Modeling Complex Tactical Team Dynamics in Observed Submarine Operations

Tara Smallidge[1], Eric Jones[2], Jerry Lamb[1], Rachel Feyre[2],
Ronald Steed[3], and Abaigeal Caras[1]

[1] Naval Submarine Medical Research Lab, Groton, CT, USA
{tara.smallidge,jerry.lamb,abaigeal.caras}@med.navy.mil
[2] Aptima, Inc., Woburn, MA, USA
{ejones,rfeyre}@aptima.com
[3] UpScope Consulting, Mystic, CT, USA
ronaldsteed@gmail.com

Abstract. Successful submarine operations—those that accomplish the mission while maintaining security and safety—depend on numerous factors. Among the most critical elements driving success are the effectiveness of team behavior and the ability to understand when this behavior breaks down such that this degradation can be mitigated or avoided. While underway, submarine Commanders and other leaders must be attuned and alert to potential precursors that may manifest in decreased performance. This paper describes a framework used to develop performance measures to support formative assessment of team behaviors and to examine team breakdown and degradation. Results are reported from two events: an observation of an operational exercise and a study at the Naval Submarine School concerning the validity and utility of the measures. This preliminary research captured essential aspects of performance and helped define future efforts to develop better tools for assessing team behavior and understanding team breakdown in our warfighters.

Keywords: performance measures, formative assessment, team effectiveness, team breakdown, submarine.

1 Introduction

Submarine crews are operating in an era of emerging complexity in both peacetime and combat operations. Complexity brings with it new, novel, and unpredictable situations which submarine tactical teams must recognize, adapt, and respond quickly and accurately in order to complete the mission or task at hand. The most critical and common complex element is that of understanding tactical team dynamics, performance, and degradation. Maintaining effective operational team performance during prolonged stressful missions is a common challenge faced by the submarine fleet. Naval Submarine Medical Research Laboratory (NSMRL) investigated approaches to further understand the details and characteristics of tactical teams.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 189–198, 2013.

Studies over the past three years at NSRML regarding submarine tactical teams have discerned team behaviors that are aligned with the unique performance needs of submarine warfighters. This research identified and validated five sustainable tactical team practices for submarine crews: Dialogue (interaction among crewmembers), Critical Thinking (how they solve problems), Use of Bench Strength (how they build and utilize all levels of the team), Decision Making (how teams distribute authority to make such decisions), and Problem Solving Capacity (an integration of the other four practices with additional behaviors that measures the degree of tactical complexity that the team can absorb successfully). [1-3]. The five practices are necessary for effective team cohesion and dynamics, and ultimately enable a team to achieve operational resilience. The research method employed a series of workshops (called COMPASS^{SM}, described below) to develop tools that capture behaviors that are aligned with the five practices. The workshops were attended by scientists, engineers, and subject matter experts in the submarine domain. The initial focus of these workshops was to develop, for each practice, a set of Performance Indicators (PIs), i.e., observable and measurable behaviors which allow an instructor or expert to recognize whether a team or individual is performing well or poorly. These PIs were then validated during a three-day observation of a command training exercise performed by an SSGN (cruise missile submarine) crew. The PIs were developed into measurement tools that improve the Submarine Force's ability to assess tactical team behaviors, enhance training through formative feedback, and thus promote successful submarine operations.

Secondly, NSMRL, the Naval Undersea Warfare Center Division Newport, Rhode Island (NUWC) and National Aeronautics and Space Administration (NASA) are proposing to further examine team performance degradation and breakdown of tactical teams during extended missions. While team breakdown is often perceived as a sudden event with a dramatic loss of effectiveness, it may, more appropriately, be viewed as a gradual or incremental process. Therefore, this research is to conduct a set of experiments that measure team performance and determine the relationship (if any) that exists between that performance and a number of variables which may or may not contribute to team performance degradation and, eventually, breakdown. By fully capturing submarine tactical team behaviors and thoroughly understanding the details and specifics of how and when a team breaks down, the Submarine Force will be more capable of resilient action as they encounter increasingly complex combat operations.

