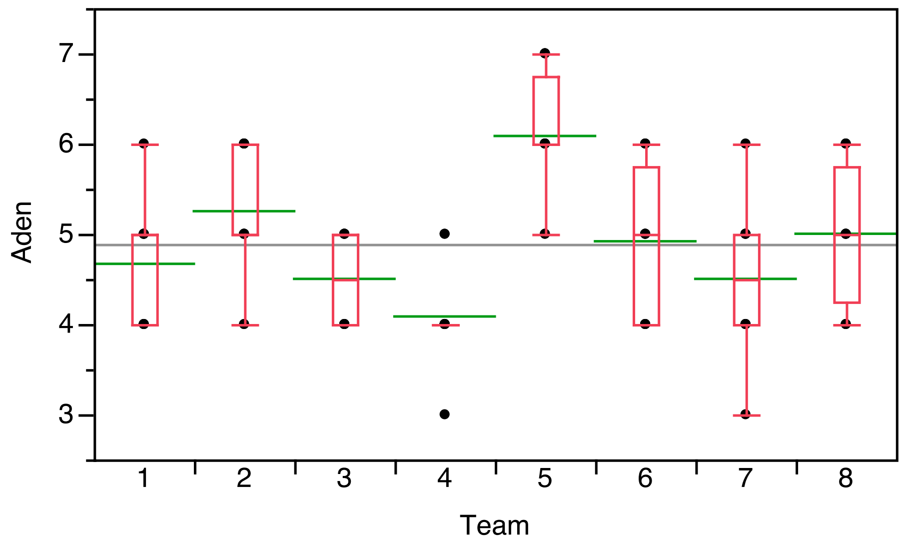

Figure 24. Boxplot Team performance, exercise Aden.

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) =37.42, $p < .0001$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found six pairs of teams to be significantly different. The rankings for Team 4 were significantly lower than the rankings for Team 2. Teams 1, 3, 4, 6 and 7 were given rankings that were significantly lower than the rankings for Team 5.

The difference in overall ranks across teams is data that can be used to discuss the tool's ability to predict performance ranking across teams, Chapter V.B.1.

c. Teamwork

The boxplot in Figure 25 reveals differences in teamwork performance across the cadet teams exemplified by the contrast between the low average scores for Teams 4 and 7 and the high scores for Teams 2, 5 and 6.

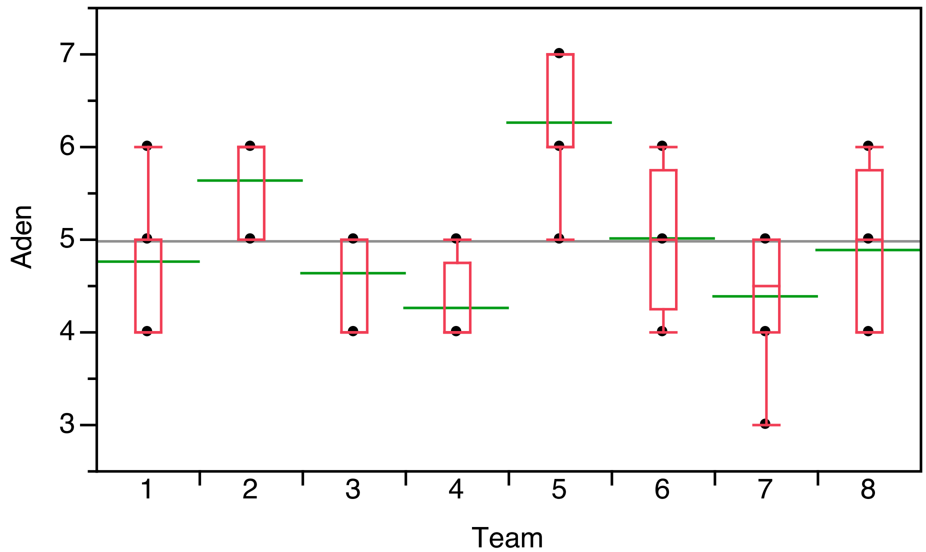

Figure 25. Boxplot Teamwork, exercise Aden.

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) = 29.48, $p < .0001$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found four pairs of teams to be significantly different. The rankings for Team 4 were significantly lower than the rankings of Team 2, and the rankings for Teams 3, 4 and 7 were significantly lower than rankings for Team 5.

The difference in teamwork average scores across teams is data used to discuss the tool's ability to predict teamwork ranking across teams, Chapter V.B.2.

d. Taskwork

The boxplot in Figure 26 reveals differences in taskwork performance across the cadet teams, exemplified by the contrast between the low average scores for Team 4 and the very high scores for Team 5.

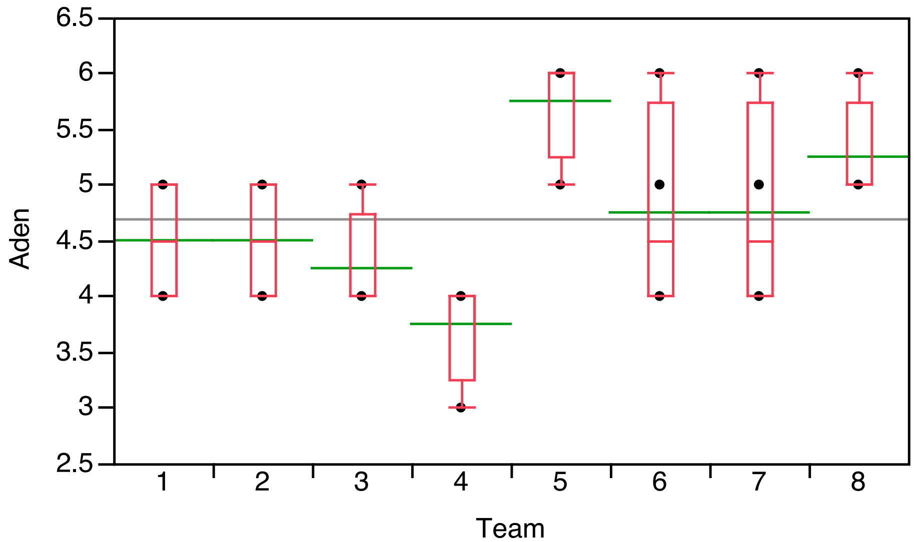

Figure 26. Boxplot Taskwork, exercise Aden.

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) =15.15, $p < .0342$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found no pairs of teams to be significantly different. The less stringent pairwise Wilcoxon ($\alpha = 0.05$) comparisons, without the Bonferonni correction, found five pairs of teams to be significantly different. The Wilcoxon test showed that the rankings for Team 4 were significantly lower than rankings for Teams 5 and 8, and the rankings for Teams 1, 2 and 3 were significantly lower than the rankings for Team 5.

The difference in taskwork average scores across teams is data used to discuss the tool's ability to predict teamwork scores ranking across teams, Chapter V.B.3.

3. Live exercise Dolphin

a. Raters and rating scales

The RNoNA assessment tool was used to rate eight cadet teams during exercise Dolphin. A total of 25 SMEs rated the teams in nine assessment events throughout the exercise. Some SMEs rated only one single team during one specific event; some rated the same team on more than one event. Other SMEs rated all eight teams performing the same event. The scoring was entered during and/or immediately after each event (Table 3).

				Team score Dolphin							
				Team 1	Team 2	Team 3	Team 4	Team 5	Team 6	Team 7	Team 8
Measurement	Teamwork	1	Team Orientation	5.78	5.33	4.67	4.16	4.78	5.13	5.56	4.89
		2	Backup Behavior	5.44	5.33	4.22	3.93	4.67	5.00	5.78	5.00
		3	Mutual Performance Monitoring/Mutual Trust	5.67	5.44	4.56	4.16	4.67	5.25	5.33	4.78
		4	Closed-loop Communication	5.22	5.44	4.11	3.82	4.22	4.38	5.11	4.44
		5	Team Leadership	5.67	5.22	4.33	3.82	4.56	5.25	5.33	5.22
		6	Shared Mental Models/Interdependence	5.11	5.67	4.33	3.71	4.44	4.75	5.22	4.56
		7	Adaptability	5.22	5.44	4.67	3.93	4.78	5.50	5.33	5.11
		8	Agility	4.78	4.89	4.67	3.60	4.22	4.75	4.67	4.44
	Taskwork	9	Creative Action	4.89	5.11	4.44	3.56	4.78	4.63	4.78	4.44
		10	Speed	5.00	5.22	5.00	3.67	4.22	5.13	4.89	4.56
		11	Thoroughness	5.11	5.11	4.44	3.89	4.56	5.13	5.11	5.00
		12	Success	5.78	5.44	4.56	3.89	4.56	5.25	5.00	5.22
			Average score: Overall	5.31	5.31	4.50	3.84	4.54	5.01	5.18	4.81
			Average score: Teamwork	5.36	5.35	4.44	3.89	4.54	5.00	5.29	4.81
			Average score: Taskwork	5.19	5.22	4.61	3.75	4.53	5.03	4.94	4.81
			Range of scores	2.3	1.9	2.0	1.8	2.3	2.4	1.8	2.1
			Rater #	1, 3, 5, 7, 9, 17, 18	1, 3, 5, 7, 10, 19, 20	1, 3, 5, 7, 11, 21, 22	1, 3, 5, 7, 12, 23, 24, 25	2, 4, 6, 8, 13, 17, 18	2, 4, 6, 8, 14, 19, 20	2, 4, 6, 8, 15, 21, 22	2, 4, 6, 8, 16, 23, 24, 25

Table 3. Assessment scores, exercise Dolphin.

