2019-03

DETERMINANTS OF SHIP-HANDLING PROFICIENCYEVIDENCE FROM FIRST-TOUR OFFICERS OF THE DECK (OODS)

Dearth, Robert W., Jr.

Monterey, CA; Naval Postgraduate School

http://hdl.handle.net/10945/62245

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

DETERMINANTS OF SHIP-HANDLING PROFICIENCY—EVIDENCE FROM FIRST-TOUR OFFICERS OF THE DECK (OODS)

by

Robert W. Dearth Jr.

March 2019

Thesis Advisor:	Jesse Cunha
Co-Advisor:	Charles P. Good

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2019	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE DETERMINANTS OF SHIP-HANDLING PROFICIENCY—EVIDENCE FROM FIRST-TOUR OFFICERS OF THE DECK (OODS)	**5. FUNDING NUMBERS** NPS-19-N082-B
6. AUTHOR(S) Robert W. Dearth Jr.	
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) Naval Surface Forces, San Diego, CA 92155	**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.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

This paper examines the determinants of a mariner's ship-handling proficiency. The understanding of this theoretical relationship affects the approach that Surface Warfare Officers School (SWOS) utilizes in preparing officers of the deck (OOD) for assignment on surface combatants. At the ship level, budget reductions may further complicate the process of developing and maintaining proficient mariners by removing opportunities available to develop experience and currency-related skills. There are no data-focused studies available to explain the mechanisms through which mariners' skills are developed or maintained in the Navy. We examined the optimal metrics for measuring OOD performance through a proficiency-prediction model, using cross-sectional data from 164 first-tour OODs who were tested across 61 ships. We find that mariners' skills, knowledge, and experience on the bridge are correlates of proficiency. This finding suggests that policies designed to encourage additional opportunities for deliberate practice mitigates skill degradation in the short term and leads to mastery of maritime skills in the long term. Policymakers should leverage simulator training to increase the proficiency of OODs through experience and currency-building evolutions. Simulators provide a substantial return on investment and offer unlimited combinations of experience-building scenarios that are difficult to duplicate in real-world practice with limited resources.

14. SUBJECT TERMS United States Navy, officer of the deck, OOD, SWO, competency, proficiency, currency, training	**15. NUMBER OF PAGES** 113
	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

THIS PAGE INTENTIONALLY LEFT BLANK

DETERMINANTS OF SHIP-HANDLING PROFICIENCY—EVIDENCE FROM FIRST-TOUR OFFICERS OF THE DECK (OODS)

Robert W. Dearth Jr.
Lieutenant, United States Navy
BSAST, Thomas A Edison State College, 2012

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
March 2019

Approved by: Jesse Cunha
Advisor

Charles P. Good
Co-Advisor

Latika Hartmann
Academic Associate, Graduate School of Business and Public Policy

ABSTRACT

This paper examines the determinants of a mariner's ship-handling proficiency. The understanding of this theoretical relationship affects the approach that Surface Warfare Officers School (SWOS) utilizes in preparing officers of the deck (OOD) for assignment on surface combatants. At the ship level, budget reductions may further complicate the process of developing and maintaining proficient mariners by removing opportunities available to develop experience and currency-related skills. There are no data-focused studies available to explain the mechanisms through which mariners' skills are developed or maintained in the Navy. We examined the optimal metrics for measuring OOD performance through a proficiency-prediction model, using cross-sectional data from 164 first-tour OODs who were tested across 61 ships. We find that mariners' skills, knowledge, and experience on the bridge are correlates of proficiency. This finding suggests that policies designed to encourage additional opportunities for deliberate practice mitigates skill degradation in the short term and leads to mastery of maritime skills in the long term. Policymakers should leverage simulator training to increase the proficiency of OODs through experience and currency-building evolutions. Simulators provide a substantial return on investment and offer unlimited combinations of experience-building scenarios that are difficult to duplicate in real-world practice with limited resources.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACDL	Army Commercial Driver's License
ADOC	Advanced Division Officer Course
AIS	Automatic Identification System
ARPA	Automatic Radar Plotting Aid
ATP	Commercial and Airline Transport License
BDOC	Basic Division Officer Course
BRM	Bridge Resource Management
CAC	Common Access Card
CDL	Commercial Driver's License
CG	Guided Missile Cruiser
CNAF	Commander, Naval Air Forces
CO	Commanding Officer
CONN	Conning Officer
DDG	Guided Missile Destroyer
DoD	Department of Defense
DoD ID	Department of Defense Identification Number
DoN	Department of the Navy
FAA	Federal Aviation Administration
FLTMPS	Fleet Management and Planning System
FMCSA	Federal Motor Carrier Safety Administration
GAO	Government Accountability Office
JOOD	Junior Officer of the Deck
LDO	Limited Duty Officer
LHD	Landing Helicopter Dock
LPD	Landing Platform Dock
LSD	Landing Ship Dock
LSO	Landing Signal Officer
LTCSS	Large Truck Crash Causation Study
MCM	Mine Countermeasure ship
NSS	Navigation, Seamanship, and Ship Handling Test