2 Measures of Resilient Submarine Tactical Team Behavior

2.1 Overview and Development Process

Prior work performed by NSMRL has identified five team practices that are integral to promoting resilient submarine team operations [1-3]. They focus on interaction among crewmembers (Dialogue), how they work together to solve problems (Critical Thinking), how they build and utilize all levels of the team (Use of Bench Strength),

how the authority to make decisions is distributed among the team (Decision Making), and the degree of tactical complexity that the team can successfully absorb (Problem Solving Capacity). Aptima, Inc.'s COMPASS^{SM} workshop process was employed to identify observable and measurable behaviors that were aligned with these five practices. These behaviors will be used for assessment and as a provision of formative feedback. Initial data collection opportunities validated the initial products of the workshops which included Likert scales, checklists, and narrative descriptions of behavior.

Aptima, Inc.'s COMPASS^{SM} workshop process is a systematic method for identifying essential knowledge and skills and then identifying observable behaviors that provide evidence of that knowledge and those skills at varying levels of expertise. The goal of COMPASS is to develop meaningful and reliable measures that are sensitive to variability in performance and are validated by their relationship to mission outcomes [4]. It does this by combining performance and psychometric theory with extensive subject matter expert input. This input is critical to developing metrics that are firmly tied to the operational domain, are clearly expressed in operationally-relevant terms, and reflect performance at multiple levels within the tactical team (i.e., individual operators, departments, leadership). Leveraging psychometric theory ensures that the resulting measures are reliable, valid, and sensitive to changes in performance (across crewmembers and across time), and that they provide meaningful and diagnostic feedback for post-exercise debriefing. COMPASS has been applied to many complex organizations across the military. In the submarine domain, measures have been developed for routine operations such as coming to periscope depth and weapons employment, while more recent efforts have focused on more interpretive assessments of topics such as Command Team decision making [5].

At the conclusion of the first workshop, approximately 75+ PIs were identified during a thorough discussion of the observable behaviors that an expert observer would expect to see during the course of four representative submarine missions: Intelligence, Surveillance, and Reconnaissance at Periscope Depth, Anti-Submarine Warfare, Strike missions, and Routine Transit. The PIs were continually condensed and refined, and the long list was culled to a much smaller subset of "high-level" PIs that adequately covered the main categories of behavior that were represented. Throughout this process, the PIs were cross-checked with the five practices to ensure that they were aligned with the initial framework. With a reduced set of high-level PIs, the research team then turned to data collection opportunities to begin validating the products of the workshops thus far.

2.2 SSGN Command Training Exercise Observation

In spring of 2012, data were collected during a three-day observation of an SSGN (cruise missile submarine) command training exercise at the Trident Training Facility in Bangor, Washington. The research team divided into two groups of three observers

each, both with a mix of submarine domain and performance measurement experts. Each day was divided into morning and afternoon sessions (eight hours each), with a scheduled two-hour overlap to meet as a group and discuss findings. Each team of three remained on either the morning or afternoon watch for the duration of the multi-day event. During each session, a single person was assigned to a specific practice, and asked to focus on the PIs associated with that Practice. The PI assignments were balanced across the teams and across the days to maximize the amount of data collected for each. All notes were unclassified, and the observers were free to use the previously-developed data-collection sheets in any manner that they considered most useful.

From these data come a number of preliminary findings. The PIs can be developed agnostic of mission, and in fact, they were determined to be nearly agnostic of prior technical knowledge of submarine operations. Even those who were not experienced submariners could pick up on many of the identified behaviors that mark both resilient and brittle teams[1]. The SSGN crew was comprised of three different Watch Sections, i.e., intact teams which are on duty for six hours at a time. In this case, the different Watch Sections provided an excellent opportunity to witness a range of brittle and resilient team behaviors. The abilities of each section seemed to align nicely with this spectrum, and provided first-hand examples of contrasting events that will inform future development of the PIs. The research team determined that there were no "missing" practices, and in fact, some of the five practices (such as Problem Solving Capacity and Decision Making) began to lend themselves to much more richness than previously thought.