As can be seen at the bottom of Table 3, the raters by average show range restrictions in assessments of the cadet teams. As in Carey and Aden, the low variability in SME scale use in Dolphin results in less information about the N=12 metrics.

b. Team Performance

The boxplot in Figure 27 reveals differences in overall team performance across cadet teams, exemplified by the contrast between the low average scores for Teams 3, 4 and 5 and the high scores for Teams 1, 2 and 7.

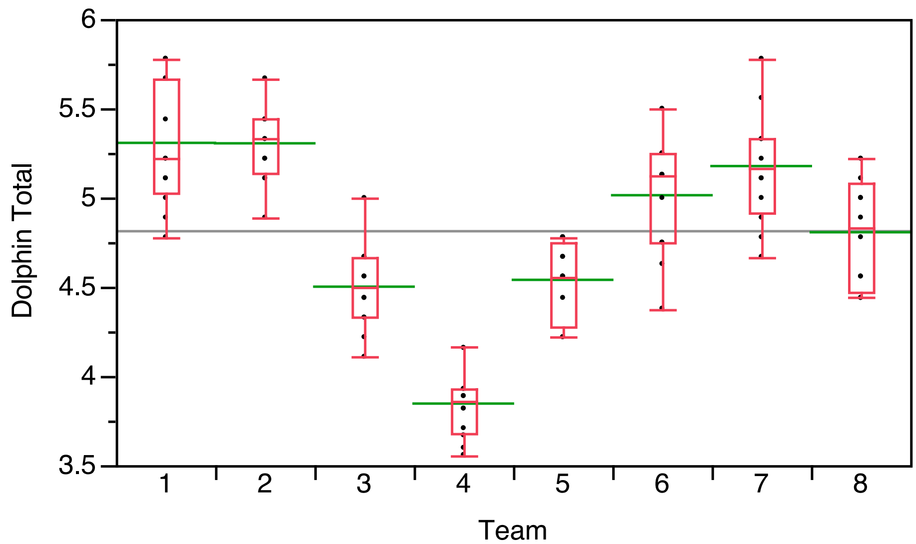

Figure 27. Boxplot Team performance, exercise Dolphin.

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) =67.64, $p < .0001$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found 15 pairs of teams to be significantly different. The ranks for Team 4 were significantly lower than the rankings for all other teams. The rankings for Team 3 were significantly lower than the rankings for Teams 1, 2, 6 and 7. The rankings for Team 5 were significantly lower than the rankings for Teams 1, 2, and 7. The rankings for Team 8 were significantly lower than the rankings for Team 2.

The difference in overall ranks across teams is data that can be used to discuss the tool's ability to predict performance ranking across teams, Chapter V.B.1.

c. Teamwork

The boxplot in Figure 28 reveals differences in teamwork performance across the cadet teams exemplified by the contrast between the low average scores for Teams 3, 4 and 5 and the high scores for Teams 1 and 2.

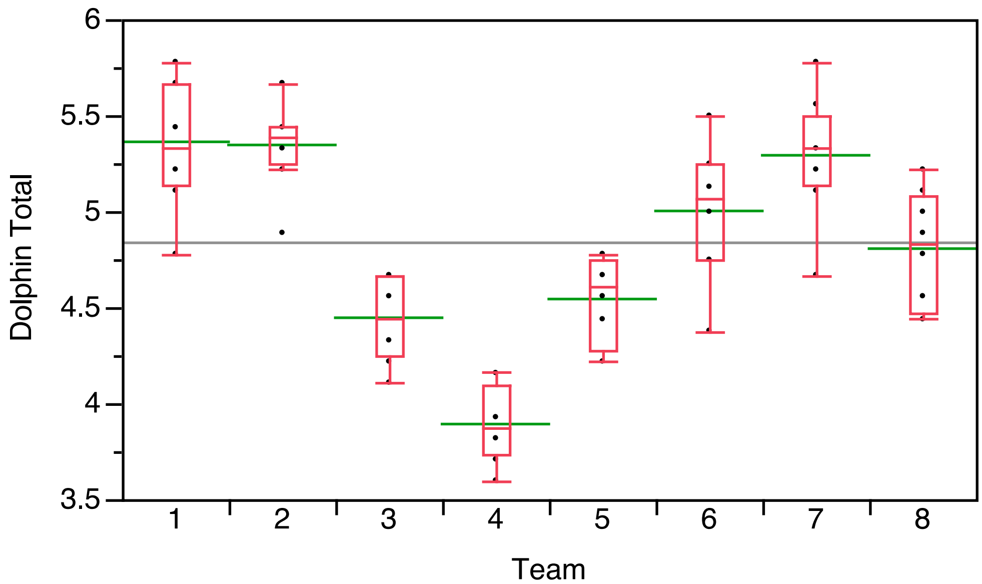

Figure 28. Boxplot teamwork, exercise Dolphin.

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) =47.26, $p < .0001$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found 12 pairs of teams to be significantly different. The rankings for Team 4 were significantly lower than the rankings for all other teams. The rankings for Team 3 were significantly lower than the rankings for Teams 1, 2 and 7. The rankings for Team 5 were significantly lower than the rankings for Teams 1 and 2.

The difference in overall ranks across teams is data that can be used to discuss the tool's ability to predict performance ranking across teams, Chapter V.B.2.

d. Taskwork

The boxplot in Figure 29 reveals differences in taskwork performance across the cadet teams exemplified by the contrast between the low average scores for Teams 3, 4 and 5 and the high scores for Teams 1 and 2.

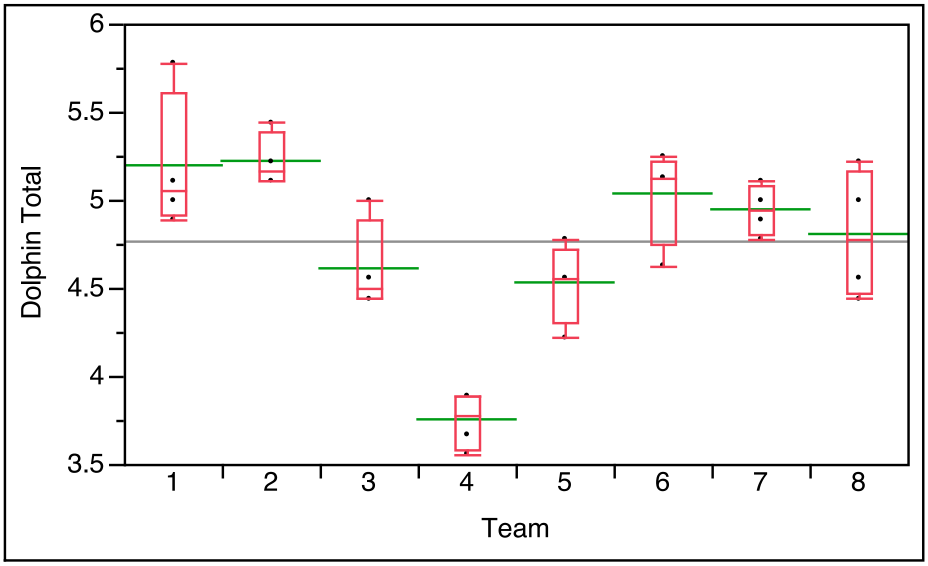

Figure 29. Boxplot taskwork assessment, exercise Dolphin.

The Kruskal-Wallis ANOVA by ranks found a significant difference across teams, Chi-square (7) =21.53, $p < .0031$. Post-hoc pairwise Steel-Dwass ($\alpha = 0.05$) comparisons found no pairs of teams to be significantly different. The less stringent pairwise Wilcoxon ($\alpha = 0.05$) comparisons, without the Bonferonni correction, found 11 pairs of teams to be significantly different. The rankings for Team 4 were significantly lower than the rankings for all other teams. The rankings for Team 5 were significantly lower than the ranks for Teams 1, 2 and 7. The rankings for Team 3 were significantly lower than the rankings for Team 2.

The difference in overall ranks across teams is data that can be used to discuss the tool's ability to predict performance ranking across teams, Chapter V.B.3.

B. RNONA ASSESSMENT TOOL PREDICTABILITY

This section presents analyses that address the research questions regarding the RNoNA team performance assessment tool. The hypotheses asked whether the tool enables the SMEs to make assessments of RNoNA cadet teams in simulator training exercises that can predict cadet team performance in a live training exercise.