NSST	Navigation, Seamanship, and Ship Handling Training
O-FRP	Optimized Fleet Response Plan
OCS	Officer Candidate School
ODS	Officer Development School
OLS	Ordinary Least Squares
OOD	Officer of the Deck
PQS	Personnel Qualification Standard
RADM	Relational Administrative Data Management
ROR	Rules of the Road
ROTC	Naval Reserve Officer Training Corps
SWO	Surface Warfare Officer
SWOS	Surface Warfare Officers School
TER	Transfer Effectiveness Ratio
TOT	Transfer of Training
TSS	Traffic Separation Scheme
USNA	United States Naval Academy
VMS	Voyage Management System
YP	Yard Patrol Craft

ACKNOWLEDGMENTS

I would like to thank the staff at Surface Warfare Officers School for their support throughout this project. I would also like to thank my advisors, Dr. Cunha and CAPT Good, for their brilliant insight, guidance, and patience. Lastly, I would not have accomplished this much in my career if it were not for the love and support from my wife, Cassie.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

In the name of efficiency, policymakers continuously seek avenues that reduce costs for their organizations. When policymakers take these reductions too far, effectiveness suffers, and the results can be catastrophic. Several major ship collisions in 2017 catalyzed a deep reflection on the constraints the Surface Navy faces. Since the 1990s, the volume of commercial maritime traffic increased by 400 percent (Department of the Navy, 2017). During the same period, officers of the deck (OODs) lost half of their opportunities to hone their maritime skills due to Surface Warfare Officer (SWO) over accession and overall reductions in Fleet steaming time. Proposed solutions to improve mariner proficiency vary significantly in cost, from the purchase of dedicated Yard Patrol craft (YP) to gain hands-on experience at sea to increased use of simulators ashore. Identifying how any solution impacts the determinants of proficiency is critical as we determine the most effective allocation of training resources for a new generation of OODs.

We generally refer to the evolution of the OOD proficiency assessment as three distinct periods throughout this paper. The reader should expect key-signal phrases to differentiate between whether the authors are referring to the past, present, or future assessments. When the authors refer to "164 OODs," "pilot study," or "original OOD proficiency assessment" the reader should interpret these context clues as the original data set that Surface Warfare Officers School (SWOS) collected in 2018 in response to several ship collisions in the Western Pacific. As the Surface Community implements these lessons learned, SWOS is continuing to assess OOD proficiency in a similar manner to this original study. The authors refer to this transitional period in the present tense because we are actively supporting this data collection effort by designing products and consulting with SWOS to resolve data collection shortfalls due to the design of the original study. Finally, the reader should look for references such as "future analysis," "future studies," or "moving forward" and interpret these contextual clues to understand that the authors are referring to the long-term OOD assessment continuum currently being resourced for full implementation in 2021.

This paper includes several chapters where the authors analyze the original study and provide recommendations for both short- and long-term solutions. We designed these recommendations to align each part of the original OOD assessment with a big-picture strategy aimed at the long-term assessment process. This strategy supports the health of the Surface Warfare community by institutionalizing a robust data collection and assessment protocol grounded in an academic approach for consistency and repeatability.

This paper analyzes cross-sectional data collected by senior SWOS instructors from 164 randomly selected OODs by Surface Force leadership in fleet concentration areas. The randomized OODs participated in a two-hour exam comprised of a ship-handling simulator scenario and two written exams. SWOS developed a moderate-density Traffic Separation Scheme (TSS) transit simulation that they designed to assess a range of mariners' skills. SWOS selected several post–sea-duty lieutenants that served as the conning officer (CONN), junior officer of the deck (JOOD), and Combat Information Center watch standers. SWOS designed this format to ensure a consistent atmosphere that could focus any variation in performance on the assessed OOD. The written portion included rules of the road (ROR) and navigation, seamanship, and ship-handling (NSS) exams.