With the data that were collected, preliminary analyses were performed that examined the frequency with which each PI was observed. Figure 1 shows the total number of observations of each high-level PI summed across all observers and all days of observation. By far, the most frequently observed high-level PI was "Decision-makers use briefs to build shared understanding." This is not surprising considering that it is an easily observable act that requires someone to communicate verbally with individual or multiple crewmembers. Although all of the high-level PIs focus on observable behaviors, some were more salient than others. For example, discussion and crew engagement can be seen during an exercise, while changes to watch team configuration manifest more slowly and need to be assessed over a longer period of time. Furthermore, some PIs are not frequently observed because they rely on infrequent opportunities to observe them (i.e., the exercise may or may not achieve the necessary conditions for activating a tripwire or pre-planned response.)

[1] Submarine Operational Resilience is a team's capacity to recognize, deep within the command structure, developing danger and opportunity under ambiguous and uncertain conditions. It is a team achievement, requiring conscious and purposeful practices and behaviors. Once a danger or opportunity is recognized, resilient teams are able to adapt and respond in ways that are safe in operations, and bold in war.

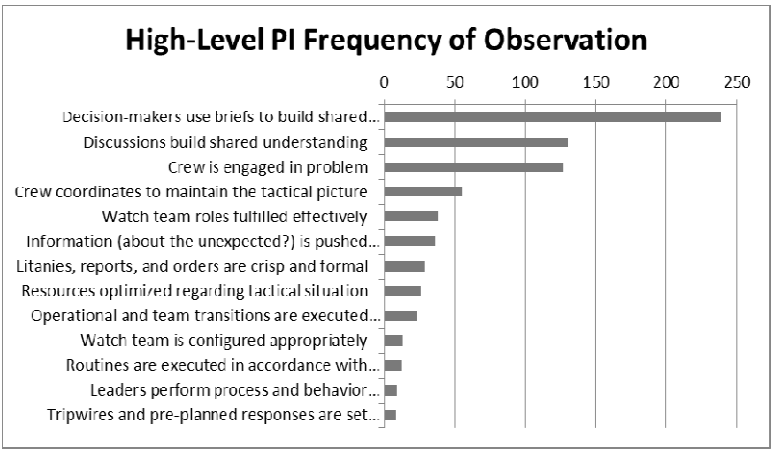

Fig. 1. The total number of observations of each high-level Performance Indicator (PI) during a three-day observation of a SSGN command training exercise

Many of the high-level PIs were associated with multiple practices, if not all of them. Figure 2 below shows the total number of times that each high-level PI was observed through the lens of each practice. Dialogue has the most number of observations, again, because it is the most salient when watching the crew over a short period of time. By contrast, very few of the observed high-level PIs were associated with Problem Solving Capacity or Decision Making. This could suggest that those practices are not yet associated with behaviors readily seen, or are not detectable over the span of a few days of observation. Regardless, data suggest these practices require additional methodological development to capture fully.

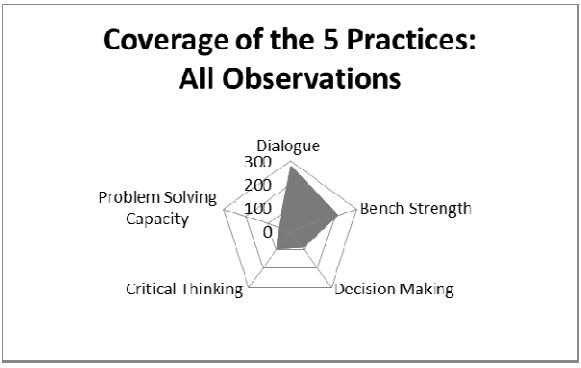

Fig. 2. The total number of times the high-level PIs that are associated with each of the five practices were observed during a SSGN command training exercise (each high-level PI was associated with one or more of the five practices)

2.3 Current State

Following the command training exercise observation, the research team continued to speak with additional experts in the submarine domain to explore ways in which the measures of tactical team behavior could be tailored to meet the needs of its future users. Through these conversations, it became clear that the comprehensive set of measures that were being developed would not be easily integrated into the existing tools and systems that are currently employed by the Fleet to train and evaluate the performance of its warfighters. One key suggestion was to summarize the set of measures so that it could be printed on paper and quickly reviewed by instructors, Commanders, and other leaders to guide their assessments and enable more formative feedback. Therefore the measures were adapted into one-page narrative descriptions of behavior, also referred to as "Team Behavior Maps," that would be much easier to use in this manner.