Ordinal ratings data were analyzed using the non-parametric Spearman rank-order correlation coefficient ρ (Siegel & Castellan, 1988). The first set of analyses (row 1) compares the overall team performance scores for the eight teams in the two simulator exercises and the live exercise. The second set of analyses (row 2) compares their teamwork scores. The third set (row 3) compares the taskwork scores. Table 4 shows the Spearman rank-order coefficients and the corresponding P-values for (1) Simulator vs. Dolphin, (2) Carey vs. Dolphin and (3) Aden vs. Dolphin. The scores in the Simulator column are the aggregated scores for both simulator exercises, presented as averaged scores for Carey and Aden.

<table>
<tr><th rowspan="2"></th><th colspan="2">Simulator</th><th rowspan="2"></th><th colspan="2">Carey</th><th rowspan="2"></th><th colspan="2">Aden</th><th rowspan="2"></th></tr>
<tr><th>ρ</th><th>P(ρ)</th><th>ρ</th><th>P(ρ)</th><th>ρ</th><th>P(ρ)</th></tr>
<tr><td>Team Performance</td><td>.49</td><td>.12</td><td rowspan="3">Dolphin</td><td>.66</td><td>.045</td><td rowspan="3">Dolphin</td><td>.29</td><td>.27</td><td rowspan="3">Dolphin</td></tr>
<tr><td>Teamwork</td><td>.50</td><td>.115</td><td>.67</td><td>.043</td><td>.33</td><td>.24</td></tr>
<tr><td>Taskwork</td><td>.48</td><td>.13</td><td>.73</td><td>.028</td><td>.05</td><td>>.50</td></tr>
</table>

Table 4. Spearman rank-order correlation and their probabilities.

1. Team Performance

Team performance assessment includes all twelve metrics. The average overall team performance in the two simulator exercises was compared with the average overall team performance in the live exercise, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .49, with a p-value of .12 (Zar, 1972). This suggests that teams that perform well (or poorly) in the RNoNA scenario-based simulator training exercises will also perform well (poorly) in the live exercise.

These findings support hypothesis H1. When the RNoNA team performance assessment tool is used in a training simulator, the average score of the twelve selected measures can be used to predict performance in a live training exercise.

2. Teamwork

Teamwork assessment includes the eight teamwork metrics. The average teamwork performance in the two simulator exercises was compared with the average teamwork performance in the live exercise, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .50, with a p-value of .115 (Zar, 1972). This suggests that teams with a high teamwork score (or low) in the RNoNA scenario-based simulator exercise have a high probability of receiving similar scores in the live exercise.

These findings support hypothesis H2. When the RNoNA team performance assessment tool is used in a training simulator, the average score of the eight teamwork measures can be used to predict the average of the eight measures in a live training exercise.

3. Taskwork

Taskwork assessment includes the four taskwork metrics. The average taskwork performance in the two simulator exercises was compared with the average taskwork performance in the live exercise, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .48, with a p-value of .13 (Zar, 1972). This suggests that teams with a high (or low) taskwork score in the RNoNA scenario-based simulator exercise have a high probability of receiving similar scores in the live exercise.

These findings support hypothesis H3. When the RNoNA team performance assessment tool is used in a training simulator, the average score of the four selected taskwork measures can be used to predict the average of the four measures in a live training exercise.

C. STRESSORS AND RNONA TOOL PREDICTABILITY

This section addresses the research question regarding the impact of stressors in the training exercises. The hypotheses asked whether the match (or difference) between the stressors built into the two simulator exercises and the live exercise had an impact on the RNoNA assessment tool's prediction of team performance and whether there is a differential impact on measures of teamwork and taskwork.

The simulator exercise design suggests that there is higher demand for resilient behavior in cadet teams in the Carey exercise than in the Aden exercise. The live exercise design suggests an even higher demand for resilient behavior in cadet teams. The expectation is therefore to find higher correlations for Carey vs. Dolphin than for Aden vs. Dolphin.

Ordinal ratings data were analyzed using the non-parametric Spearman rank-order correlation coefficient ρ (Siegel & Castellan, 1988). The first set of analyses compares the assessments made in the Carey simulator exercise and the Dolphin live exercise. The second set of analyses compares the assessments made in the Aden simulator exercise and the Dolphin live exercise.

1. Does assessment in Carey predict assessment in Dolphin?

The Carey scenario-based simulator exercise represented uncertainty, vulnerability, ambiguity and dilemma situations in a complex maritime environment, placing a high demand on resilient behavior in the cadet teams.

The Dolphin live exercise is basically a set of events in a maritime military setting representing physical and mental challenges, uncertainty, vulnerability, time pressure, dilemma situations, boredom, ambiguity and danger, placing a very high demand on resilient behavior in the cadet teams.

a. Team Performance

Team performance assessment includes all twelve metrics. The average overall team performance in Carey was compared with the average overall team performance in Dolphin, Table 4. The Spearman rank-order correlation coefficient ($n = 8$ teams) is .66, with a p-value of .045 (Zar, 1972). This suggests that teams that perform well (or poorly) in Carey will also perform well (poorly) in Dolphin.

These findings support hypothesis H4. The overall team performance assessment predictability is high when the stressors built into the training exercise match the stressors in the live exercise.

b. Teamwork

Teamwork assessment includes the eight teamwork metrics. The average teamwork performance in Carey was compared with the average teamwork performance in Dolphin, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .67, with a p-value of .043 (Zar, 1972). This suggests that teams with a high teamwork score (or low) in the Carey have a high probability of receiving similar scores in the Dolphin.

These findings support hypothesis H5. The teamwork assessment predictability is high when the stressors built into the training exercise match the stressors in the live exercise.

c. Taskwork

Taskwork assessment includes the four taskwork metrics. The average taskwork performance in Carey was compared with the average taskwork performance in Dolphin, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .73, with a p-value of .028 (Zar, 1972). This suggests that teams with a high taskwork score (or low) in Carey have a high probability of receiving similar scores in Dolphin.

These findings support hypothesis H6. The taskwork assessment predictability is high when the stressors built into the training exercise match the stressors in the live exercise.

2. Does assessment in Aden predict assessment in Dolphin?

The Aden scenario-based simulator exercise represented time pressure, ethical and tactical dilemma situations, and extensive use of technology in an overt modern allied naval operation, placing a moderate demand on resilient behavior in the cadet teams.

In contrast, the Dolphin live exercise is basically a set of events in a maritime military setting representing physical and mental challenges, uncertainty, vulnerability, time pressure, dilemma situations, boredom, ambiguity and danger, placing a very high demand on resilient behavior in the cadet teams.

a. Team Performance

Team performance assessment includes all twelve metrics. The average overall team performance Carey was compared with the average overall team performance in Dolphin, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .29, with a p-value of .27 (Zar, 1972). This suggests that team performance assessments made in Aden have only moderate predictability for performance in Dolphin.

These findings support hypothesis H4. The team performance assessment predictability is low when the match between stressors built into the training exercise and the stressors in the live exercise are low.

b. Teamwork

Teamwork assessment includes the eight teamwork metrics. The average teamwork performance in Carey was compared with the average teamwork performance in Dolphin, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .33, with a p-value of .236 (Zar, 1972). This suggests that teamwork assessments made in Aden have only moderate predictability for performance in Dolphin.

These findings support hypothesis H5. The teamwork assessment predictability is low when the match between stressors built into the training exercise and the stressors in the live exercise are low.

c. Taskwork

Taskwork assessment includes the four taskwork metrics. The average taskwork performance in Carey was compared with the average taskwork performance in Dolphin, Table 4. The Spearman rank-order correlation coefficient (n = 8 teams) is .05, with a p-value above .50 (Zar, 1972). This suggests that taskwork assessments made in Aden have very low predictability for taskwork performance in Dolphin.

These findings support hypothesis H6. The taskwork assessment predictability is low when the match between stressors built into the training exercise and the stressors in the live exercise are low.

D. ANALYSIS OF INDIVIDUAL METRICS IN THE RNONA TOOL

In an effort to bridge the gap between academicians and practitioners, and to encourage longitudinal team performance research, Salas et al. (2005) argued that it was possible to condense the knowledge of teamwork into five core components called the "Big Five" in teamwork: (1) team leadership, (2) mutual performance monitoring, (3) backup behavior, (4) adaptability and (5) team orientation. According to Salas et al., factors that mediate between the "Big Five" are coordinating mechanisms of (a) shared mental models, (b) closed-loop communication and (c) mutual trust.

The work by Salas et al. (2005) inspired the design of the RNoNA assessment tool, and the five core components together with the three mediating factors are all metrics in the tool. The theoretical framework of the "Big Five" provided ten research propositions, some of which found support in this study. The findings are shown in Table 5 as Spearman correlation coefficients between rankings across teams in each exercise.