A post-command Commander or Captain served as the Commanding Officer (CO) of the simulated ship and the OOD's evaluator. This senior assessor received contact reports and equipment malfunction reports from the OOD to simulate a real shipboard environment. The OOD completed the 45-minute simulation while the evaluator followed a pre-planned script with a checklist to score each decision made while navigating the simulation's decision points. The evaluator assessed each OOD's proficiency on a three-point scale as "significant problems," "completed with concerns," or "completed with no concerns." The senior assessor also evaluated each decision the OOD made under one of several categories, which included management of the bridge team (BTM), managing bridge resources (BRM), leadership, application of the ROR, and performance under stress. The senior assessor evaluated these categories as either "unsatisfactory, requires improvement, meets standards, or exceeds standards."

SWOS collected this data for two primary reasons. First, mandates from Congress, the Secretary of the Navy, the Chief of Naval Operations, and United States Fleet Forces required SWOS to develop a standard to evaluate the proficiency of the fleet. This pilot-study assessment was a first step towards meeting those requirements. Second, SWOS wanted a benchmark to evaluate future program performance as policymakers allocate more resources to the OOD training pipeline. The data collected thus far serves as a baseline for additional data collection opportunities so SWOS can measure the effectiveness of fleet training efforts across future cohorts. The challenges SWOS faced with coordinating these checks across seven homeports demonstrated that the path ahead necessitates a localized-testing approach in Newport, RI.

SWOS intends to conduct future proficiency assessments centrally at Newport, RI, until OOD Phase I and II courses come online in Norfolk, VA, and San Diego, CA. This approach allows SWOS to expend fewer man-hours to assess OODs by ensuring that all assessments occur "in house" at SWOS learning sites. This approach requires incoming cohorts to arrive with the requisite experience levels documented to satisfy training requirements for career progression.

OODs document these experiences on newly developed and individually issued Mariner Skills Logbooks.[1] Modeled after the aviation community's pilot logs, this record covers individual ship-handling experiences, special evolutions, simulator training, Commanding Officer quarterly endorsements, and summarizes the accumulation of ship-handling experiences. Moving forward, OODs will complete proficiency assessments and future studies will examine the correlates to proficiency by analyzing the experiences documented in these Mariner Skills Logbooks. Eventually, the data collected in these logs will offer additional value to the Surface community.

Navy Military Personnel Command will receive an end-of-tour compilation, which summarizes the experience OODs acquired during their assignment. This information will provide detailers with experience information that they can leverage to place officers into commands that can best utilize the officer's prior ship-handling experience. Additionally, SWOS can use the logbooks to tailor advanced training opportunities to the individual level, which ensures officers earn the necessary experience to remain successful as a SWO. This is a step forward in the detailing and training process that services ships and their officers today. The current model focuses on projected rotation dates that align a new officer and the gaining

---

[1] An abbreviated version of the Surface Warfare Mariner Skills Logbook is included in Appendix H.

command while not accounting for the ship-handling experience needs of the ship and the individual officers because the community as not collected the necessary data until now.

Promotion boards will consider the results of proficiency assessments but these assessments could affect OODs at certain milestones called “Go, no-Go” assessments.[2] Students must pass certain practical exams that demonstrate a rudimentary understanding of technical ship-handling skills. If an OOD fails one of these assessments and SWOS leadership determines the student is beyond the scope of remediation, then the officer will not be permitted to progress through the SWO training pipeline. This outcome results in the officer re-designating into another career field or separating from the Navy.

SWOS should consider processes designed to standardize the assessment because they intend to implement OOD Phase I and II courses in Norfolk, VA and San Diego, CA. This shift in location away from Newport, RI, may create difficulties separating assessment-site grader effects from the effects due to individual homeport effects during the first and third assessments under this new assessment framework for SWO career progression. Standardization between Norfolk, VA, and San Diego, CA, is especially critical since the third assessment point is a “Go, no-Go” assessment. We recommend that SWOS develop a protocol that minimizes the effects of differential grading between the two fleet concentration sites for OOD Phase II, because it may disadvantage officers at one of the assessment sites.