Each of the five practices (Dialogue, Critical Thinking, Use of Bench Strength, Decision Making, and Problem-Solving Capacity) has its own one-page Team Behavior Map. To develop each of them, the PIs and performance measures for each practice were distilled into a set of observable behaviors that were placed along a continuum of "brittle," "average," and "resilient" behavior. This continuum was divided into five distinct levels of performance that map to this range, and the behaviors were binned into one of the five categories. Observers are then able to assess where a team exists along this continuum of performance by matching observed behaviors to those in each category along the Team Behavior Map scale.

3 Team Performance Degradation and Breakdown

As mentioned earlier, maintaining effective operational team performance during prolonged stressful missions is a common challenge faced by the submarine fleet. While the behavior maps place the teams on a brittle-resilience scale, the ability to assess the change in team performance during increasing stress (complexity) is also necessary. Teams will eventually fail to accomplish their tasking, but such breakdown is preceded by other observable behavior changes. This team degradation and breakdown can be seen during observations of tactical submarine teams when the difficulty of the mission overwhelms their capacity to absorb the complexity. But while team breakdown is often perceived as a sudden event with a dramatic loss of effectiveness, this decrease in performance may in fact be a gradual or incremental process that is presently undetectable. By understanding the specific precursors prior to breakdown, the Submarine Force will be able to design technology and training that more effectively detects and mitigates its impact.

One approach to building an understanding of team degradation and breakdown is to continue gathering data while observing training and at-sea exercises. In addition, NSMRL, in collaboration with NUWCDIVNPT and NASA, propose a series of experiments that will continuously measure team performance while collecting several dependent variables (physiological and behavioral) that are identified as potential early indicators of breakdown. Assessing these variables before and during the

experiment will illustrate when and what variables contribute to team degradation. Another component of this work will research the factors that predict continued effective team performance during prolonged stress. For example, the theoretical underpinnings of this resilience; personality types that are most robust in these environments; the psychological effects of these conditions; crew selection techniques for mitigating breakdown; and, training to build team resilience. Finally, this research aims to gain firm understanding of how a team recovers effective functioning after a breakdown.

A notable factor in studying these types of teams is the flexibility by which they replace losses and augment the team during or after the stressful event; this concept of utilizing reserve capacity is reflected in the practice "Use of Bench Strength," discussed earlier. Although the mission can be accomplished with added support, the original, remaining team must continue functioning and ultimately recover from the breakdown. While stressor types vary between the missions and organizations, varying stress levels should induce team performance degradation or breakdown regardless of the specific mission or organization that is involved. Studying this process requires three critical elements; preliminary data collection, a set of metrics that illuminate precursors to the breakdown, and an operationally realistic environment where the team can perform for days or weeks at a time. Figure 3 below shows two studies that are needed to initially address the team questions that are posed.

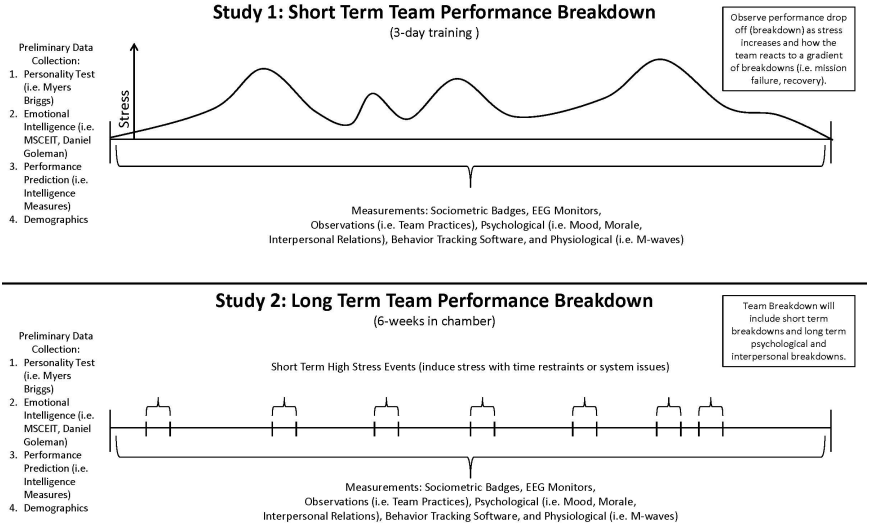

Fig. 3. A sample measure that assesses how a team builds an operational picture and how this assessment may correlate to the proactive transfer of information