Measurement	Carey			Aden		
	ρ	P(ρ)		ρ	P(ρ)	
Team Orientation	.14	> .50	Dolphin	.10	> .50	Dolphin
Backup Behavior	- .15	> .50		- .22	> .50	
Mutual Performance Monitoring/Mutual Trust	.44	.16		.07	> .50	
Closed-loop Communication	.58	.08		.45	.15	
Team Leadership	.10	> .50		- .10	> .50	
Shared Mental Models/Interdependence	.18	> .50		- .15	> .50	
Adaptability	.78	.02		- .01	> .50	
Agility	.55	.09		.44	.16	
Creative Action	.60	.07		- .24	> .50	
Speed	.31	.24		- .03	> .50	
Thoroughness	.41	.18		.02	> .50	
Success	.41	.18		.46	.15	

Table 5. Ranks for individual metrics across teams for each exercise.

1. Carey and Dolphin correlations

The propositions made by Salas et al. (2005) included seven of the twelve metrics in the RNoNA tool. The correlations between Carey and Dolphin support three of the Salas et al. (2005) propositions: (1) mutual performance monitoring/mutual trust, $P(\rho) = .16$, (2) closed-loop communication, $P(\rho) = .08$ and (3) adaptability, $P(\rho) = .02$. The results indicate that effective mutual performance monitoring occurs in teams with a climate of trust and, along with closed-loop communication and adaptability, contribute to team effectiveness in exercises that place a high demand for resilient behavior in cadet teams.

Four other propositions by Salas et al. (2005) were found not to be significant, $P(\rho) > .5$: (a) team orientation, (b) backup behavior, (c) team leadership and (d) shared mental models. Further, two items in the RNoNA tool not included in the propositions, (1) agility, $P(\rho) = .09$ and (2) creative action, $P(\rho) = .07$, were found to strongly correlate between Carey and Dolphin. This result suggests that agility and creative action may be better predictors of team performance than the four constructs proposed by Salas et al. (2005) when the demand for resilient behavior in cadet teams is high. In addition, thoroughness, $P(\rho) = .18$, and success, $P(\rho) = .18$, showed positive correlations. The overall high number of strongly positive correlations (eight out of twelve) between

RNoNA cadet team performance in Carey and Dolphin suggests that task demands were matched between these two exercises.

2. Aden and Dolphin correlations

The lack of matching task demands between Aden and Dolphin led to a lack of significant correlations between metrics in the RNoNA tool. The correlations between Aden and Dolphin support only one of the Salas et al. (2005) propositions, namely closed-loop communication, $P(\rho) = .15$. This result suggests that closed-loop communication is a reliable metric for team performance. Two strongly positive correlations of metrics not included in the propositions (Salas, Sims, & Burke, 2005) were agility, $P(\rho) = .16$ and success, $P(\rho) = .15$. The low number of positive correlations (three out of twelve) between measures in Aden and Dolphin suggest that task demands were not matched between these two exercises.

3. Is there a Big Five in team performance assessment?

Closed-loop communication, agility and success correlate across teams independent of simulator exercise with Dolphin, suggesting that they are reliable predictors of team performance, regardless of the match of task demands. In addition, mutual performance monitoring/mutual trust, adaptability, creative action, speed and thoroughness correlated between Carey and Dolphin, suggesting that they are reliable predictors when task demands are matched. The "Big Five" proposals (Salas, Sims, & Burke, 2005) team orientation, backup behavior, team leadership, and shared mental models were not supported. These findings may suggest that the agility and success metrics should be included in the "Big Five," replacing other, weaker predictors of team performance.

The analysis shows significant distinctions between correlations of teamwork and taskwork metrics. The many positive correlations for taskwork metrics indicate that team performance assessments cannot rely on teamwork metrics alone, but must also include measures of taskwork metrics to reliably predict future team performance. This can be illustrated as: Team Performance assessment = Teamwork assessment + Taskwork assessment, where effective team performance is achieved through the integration of teamwork and taskwork skills (Flin, O'Connor, & Crichton, 2008).

Even though the study provided evidence of the usefulness of the RNoNA tool, the findings are based on a small sample, with only one set of observations. Additional data is therefore needed before a true test of hypotheses by Salas et al. (2005) can be run.

E. SUMMARY

Team performance, teamwork and taskwork measures for both Carey and Aden are positively correlated with measures in Dolphin. The correlations with Dolphin are stronger for Carey than Aden, largely explained by similarities in demand for resilient behavior. The difference was especially visible for measures of taskwork metrics: Carey ($\rho = .73$) and Aden ($\rho = .05$).

Individual metrics in the RNoNA tool, (1) closed-loop communication, (2) agility and (3) success, correlate strongly and positively with Dolphin across teams independent of simulator exercise, suggesting that they may be reliable predictors of team performance for any level of matching task demands. The high correlation for these metrics suggests consistent cadet team performance, which also was consistently scored by the SMEs. Other metrics in the RNoNA tool, (1) mutual performance monitoring/mutual trust, (2) adaptability, (3) creative action, (4) speed and (5) thoroughness, correlated between Carey and Dolphin, suggesting that they are reliable predictors when task demands are matched.

The findings support the central hypothesis of this research: The RNoNA tool enables SMEs to make assessments of RNoNA cadet teams in simulator training exercises to reliably predict, $P(\rho) = .12$, cadet team performance in a live training exercise. Further, the match (or difference) between the stressors built into the two simulator exercises and the live exercise impact the RNoNA tool’s prediction of team performance. There is a differential impact on measures of teamwork and taskwork depending on the match of stressors and demand for resilient behavior in cadet teams, supporting the hypothesis that scenario matters.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. DISCUSSION

A. SUMMARY OF FINDINGS

The purpose of this paper was to present and evaluate a tool designed to assess the performance of military teams participating in complex military training exercises and to investigate the effectiveness of simulator training and live training from the matching of inherent stressors.

There are significant findings regarding teamwork and taskwork measures and of the fit between the simulator exercises and the live exercise, as explained in Chapter V. The RNoNA tool was found to be effective at predicting future team performance based on assessments made in two simulator exercises. Carey was found to predict Dolphin better than Aden due to the match of stressors and task demands. Team performance assessments were found to require both teamwork and taskwork measures to be reliable.

The researchers had hoped to discuss which of the twelve items in the tool was most influential, but this was made difficult due to the range restrictions in assigned scores for the metrics. The tool was shown to be useful and can only improve when raters are encouraged to use the full range of the scales. The current restrictions found with SME scale-use has prompted follow-on efforts to improve both the tool and the assessment procedure, which is further discussed in Section C.

B. RESEARCH GOALS (FINDINGS)

The analyses shown in Table 4 revealed that team performance and teamwork in both simulator exercises were positively correlated with team performance and teamwork in the live exercise, with Carey being more strongly correlated than Aden. In contrast, the patterns of correlations for taskwork, Table 4, are distinctly different. The weak correlations between Aden and Dolphin reveal that the assessment tool may have uncovered a mismatch between the measures in the Aden exercise and the Dolphin exercise.

The apparent mismatch may be related to the degree of similarity in mission context and the demand for resilient behavior in cadet teams. Both Carey and Dolphin are strongly influenced by uncertainty and the fog of war. Perceived and real threats pose dire consequences, and the expected unfolding of events is not realized. The taskwork categories in the tool echoed these dynamics and the demand for resilient behavior in cadet teams.

In contrast, the Aden scenario adheres to a more familiar military command structure, with known orders, SOPs and tactics (Flin, O'Connor, & Crichton, 2008). These are factors that might lead to a lesser need for resilience competencies in Aden than in Carey and Dolphin. As result, the taskwork categories in the assessment tool that emphasize resilience may not have reflected the rule-following processes required by the Aden scenario.

The difference in demands for resilience between Carey and Aden and the similarity in the demand for resilience between Carey and Dolphin suggest that scenario matters. Performance in the simulator will be more like performance in the field when the simulator experience better matches the requirements of the field. An implication for practice is that the teamwork categories can be kept, but that the taskwork categories measured in a simulator exercise should be attuned to match the mission essential competencies (MECs) required on the job (Salas, Rosen, Held, & Weismuller, 2009). The RNoNA assessment tool can then be used to test whether the demands designed into a simulator exercise match the requirements for training objectives in live exercises.

C. IMPROVEMENTS TO THE RNONA TOOL

The assessment tool was shown to be effective in predicting team ranking assessed through teamwork and taskwork behavior in both training and live environments. However, inspection of the observed ratings (Tables 1, 2 and 3) reveals range restriction. The majority of raters did not use the available 7-point scale to its full extent. An example from the Dolphin exercise is how one SME rater employed a total range of 2, whereas another used a range of 4 for the same team. The tool supported the hypothesis for rank-order correlations, but the lack of range used by SMEs made it difficult to conclude if any of the teamwork and taskwork metrics had greater impact on or indication of performance than others. The limited scale use prompts a need for changes in the tool itself and is addressed in Chapter VI.C.1.