We consulted the military- and civilian-sector bodies of literature to gain insight on how they have addressed similar challenges to what the Navy faces in a crowded maritime domain. Our literature review focuses on research related to the initial selection, training, periodic assessments, and remedial training necessary to ensure the aviation and long-haul trucking industries develop proficient pilots or drivers. We also consulted the simulator-relevant literature to determine if academic research supports alternatives to the traditional training process for building, maintaining, and testing proficiency. We found that both of these industries utilize simulation extensively, and this has reduced overall-training costs for their respective fleets.

---

[2] Figure 1 in Chapter V illustrates both the SWO career progression and associated “Go, no-Go” assessment schedule.

Since the 1920s, the aviation community has invested heavily into answering questions about how much experience is necessary to establish a minimum level of proficiency and what characteristics made the best pilot candidates (Siem & Murray, 1997), (Seltzer & McBrayer, 1971), (Kavanagh, 2005), and (Martinussen, 1996). Martinussen (1996) conducted a meta-analysis that analyzed 50 independent aviation studies designed to predict pilot performance. They found that previous training experience and cognitive testing are the best predictors of pilot performance. Hammon and Horowitz (1990) found that simulated or actual aircraft training opportunities affect pilot performance through improvements in the short and long run. They determined that experience serves to hone critical skills while minimizing skill deterioration in the short run. Similarly, they found that the long-run improvements are the product of pilots achieving a mastery of skills at a higher level.

In the trucking industry, a lack of sufficient driver-level data historically inhibited meaningful research on the relationship between driver experience and performance. This changed in the late 1990s after responsible organizations compiled better databases to correlate the effects of experience and driver behavior to safety-related outcomes (Federal Motor Carrier Safety Administration [FMCSA], 2006). Experience- and driver-related behaviors are significant predictors of safety and performance across the literature (Blower & Campbell, 2002), (Cantor, Corsi, Grimm, & Ozpolat, 2010), (Hallmark, Hsu, Maze, McDonald, & Fitzsimmons, 2009), (Lueck & Murray, 2005), (Lueck & Murray, 2011), and (Monaco & Williams, 2000).

Advances in high-fidelity simulator technology are responsible for significant cost reductions across numerous military and civilian organizations (Schank, Thie, Beel, & Sollinger, 2002). Although the magnitude of the effect varies, they conclude that simulators facilitate additional training opportunities and those hours of training improve operator proficiency. The authors report that tightening budgets and access to training platforms, which affect money for fuel, have reduced the Navy's ability to conduct unit-level training. Additionally, they noted that as fuel availability for OOD underway training diminishes, the return on investment for simulator utilization increases proportionally.

The literature does not answer the question of how skills, knowledge, experience, and currency contribute to proficiency in our ship handlers. Hammon and Horowitz (1990) examined the small body of literature covering experience and currency in the aviation community, and expanded it by conducting three additional studies. These studies examined the correlation between career experience, currency, and pilot performance. Their research is important to policymakers because they found that pilot skills degrade quickly when training opportunities are unavailable, and it illustrates the importance of simulator training's value to overcome experience deficits and maintain currency.

With the knowledge of the approaches demonstrated in the literature, we examined the correlates of OOD proficiency using an ordinary least squares (OLS) model. This method allows us to estimate the partial correlations of explanatory factors with OOD proficiency, holding all other factors constant. Partial correlations are the appropriate statistic from a policy point of view as they allow policymakers to focus on the specific factors that they can influence when considering the reallocation of resources for training or officer selection.

Our results that dealt with the functional inputs of proficiency are generally consistent with the literature. We find that skills, knowledge, experience, and training offered at commissioning source are correlates of proficiency, while ship type and currency are not correlates of proficiency. While ship type is not a surprising result, the lack of correlation between currency and proficiency contradicts the literature. It is important to note that the only currency-related information collected in this pilot study (time since last BRM course) is a poor proxy for the actual information that explains the ship-handling experiences that an OOD accumulated within the last three months. For the initial 164 data points, currency may not be a correlate of proficiency only due to the low-quality currency proxy available in this data.