In both studies the researchers will collect preliminary data on personality traits (e.g., Myers Briggs); emotional intelligence (e.g., MSCEIT, Daniel Goleman); performance predictors (e.g., Intelligence Measures); and demographics. The researchers plan to correlate these data with team performance degradation and breakdown as

measured through various tools: observer-based measurements (e.g., the measures of Resilient Submarine Tactical Team Behavior), physiological measures (e.g., M-waves), psychological measures (e.g., Mood, Morale, Interpersonal Relations), Electroencephalogram (EEG) monitors, Sociometric Badges (described below), and behavior tracking software. Some of these tools are being used in ongoing research efforts which are examining the behavioral, physiological, and neurological factors that enable a more nuanced assessment of team performance than is currently available.

Several of these ongoing studies provide both insight and techniques for the experimentation that is proposed. For example, the measures of resilient submarine tactical team behavior discussed above potentially provide a sensitive mechanism through which to assess incremental team degradation. Another study with direct relevance is a Defense Advanced Research Projects Agency (DARPA) program with the University of California, Los Angeles (UCLA), Submarine Learning Center (SLC), and NSMRL to wirelessly monitor EEG signals from six submarine team members performing a navigational task. The signals are then time-correlated and assessed using a measure called neurosynchonicity (i.e., a measure of how engaged and collaborative a team is at any given time). This measure has been shown to correlate with scenario events and has been used to identify differences between the ad hoc and mature teams [6].

Physiologically, NSMRL has been conducting at-sea tests of circadian rhythm and lighting by collecting salivary, melatonin, cortisol and alpha-amylase, which are important biomarkers of stress. Specific performance methodologies that have been used include the Multi-Attribute Task Battery, which incorporates tasks that are analogous to activities that aircraft crew members perform in flight, and the Psychomotor Vigilance Task which is a sustained-attention, reaction-timed task that measures the speed at which subjects respond to a visual stimulus.

In addition, NSMRL has the responsibility for testing prospective submariners for suitability for submarine service. The test used, SUBSCREEN (an NSMRL-developed instrument) has been shown to predict losses during the first enlistment. The test is currently being reanalyzed to better predict retention losses. If it is effective, it could possibly become a component of a selection process for effective teams, in addition to standardized test like the Myers-Brigg. Psychological measures will also be included in this study for use in measuring Mood, Morale, and Interpersonal Relations over time. The measures being evaluated during the prolonged stress-induced task will give an indication of how/when a team potentially breakdown in order to provide guidance to our warfighters.

Manually assessing team interactions can, at best, be resource intensive, and for certain team sizes and lengths of time, intractable. Automating such assessments reduces the resources required to do so by decreasing both the number of observers that are required and the time spent manually coding interactions. By removing these constraints, it also increases the amount of data that can be gained because now a group of practically any size can be instrumented to collect data over any length of time. Sociometric Badges (produced by Sociometric Solutions, Inc. [SSI] and the Massachusetts Institute of Technology [MIT]) are small, unobtrusive pieces of hardware that are worn around a person's neck and employ multiple sensors that collect various

types of data as teams of people interact in complex mission environments. The types of data that are recorded include artifacts of speech, face-to-face interactions, body movements, and the proximity of people with respect to one another. In prior experiments the data were analyzed to assess the ability of the Sociometric Badges to automatically and reliably detect behaviors that correlate to team performance [7].

Another tool to assist in behavioral observation and coding is the NASA Behavior Tracking Software developed by Horizon Performance through a Small Business Innovative Research (SBIR) grant. This software, originally designed for the Department of Defense and modified for monitoring astronauts, allows users to track and code human behavior in real time or post hoc using video. The software allows users to timestamp, tag, and rate behaviors as they occur and are observed. These behaviors can then be linked to other data sources. The software can also be used to generate near real-time reports for use by observers or the individuals being observed.