Salas et al. (2009) suggest that an assessment tool should be accompanied by an instruction to guide and encourage the SMEs to improve the consistency and quality of data gathered during observation sessions. Accordingly, Chapter VI.C.2 addresses necessary changes in the procedure.

1. Changes in the tool

a. Scale and metrics

SME feedback indicates that the initial anchoring gives a third-person view and can produce impersonal ratings. As human beings, we are more attuned to personal cues than contextual cues (Gladwell, 2000). To encourage raters to avoid range restrictions, new anchors have been applied to a revised version of the RNoNA tool (Figure 30). “Strongly disagree,” “Partially agree” and “Strongly agree” are replaced with "Unacceptable," “Below expectations,” “Meets expectations,” “Above expectations” and "Outstanding" in the revised version. The scales of 1–7 are explained in an SOP that will accompany the rater form on a separate page (Chapter VI.C.2).

The purpose of the new anchors is to give the rater a first-person view of observed behavior to make the rating process more personal, and thereby encourage full range of scale-use. Team behavior is also expected to be easier to assess when the SME can relate his or her personal experience to rate observed performance with expected performance (Brannick, Salas, & Prince, 1997).

Discoveries made in the research have led to some changes in the metrics found in RNoNA tool. The metric "mutual performance monitoring/mutual trust" has been split into two parts: (1) mutual performance monitoring and (2) mutual trust. In addition, changes in wording have been made to six of the metrics: (3) shared mental models (4) adaptability, (5) agility, (6) creative action, (7) speed and (8) thoroughness.

The changes are described in the next paragraph and reflected in the revised RNoNA tool (Figure 30).

(1) Mutual performance monitoring

Old version: The team adjusted and reinforced each other (feedback when wrong or right was accepted and implemented by team members).

New version: The team adjusted and reinforced each other (feedback when right or wrong was offered and accepted by team members).

There are two changes in mutual performance monitoring. One is the split from mutual trust, which is explained in bullet (2). The other change is found within the parentheses. The old version told the SME to assess whether feedback was implemented, which is difficult to observe. The new version lists observable actions of feedback: "offering" and "acceptance."

(2) Mutual trust

Old version: The team adjusted and reinforced each other (feedback when wrong or right was accepted and implemented by team members).

New version: The team trusted one another (information was freely shared, no reprisals for sharing, confident in each other's ability to perform tasks).

The reason for the split is to isolate the two metrics "mutual performance monitoring" and "mutual trust." The combination of the two metrics may well have confounded the SMEs, and the new version will resolve this. The split has increased the number of metrics from twelve to thirteen.

(3) Shared Mental Models

Old version: Shared Mental Models/Interdependence

New version: Shared Mental Models

There is no change in content, only to the heading. Feedback from SMEs suggested that the addition of "interdependence" was more confusing than clarifying.

(4) Adaptability

Old version: The team showed the ability to adjust strategies (dynamic co-ordination to meet shifting internal and external needs).

New version: The team showed the ability to recognize mismatches and adjust strategies to fit the situation (coordination to meet shifting internal and external needs).

The new version includes the ability to recognize mismatches in the environment and/or situation that represent the need to change. If a mismatch is not recognized, then there is no apparent need to adapt. This addition will hopefully guide the SME to assess team adaptability as a function of both monitoring and decision making.

(5) Agility

Old version: The team showed the ability to rapidly change their orientation in response to what is happening (monitor, detect and respond to resource allocation needs, *e.g., alert and ready to move*).

New version: The team showed the ability to rapidly change their orientation in response to what is happening (the team actively interacted with the environment, taking immediate actions to changes, *e.g., alert and ready to move*).

The change is found within the parentheses. The previous text seemed too static, which is contradictory to a metric that actually measures readiness and motion. The new text has more flow to it and is meant to encourage the SME to address motion and action when assessing agility.

(6) Creative Action

Old version: The team was creative in their actions (taking action to generate and exploit advantage over the situation/opponent to achieve their objectives, *e.g., cause friction to opponent, "command both sides"*).

New version: The team was proactive in their actions to generate unexpected changes (taking action to create and exploit an advantage, *e.g., shift friction from oneself to the opponent, "command both sides"*).

The changes are meant to increase focus on proactive actions. There is a difference between consciously applying creative actions to gain an advantage (active behavior) and applying creative actions because you are pushed into a corner (reactive behavior). The idea is to identify the teams who proactively employ creativity.

(7) Speed

Old version: The team was effective to complete assignments (short time, appropriate method and strategy).

New version: The team showed correct and timely coordination of planning and actions (short time, appropriate method and strategy, valuable time was not wasted, acting faster than the opponent).

The feedback from the SMEs suggested that the old version was unclear. The new version is meant to illustrate that the gap between the planning and the start of an action is considered a time delay in military operations, and that high performing teams act faster to achieve objectives.

(8) Thoroughness

Old version: The team was thorough in their assignments (solutions and actions that fit with the stated plan).

New version: The team maintained commitment and determination to challenge the situation (bounced back from pressure).

The previous version was too vague. The idea behind the new version is to bring team self-efficacy and resilient behavior to the forefront of desired behavior.

RNoNA Team Performance Assessment **Team:** **Rater:**

Teamwork

1. **Team Orientation:**
The team showed a high degree of involvement (team members monitored and paid attention to other team members, not many "free riders" in the teamwork process)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

2. **Backup Behavior:**
The team showed a high degree of backup behavior (team members helped/assisted without being asked, push of information)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

3. **Mutual Trust:**
The team trusted one another (information was freely shared, no reprisals for sharing, confident in others ability to perform tasks)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

4. **Mutual Performance Monitoring:**
The team adjusted and reinforced each other (feedback when right or wrong was offered and accepted by team members)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

5. **Closed-loop Communication:** The team exchanged information and coordinated actions through feedback and response
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

6. **Team Leadership:** The leader was effective at solving team problems (roles and responsibilities were distributed in the team)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

7. **Shared Mental Models:** The team showed the ability to create a common outlook (all team members were kept updated on the objectives, situation and priorities, both for teamwork and taskwork objectives, "what if"-processes)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

8. **Adaptability:** The team showed the ability to recognize mismatches and adjust strategies to fit the situation (coordination to meet shifting internal and external needs)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

9. **Agility:** The team showed the ability to rapidly change their orientation in response to what is happening (the team actively interacted with the environment, taking immediate actions to changes, *e.g. alert and ready to move*)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

Taskwork

10. **Creative Action:** The team was proactive in their actions to generate unexpected changes (taking action to create and exploit an advantage, *e.g. shift friction from oneself to the opponent, "command both sides"*)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

11. **Speed:** The team showed correct and timely coordination of planning and actions (short time, appropriate method and strategy, valuable time was not wasted, acting faster than the opponent)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

12. **Thoroughness:** The team maintained commitment and determination to challenge the situation (bounced back from pressure)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

13. **Success:** The team successfully accomplished the task/mission (when compared to training/mission objectives for the exercise)
Unacceptable *Below expectations* *Meets expectations* *Above expectations* *Outstanding*
1 2 3 4 5 6 7

Comments (fill in additional information on team behavior, special assignments that can explain scores, overheard quotes, etc. that can further describe your assessment):

Figure 30. Revised RNoNA Team Performance Assessment Tool, first page.

b. Behavioral markers

Following the suggestions of Salas et al. (2009), the RNoNA tool now includes a more comprehensive description of the thirteen metrics and examples of behavioral markers the SMEs should look for (Figure 31). The behavioral markers have been added to page 2 of the tool, following the same order as on the front-page. Many descriptions are from previous work by other researchers, divided into sections of Teamwork (Alberts, 2007; Boyd, 2005; Cannon-Bowers & Salas, 1998; Entin & Serfaty, 1999; Espevik, Johnsen, & Eid, 2011; Paris, Salas, & Cannon-Bowers, 2000; Salas, Sims, & Burke, 2005; Wilson, Salas, Priest, & Andrews, 2007) and Taskwork (Bandura, 1977; Bartone, 2006; Boyd, 2005; Brehmer, 2005; Hollnagel, Woods, & Leveson, 2006; Mjelde, 2013; Osinga, 2005).

This page contains description of behavioral markers for the 13 team-performance categories.

Teamwork	**Behavioral Markers**
Team orientation	Team goals are set before individual goals. Team members show motivation and involvement to cooperate. Team members are encouraged to provide alternative solutions to determine the best course of action. Team members value each other's perspectives.
Backup behavior	Team members provide and request assistance when needed. Team members assist each other in completion of tasks. The team shifts workload among team members to achieve a more balanced distribution. Team members with positive behavior and performance are recognized and acknowledged.
Mutual performance monitoring	Team members observe each other's performance while conducting their own tasks. Team members recognize and identify mistakes and lapses in other team members' actions. The team encourages mutual feedback on performance to facilitate self-correction.
Mutual trust	Team members protect the interests of others in the team. Team members accept the risk of being vulnerable to others in the team. Team members show willingness to admit mistakes and accept feedback. Team members freely exchange information with the team. Team members confront each other in a constructive manner without fear of reprisals. Team members trust each other's abilities to perform team tasks without double-checking.
Closed-loop communication	Team members acknowledge requests from others. Team members acknowledge receipt of a message. Team members clarify with the sender that the message was received and interpreted as expected.
Team leadership	Team leader provides direction. Team leader coordinates team member tasks. Team leader clarifies team member roles. Team leader synchronizes and combines individual team member contributions to achieve team goals. Team leader provides performance expectations and acceptable interaction patterns. Team leader engages in feedback sessions with the team. Team leader motivates team members. Team leader facilitates team problem solving.
Shared Mental Models -Teamwork -Taskwork	Team members share understanding of team goals and mission objectives. Team members communicate and coordinate implicitly rather than explicitly. Team members understand each other's tasks, responsibilities and roles. Team members anticipate and predict each other's needs. The team identifies changes in the environment, task, or with teammates and adjust strategies as needed. The team is able to create a common outlook. The team uses available time to provide "big picture" situation updates of the task and the environment.
Adaptability	The team can alter a course of action in response to changing conditions, internal and external. The team can adapt to meet the demands of the situation by changing teamwork processes (e.g. different communication style or restructure of team roles). Team members pick up cues that a change has occurred, assign meaning to that change, and develop a new plan to deal with the changes.
Agility	The team accepts and expects that changes are inevitable in military missions. Team members keep attention to internal and external changes and are ready to act ("staying light on their feet"). The team responds rapidly to changes in the environment.
Taskwork	**Behavioral/effectiveness Markers**
Creative Action	The team increases friction for opponent. The team decreases friction for own team and task. The team seeks innovative thinking and acts on creative solutions. The team changes anticipated rules/actions characteristic of that domain to fit team goals.
Speed	The team performs situation assessment in short time. The team quickly executes decisions. The team acts faster than the opponent. The team completes tasks without wasting valuable time. The team reallocates resources quickly within team.
Thoroughness	The team employs solutions and actions that fit with the stated plan. The team is committed to the task. Team members express belief they can influence the situation. The team monitors the outcome of actions. The team continues to challenge mission objectives even if failure threatens.
Success	The team achieves criteria set for the training/mission. The team establishes a distinctive advantage. The team's opponent surrenders. Team goals are achieved through actions and skills, not by doing the "right" things for the wrong reasons.

Figure 31. Revised RNoNA Team Performance Assessment Tool, back page.

2. Changes in the procedure

Given the restricted range of scale use, it appears that SMEs need better guidance and assistance in rating the cadet teams. One solution is to provide an SOP for how to use the RNoNA assessment tool together with SME training; another is to reduce workload and improve accuracy by increasing the number of SMEs.

a. Guidelines and SME training

According to Proctor and Van Zandt (2008), "an effective performance appraisal begins with an understanding of what it is that must be evaluated." This can be accomplished by constructing an SOP and provide SME training. At a minimum, the SOP should provide an overview of the exercise, explain the grading process, describe the reference points for the grading scales and apply it to the evaluation form. Additionally, the SOP should identify "critical behaviors and/or responsibilities" that will set a baseline for "good" team performance and define the different levels of team performance (Salas, Rosen, Burke, Nicholson, & Howse, 2007).

SME training should reinforce the SOP guidelines by identifying ways to minimize biases, incorporating practice scenarios and providing an opportunity for the SME to ask questions (Proctor & Zandt, 2008).

b. Standard operating procedure for the RNoNA tool

Figure 32 shows the standard operating procedure that will be a part of the RNoNA tool from now on. The SOP provides an introduction to the SME task and explains the grading process and standards.

RNoNA Team Performance Assessment Tool
Standard Operating Procedure

You are one of the subject matter experts (SME) who will assess team performance by observing the way the team performs tasks. The objective of the assessment is to expand our understanding of how we can identify and train team performance, and ultimately improve military team effectiveness. The RNoNA tool is designed to assess the performance of military teams participating in complex military training exercises representative of actual military operations, executed in a controllable training environment.

The tool's first page is where you score team behaviors using the metrics and scales found there. The data collected are measures of teamwork and taskwork. The assessment tool includes nine measures of teamwork and four of taskwork. The teamwork characteristics refer to interactions team members must develop and perform to function effectively as a team. The taskwork characteristics refer to resilient behaviors related to the operational activities in a complex and stressful environment.

The tool's back page contains further descriptions of the thirteen metrics, and includes examples of team behaviors you should look for. You are encouraged to combine your knowledge of the training objectives for this particular exercise with your own experience when you rate the team.
You are highly encouraged to make full use of the available scale from 1-7, and avoid restricting your scores to just a few numbers. Too little variation in scores results in poor information about team performance. The scale standards are listed below.

7	Outstanding	The best performance a team can possibly have. Performance rank in the top 5%.
6	Above expectations	Performance clearly exceeds expected performance.
5		Performance is higher than expected.
4	Meets expectations	Performance is at or above minimum standards. This level is what is expected from most teams.
3	Below expectations	Performance is lower than expected.
2		Performance is clearly much lower than expected.
1	Unacceptable	The performance is clearly unsatisfactory. It is questionable whether the team can improve to meet minimum standards.

Note: Effectiveness criteria in a training exercise are dependent on the training objectives and on the teams' expected level of proficiency at the time of assessment.

Figure 32. Revised RNoNA Team Performance Assessment Tool, SOP.

c. Workload and accuracy

To improve SME performance, an evaluation system must minimize distractions and overloading (Wilson & Corlett, 2005). Their research suggests that an evaluator should only analyze "four to five dimensions" of teamwork due to its complex and dynamic nature.

According to Wilson and Corlett (2005), "it takes a team to measure a team accurately," and they recommend one SME for every two team members. Direct observation has been the method used in this study to evaluate team performance; however, Wilson and Corlett suggest videotaping as a means to reduce workload and to provide a useful instructional tool for feedback as well. The two simulator exercises, Carey and Aden, can be videotaped, but the low luminance level in the cubicles may result in poor video resolution making it difficult to identify team members, their actions and behaviors. The live exercise cannot be videotaped, however, due to practicality reasons and military restrictions.

An evaluation system must accurately capture the evolution of team performance from start to finish (Wilson & Corlett, 2005). They claim that one evaluation at the end of an exercise is not enough and that multiple measures should be taken throughout an observation to gather a representative picture. During the exercises observed in this study, the SMEs must attend to several sources of information simultaneously. The SMEs may choose to rate the cadet teams during or shortly after the completion of single events. In relation to the signal detection theory (Proctor & Zandt, 2008), the SME's divided attention may overlook ("miss") critical team interactions. Based on this finding, multiple measures made during an exercise may improve accuracy (Wilson & Corlett, 2005) and provide more opportunities for "hits" (Proctor & Zandt, 2008).

Measurements of team performance can be quantitative, qualitative or a combination of both (Kiekel & Cooke, 2011). Quantitative measures can be performed automatically, but qualitative measures are better performed through observations or after-action reviews. The simulator exercises can include two acknowledged observers (Stangor, 2011) to be present for each team without risking reactive behavior from the cadet team members. This could improve the consistency and quality of data gathered during observation sessions and would produce data for inter-rater reliability analysis. Another way of achieving multiple measures in the simulator exercises is through quantitative and/or automatic ratings of behavior. Such data collection will not necessarily capture the fleeting knowledge that occurs during a dynamic event, and it is difficult to know if the behaviors observed are representative of actual experience within the team (Cooke, Salas, Kiekel, & Bell, 2004). This challenge will require measures of team performance that can be administered automatically and scored in real-time as the task is performed and events unfold (Kiekel & Cooke, 2011). An example of quantitative measures could be to measure the frequency and duration of voice communication, but that could leave out the qualitative measure of the content in the communication. For the live exercise, the answer is not so simple. Too many observers, even if they are acknowledged, will reduce the realism of the training event, and thereby reduce the learning experience for the cadet teams.

D. THE PLANS FOR THE RESEARCH PROGRAM (FUTURE WORK)

The study has identified needs for improvement in both the RNoNA tool itself and the implementation of the tool. This knowledge has sparked an interest for future research on team performance assessment.

1. Future use of the revised RNoNA tool

The revised tool (Figures 30, 31 and 32) will be used in a longitudinal study to assess RNoNA cadet team performance in simulator and live exercises for RNoNA cadets in 2014–2017. Even though the cadet teams change every year, their knowledge, skills and abilities are found to be homogeneous across years (Royal Norwegian Naval Academy, 2009) to allow for a pooling of data over two to four years. Four years of data collection will include approximately 240 cadets and 32 teams. This number of observations will increase statistical power. The research method will adhere to the structure described in this study.

The increased number of observations will hopefully enable more rigorous testing of the Big Five hypotheses by Salas et al. (2005) and increase the understanding of how matching task demands in simulator and live exercises can be used to improve training effects for military teams.