4 Discussion and Next Steps

Team behaviors are crucial to successful submarine operations. If validated, the team practice behavior maps would allow for the accurate evaluation of tactical teams' behaviors, the precious identification of problem areas, and the targeted delivery of formative feedback to communicate, precisely, how a team can improve its resilience. Validation will require disparate observers who are using the tools to record similar assessments (reliability), descriptions that sufficiently capture the range of behaviors that define the team's performance (sensitivity), as well as accurate correlations between Team Behavior Map rankings and team performance (validity). If the Behavioral Practices are diagnostic of resilient submarine team behavior, then higher scores should correlate with successful performance. The Team Behavior Maps will be further validated by comparing the assessments of the observers who are using them (i.e., members of the research team, Navy personnel and/or contractors). If the worksheets are a reliable assessment tool, the individuals' ratings should be consistent. The Team Behavior Maps will be instrumental in future research efforts, such as the proposed experimentation to examine team breakdown and degradation.

Team breakdown and degradation is often perceived as a sudden event with a dramatic loss of effectiveness, however the breakdown of a team may in fact be a gradual, observable process. As is typically seen in the submarine domain, effective operational team performance is difficult to maintain, especially during prolonged stressful missions. By understanding the proposed theoretical underpinnings of the effect these missions have on teams—including the personality types that are most robust in these environments, the psychological effects of these conditions, crew selection techniques for mitigating breakdown, and training to build team resilience—the Submarine Force will be able to design technology and training that more effectively detects and mitigates their impact.

Overall, this program of research will assist the Submarine Force as they encounter the increasing complexity of combat operations, serve to improve the individual and team selection and screening process, and evaluate intact teams for potential vulnerabilities such that they can be trained to be more resilient when faced with these challenges.

Acknowledgements. This work was completed under a contract with the Naval Submarine Medical Research Laboratory (NSMRL). Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSMRL. We gratefully acknowledge the support and assistance of the crew that participated in the study, the Trident Training Facility (Bangor, WA), the Naval Submarine School (Groton, CT), and the Naval Undersea Warfare Center Division Newport.

References

1. Steed, R., Lamb, C., Severinghaus, R.: Submarine Operational Resilience: Team Performance Behaviors during Underway Tactical Operations. Paper presented at the National Defense Industrial Organization Joint Undersea Warfare Technology Conference, Groton, CT (September 2011)
2. Lamb, C., Steed, R., Severinghaus, R.: Improving Mission Effectiveness by Increasing Team Problem-Solving Capacity. Paper presented at the National Defense Industrial Organization Joint Undersea Warfare Technology Conference, Groton, CT (September 2012)
3. Steed, R., Lamb, C., Severinghaus, R., Lamb, J., Fields, E., Caras, A.: Team Performance Behaviors during Underway Tactical Operations (NSMRL//TR–2012-0001). Naval Submarine Medical Research Laboratory, Groton (2012)
4. MacMillan, J., Entin, E.B., Morley, R.M., Bennett Jr., W.R.J.: Measuring team performance and complex and dynamic military environments: The SPOTLITE method. Military Psychology (in press)
5. Jones, E., Steed, R., Diedrich, F., Armbruster, R., Jackson, C.: Performance-based metrics for evaluating submarine command team decision-making. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) FAC 2011, HCII 2011. LNCS, vol. 6780, pp. 308–317. Springer, Heidelberg (2011)
6. Stevens, R., Galloway, T., Wang, P., Berka, C., Tan, V., Wohlgemuth, T., Lamb, J., Buckles, R.: Modeling the NeurodynamicComplexity of Submarine Navigation Teams. Computational & Mathematical Organization Theory 18, 1–24 (2012)
7. Jones, E., Lansey, J., Kern, D., Whetzel, J., Diedrich, F., Haass, M.: Automated assessment of submarine team performance using Sociometric Badge technology. Paper presented at the National Defense Industrial Organization Joint Undersea Warfare Technology Conference, Groton, CT (September 2012)