2. Translate the tool into Norwegian

Every SME in the study is a Norwegian military officer. English is taught as a second language in Norwegian schools starting in the second year of elementary school. The result is that most Norwegians are bilingual, and Norwegian officers use English frequently in their job assignments.

The RNoNA tool was purposefully written to accommodate the English skills anticipated to exist among Norwegian officers. The range restriction seen in the data collected in 2011 however may be a result of a language barrier and warrants attention.

There is a need to study this phenomenon in a separate study. First, the RNoNA tool will be translated into Norwegian. Second, raters of similar military backgrounds (SMEs) will be assigned to view a video of a cadet team performing one of the aforementioned exercises. Third, the SMEs will then assess cadet team performance using the RNoNA tool, randomly assigned to the English and the Norwegian versions.

The results will then be analyzed to identify and recommend future assessment procedures.

3. Introduce quantitative data (count of behavior)

In addition to the ordinal team performance ratings, the data collection can include quantitative measures. The benefit of introducing quantitative data is the ability to use parametric statistics in the analysis.

One way to collect quantitative data is to introduce a tool that counts team actions (Parker, Flin, McKinley, & Yule, 2012) related to the thirteen metrics in the RNoNA tool. The ratings can be analyzed for every observed team performance behavior, rather than a global rating assigned for the entire case. The benefit of this can be a more precise assessment of the metrics during the exercise.

4. Introduce a new live exercise

The analysis of the current design of two simulator exercises and one live exercise suggests that performance in the Carey exercise predicts performance in the Dolphin exercise better than the Aden exercise does, due to the level of matching stressors and task demands.

Future research should include a second live exercise with stressors closer to those found in the Aden exercise. This could lead to a comparative study as indicated in the bullets below:

- Carey vs. Dolphin
  - Both exercise designs are kept as before. Performance in Carey is expected to correlate strongly with performance in Dolphin.
- Aden vs. Dolphin
  - Both exercise designs are kept as before. Performance in Aden is expected not to correlate strongly with performance in Dolphin.
- Carey vs. New live exercise
  - The design for Carey is kept as before. The new exercise is closer to task demands found in Aden rather than Carey. Performance in Carey is therefore not expected to correlate strongly with performance in the new exercise.
- Aden vs. New live exercise
  - The design for Aden is kept as before. The new exercise is closer to task demands found in Aden rather than Carey. Performance in Aden is therefore expected to correlate strongly with performance in the new exercise.

E. SUMMARY

The exercises presented in this study make it possible for SMEs to monitor existing military teams operating in training environments that exhibit stressors found in real combat environments. The RNoNA tool enabled these SMEs to make assessments of RNoNA cadet teams in simulator training exercises that predicted cadet team performance in a live training exercise. Team performance, teamwork and taskwork measures for both simulator exercises are strongly and positively correlated with measures in the live exercise, $P(\rho) = .12$. Further, the match (or difference) between the stressors built into the two simulator exercises and the live exercise was shown to impact the RNoNA tool's prediction of team performance. The correlations with Dolphin are stronger for Carey than for Aden, supporting the hypothesis that scenario matters.

Team performance assessments were found to require both teamwork and taskwork measures to be reliable. Individual metrics in the RNoNA tool, (1) closed-loop communication, (2) agility and (3) success, were found to be reliable predictors of team performance regardless of matching task demands.

The study shows that the RNoNA assessment tool can (1) measure team performance in simulator training exercises and predict which team will perform better (or worse) in a subsequent live training exercise, and (2) that scenario-based simulator training can realistically represent training demands for live operations when there is a match between stressors and demand for resilient behavior in both training domains. The RNoNA tool was proven to reliably assess the performance of military teams participating in complex military training exercises representative of actual military operations, executed in a controlled training environment.

The demands for operational effectiveness and competitive advantage on the battlefield create a need for effective team training exercises and team assessment tools. This paper presents a tool that can be easily applied, within a short timeframe, and provide a meaningful assessment of a team's future performance. To meet this demand, the revised version of the RNoNA tool will be used to collect team performance data in the timeframe of 2014–2017. The collected data will be thoroughly analyzed to continue refining the RNoNA tool.

> The General, who in every specific case takes, if not the best dispositions, at least efficient dispositions, has always a prospect of attaining his objective. Marshal Von Moltke (Foch, 1918)

LIST OF REFERENCES

Adams, B. D., & Webb, R. D. (2002). Trust in Small Military Teams. *7th International Command and Control Technology Symposium.* Quebec City, Canada.

Alberts, D. S. (2007). Agility, Focus and Convergence: The future of Command and Control. *The International C2 Journal*, *1*(1), 1–30.

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review, 84*(2), 191–215.

Bartone, P. T. (2006). Resilience Under Military Operational Stress: Can Leaders Influence Hardiness? *Military Psychology, 18*, 131–148.

Boyd, J. R. (2005). *Patterns of Conflict.* (C. Richards, C. Spinney, & G. Richards, Eds.) Atlanta, Georgia, USA: Defense and the National Interest.

Brannick, M. T., Prince, A., Prince, C., & Salas, E. (1995). The Measurement of Team Process. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(3), 641–651.

Brannick, M. T., Salas, E., & Prince, C. (1997). *Team performance assessment and measurement.* New Jersey: Lawrence Erlbaum Associates.

Brehmer, B. (2005). The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Appraoch to Command and Control. *The Future of C2: 10th International Command and Control Research and Technology Symposium*. McLean, Virginia.

Cannon-Bowers, J. A., & Bowers, C. Z. (2011). Team development and functioning. In *APA Handbooks in Psychology, 1*, 597–650.

Cannon-Bowers, J. A., & Salas, E. (1998). *Making Decisions Under Stress.* Washington DC, USA: American Psychological Association.

Civil Aviation Authority. (2006). *Crew Resource Management (CRM) Training* (2 ed.). Gatwick, West Sussex, UK: The Stationary Office.

Coleridge, S. T. (1817). *Biographia Literaria*. London, UK.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring Team Knowledge. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *42*(1), 151–173.

Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in Measuring Team Cognition. In E. Salas, & S. M. Fiore, *Team cognition: Understanding the factors that drive process and performance* (pp. 83–106). Washington, DC: American Psychological Association.

Dalton, B. (2004). Creativity, Habit, and the Social Products of Creative Action: Revising Joas, Incorporating Bourdieu. *Sociological Theory*, *22*(4).

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, *95*, 32–53.

Dickinson, T. L., & McIntyre, R. M. (1997). A Conceptual Framework for Teamwork Measurement. In M. T. Brannick, E. Salas, & C. Prince, *Team Performance Assessment and Measurement* (pp. 19–43). New York: Psychology Press - Taylor & Francis Group.

Driskell, J. E., Salas, E., & Hughes, S. (2010). Collective Orientation and Team Performance: Development of an Individual Differences Measure. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(2), 316.

Entin, E. E., & Serfaty, D. (1999). Adaptive Team Coordination. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*, 312–325.

Espevik, R., Johnsen, B. H., & Eid, J. (2011). Outcomes of Shared Mental Models of Team Members in Cross Training and High-Intensity Simulations. *Journal of Cognitive Engineering and Decision Making*, *5*(4), 352–377.

EUNAVFOR. (2013). *http://eunavfor.eu/mission/*. Retrieved July 3, 2013, from EUNAVFOR: http://eunavfor.eu/mission/

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the Sharp End: A Guide to Non-Technical Skills.* Surrey: Ashgate Publishing.

Foch, F. (1918). *The Principles of War, translated by J deMorinni.* New York: The HK Fly company.

Ford, C. M. (1996). A Theory of Individual Creative Action in Multiple Social Domains. *The Academy of Management Review*, *21*(4), 1112–1142.

Foster, J. (2010). *Piracy in the Gulf of Aden and Indian Ocean.* MSCHOA. Warsash, Poland: Maritime Security Centre.

Gladwell, M. (2000). *The tipping point: how little things can make a big difference.* New York: Back Bay Books.

Hollnagel, E., Woods, D. D., & Leveson, N. (2006). *Resilience Engineering, Concepts and Precepts.* Aldershot, UK: Ashgate Publishing.

Irvine, J. W. (1988). *The Waves are Free* (1st Edition ed.). Lerwick, Shetland: Shetland Publishing.

Jacobsen, M. (1982). Looking for Literary Space: The Willing Suspension of Disbelief Re-Visited. *Research in the Teaching of English, 16*(1), 21–38.

Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and Trust in Global Virtual Teams. *Organization Science, 10*(6), 791–815.

Johnston, J. H., Serfaty, D., & Freeman, J. T. (2003). *Performance Measurement for Diagnosing and Debriefing Distributed Command and Control Teams.* NAVAIR Warfare Center, Orlando Training Systems Division. Orlando, FL: NAVAIR.

Kiekel, P. A., & Cooke, N. J. (2011). Human Factors Aspects of Team Cognition. In R. W. Proctor, & K. P. Vu, *Handbook of Human Factors in WEB design* (2 ed., pp. 107–119). Boca Raton, FL: Taylor and Francis Group.

Klein, G. A. (1993). A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. Zsambok, *Decision making in action: models and methods* (pp. 138–147). Norwood, CT: Ablex.

Kozlowski, S. W., Gully, S. M., Salas, E., & Cannon-Bowers, J. A. (1996). Team Leadership and Development: Theory, principles and guidelines for training leaders and teams. *Advances in Interdisciplinary Studies of Work Teams, 3*, 253–291.

Maddi, S. R., Matthews, M. D., Kelly, D. R., Villarreal, B., & White, M. (2012). The Role of Hardiness and Grit in Predicting Performance and Retention of USMA Cadets. *Military Psychology, 24*(1), 19–28.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2000). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*, 356–376.

Mathieu, J. E., Goodwin, G. F., Heffner, T. S., Salas, E., & Cannon-Bowers, J. A. (2000). The Influence of Shared Mental Models on Team Process and Performance. *Journal of Applied Psychology, 85*(2), 273–283.

Mathis, R. L., & Jackson, J. H. (2011). *Human Resource Management* (13th Edition ed.). Mason, OH, USA: South-Western Cengage Learning.

Matthews, M., Eid, J., Johnsen, B., & Bøe, O. (2011). A Comparison of Expert Ratings and Self-Assessments of Situation Awareness During a Combat Fatigue Course. *Military Psychology 23:2*, 125–136.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. G. (Eds.), *Team effectiveness and decision making in organizations* (pp. 149–203). San Francisco: Jossey-Bass.

Mjelde, F. V. (2013). Performance assessment of Military Team-Training for Resilience in Complex Maritime Environments. *Proceedings of the HFES 57th Annual Meeting (2013). In press.* San Diego: HFES.

Modeling and Simulation Coordination Office (M&SCO). (n.d.). *M&SCO*. Retrieved July 24, 2012, from http://www.msco.mil: http://www.msco.mil/index.html

Myran, R. (2008). Team-training at The Royal Norwegian Naval Academy. (F. Mjelde, Ed.) Bergen, Norway.

Naval Postgraduate School. (2013, Mar 5). *Human Systems Integration*. Retrieved Apr 23, 2013, from Naval Postgraduate School: http://www.nps.edu/Academics/Schools/GSOIS/Departments/OR/HSI/index.html

Orasanu, J. (1990). *Shared mental models and crew performance.* Princeton University, Cognitive Science Laboratory. Princeton, NJ: Princeton University.

Orasanu, J. (1995). Training for Aviation Decision Making: The Naturalistic Decision Making Perspective. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *39*(20), 1258–62.

Osinga, F. (2005). *Science, Strategy and War.* Delft, The Netherlands: Eburon Academic Publishers.

Oxford Dictionaries. (2013). *Oxford Dictionaries/Dictionary*. Retrieved June 28, 2013, from Oxford Dictionaries: http://oxforddictionaries.com/us/definition/american_english/agile?q=agility#agile__10

Paris, C. R., Salas, E., & Cannon-Bowers, J. A. (2000). Teamwork in multi-person systems: a review and analysis. *Ergonomics*, *43*(8), 1052–1075.

Parker, S. H., Flin, R., McKinley, A., & Yule, S. J. (2012). Using videos to determine the effect of stage of operation and intraoperative events on surgeons' intraoperative leadership. *Proceedings of the Human Factors and Ergonomics Society 56th annual meeting*, *56*, pp. 941–945.

Proctor, R. W., & Zandt, T. V. (2008). *Human factors in simple and complex systems* (2 ed.). Boca Raton, FL: CRC Press.

Reivich, K. J., Seligman, M. E., & McBride, S. (2011). Master Resilience Training in the U.S. Army. *American Psychologist*, *66*(1), 25–34.

Richards, C. (2005). *Certain to Win* (2 ed.). J. Addams & Partners.

Robertson, M. M. (2002). Macroergonomics of Training Systems Development. In H. W. Hendrick, & B. M. Leiner, *Macroergonomics - Theory, Methods, and Applications* (pp. 249–272). Boca Raton: CRC Press.

Ross, K. G., Phillips, J. K., Klein, G., & Cohn, J. (2005). *Creating expertise: A framework to guide technology-based training.* Marine Corps Systems Command, Program manager for training systems. Orlando: US DoD/MARCORSYSCOM.

Rosseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so Different after All: A Cross-Discipline View of Trust. *The Academy of Management Review*, *23*(3), 393–404.

Royal Norwegian Naval Academy. (2010). *Concept of Operations for Training Exercise Telemakos.* Bergen: The Royal Norwegian Naval Academy.

Royal Norwegian Naval Academy. (2009). *Man the Braces! - Leadership training philosophy of the Royal Norwegian Naval Academy* (2 ed.). (R. Espevik, & O. Kjellevold-Olsen, Eds.) Bergen, Norway: The Royal Norwegian Naval Academy.

Salas, E., & Fiore, S. M. (2004). *Team Cognition: understanding the factors that drive process and performance.* Washington DC: American Psychological Association.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008, June). On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society 2008 50: 540* .

Salas, E., Rosen, M. A., Burke, C., Nicholson, D., & Howse, W. R. (2007). Markers for Enhancing Team Cognition in Complex Environments: The Power of Team Performance Diagnosis. *Aviation, Space, and Environmental Medicine*, *78*(5), Section II.

Salas, E., Rosen, M. A., Held, J. D., & Weismuller, J. J. (2009). Performance Measurement in Simulation-Based Training : A Review and Best Practices. *Simulation & Gaming*, *40*(3), 328–376.

Salas, E., Sims, D. E., & Burke, S. C. (2005). Is there a "Big Five" in teamwork? *Small group research*, *36*, 555–599.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Serfaty, D., Entin, E. E., & Johnston, J. H. (1998). Team Coordination Training. In J. A. Cannon-Bowers, & E. Salas, *Making decisions under stress* (pp. 221–245). Washington, DC: American Psychological Association.

Shetlopedia. (2011, November 1). *The Norwegian MTB Flotilla in Shetland*. (Shetlopedia.com) Retrieved January 29, 2012, from The Shetland Encyclopaedia: (http://shetlopedia.com/The_Norwegian_MTB_Flotilla_in_Shetland)

Shufelt, J. J. (2006). A Vision for Future Virtual Training. *Virtual Media for Military Applications* (pp. KN2-1 - KN2-12). Neuilly-sur-Seine, France: NATO.

Siegel, S., & Castellan, N. (1988). *Nonparametric statistics for the behavioral siences* (2 ed.). Singapore: McGraw-Hill.

Stangor, C. (2011). *Research Methods for the Behavioral Sciences* (4 ed.). Belmont, CA: Wadsworth, Cengage Learning.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, Shared Mental Models, and Coordinated Performance: An Empirical Link Is Established. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41*(1), 61–71.

Sun Tzu, Sawyer, R. (1994). *Art of War*. Boulder, Colorado: Westview Press.

Testa, J. J., Aldinger, M., Wilson, K. N., & Caruana, C. J. (2006). Achieving Standardized Live-Virtual Constructive Test and Training Interactions via TENA. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).* I/ITSEC.

U. S. Office of Personnel Management. (2007). *Assessment Decision Guide.* Personnel and Slection Resource Center. Washington DC: OPM.

U.S. Army. (2005). *MANPRINT Handbook - Manpower and Personnel Integration.* Washington D.C.: Office of the Deputy Chief of Staff.

Urban, J. M., Weaver, J. L., Bowers, C. A., & Rhodenizer, L. (1996). Effects of Workload and Structure on Team Processes and Performance: Implications for Complex Team Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(2), 300–310.

USN Office of Naval Research. (2011). *Live, Virtual and Constructive (LVC) Training Fidelity.* Arlington, VA: Department of the Navy.

Wedervang, T. T. (2009). Naval Operational Leadership and Leadership Training. In RNoNA, R. Espevik, & O. Kjellevold-Olsen (Eds.), *Man the Braces!* (Vol. 2, pp. 5–8). Bergen, Norway: RNoNA.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3 ed.). Upper Saddle River, NJ: Prentice-Hall.

Wikipedia. (2013, March 14). *Fairmile D motor torpedo boat*. Retrieved July 3, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Fairmile_D_motor_torpedo_boat

Wikipedia. (2012, 12 11). *Wikipedia*. Retrieved 12 12, 2012, from Suspension_of_disbelief: http://en.wikipedia.org/wiki/Suspension_of_disbelief

Wilson, J., & Corlett, N. (2005). *Evaluation of Human Work.* Boca Raton, FL: CRC Press.

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the Heat of Battle: Taking a closer look at Shared Cognition Breakdowns Through Framework. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(243), 243–256.

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *The Leadership Quarterly, 12*, 451–483.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California