Our results suggest several ways in which policymakers can affect the proficiency of ship handlers. For example, if a series of near misses due to low-visibility conditions plague the fleet in the future, a prescriptive solution might require all OODs to log an additional specified number of ship-handling hours in simulated low-visibility simulations dedicated to address the reported deficiencies. This prescription has value because we now know what experience as an OOD is worth, on average, when holding all other factors constant. Additionally, the evidence in the literature suggests that improvements to the currency component of proficiency provides benefits due to the overall experience gain and likely aids in breaking the chain of forgetting that occurs when we do not routinely practice mariners' skills. The results suggest that changes in experience level can dramatically improve the safety of our fleet if we target training gaps to bolster an individual's weaknesses.

We recommend that future research focus on the better data reliability that the introduction of the Mariner Skills Logbook will produce. SWOS collected individual data in the pilot study through a survey that required OODs to estimate their experience levels dating as far back as two years. Although this may be accurate for an officer with few days underway or who had recently returned from an extended deployment, it is likely to be unintentionally misreported for the majority of officers in between these extremes. This data-reliability problem should not exist as much in future data samples assuming the compliance rate for logging formal data is sufficient. SWOS intends to mandate a minimum number of hours of experience or a minimum number of evolutions that each officer must complete prior to returning for advanced training and proficiency assessments. This requirement will serve as an incentive for officers to log the minimum hours because failing to do so could otherwise negatively affect their opportunities for career progression. COs are required to certify the logs quarterly and this oversight requirement should encourage accuracy and serve to mitigate erroneous log entries aimed at satisfying the minimum requirements.

Data that is more reliable will allow analysts to refine the estimated effects in our model. The logbooks require officers to log more independent variables than the surveys collected for the pilot-study data set. This means that researchers will have access to more information to develop better models that explain the variations in proficiency. For example, officers that commissioned through Officer Candidate School (OCS) outperformed other commissioning sources in some parts of the assessment. This result was unexpected because OCS graduates acquire the least amount of ship-handling experience prior to their first-tour assignment. We expected to see Naval Academy graduates outperform all other commissioning sources because many of them spend at least one summer training in YPs where they receive extensive ship-handling experience.

YPs are 80 to 100 feet long vessels that the Navy has used to teach navigation and seamanship to Naval Academy graduates since World War II. Many retired and senior officers have touted these platforms as the preferred training solution for OODs in the wake of several ship collisions in the Western Pacific. The pilot-study data suggests that YPs may not be the best allocation of training resources because OCS officers with the least ship-handling experience prior to commissioning perform the best, on average, while holding other factors constant, in the pilot-study assessment. This counter-intuitive result suggests that an omitted variable like age is positively correlated with both commissioning source and the proficiency dependent variable. Since SWOS did not collect age on the survey, the information it would contain may be partially flowing through commissioning source, which is undeservedly getting the credit.

We recommend modest revisions to the Mariner Skills Logbook with the understanding that future analysis will be limited in scope to answering policy-related questions supported by the framework that these logs establish. This logbook revision ensures that researchers have access to important information like age, skills, knowledge, experience, and currency so we can include direct data or the best proxies for that information in future analysis. Future models will better explain the relationship between these factors and proficiency, which enables policymakers to allocate resources effectively.

The remainder of this paper proceeds as follows: Chapter II provides background on Surface Warfare institutional changes as well as a review of the literature on the aviation industry, long-haul trucking industry, and simulators role in training programs. Chapter III describes the pilot study that SWOS developed to assess OOD proficiency. Chapter IV describes the pilot-study data and details our empirical framework with results. Chapter V discusses continuous data collection and testing mechanisms. Chapter VI concludes the thesis.

II. BACKGROUND

This chapter describes some policy changes in the Surface Warfare Officer (SWO) community prior to and after several high-profile ship collisions in 2017. Additionally, we describe the academic research that informed our approach. We drew insights from the aviation and long-haul trucking industries to identify how they develop, maintain, and test proficiency skills through training.

A. SURFACE WARFARE INSTITUTIONAL CHANGES

Technology improvements, such as electronic charting and smart ship installs, on U.S. Navy vessels have forced officers and Sailors to focus more attention on training and additional qualifications in recent decades (Department of the Navy (DoN), 2017). During this time, the SWO community also accepted too many new junior officers for the number of billets available on ships, which the community refers to as "over accession." This practice was intended to mitigate shortfalls for department head-level officers in the future, but the unintended consequence of this practice is a reduction in opportunities for individual SWOs to hone their ship-handling proficiency. This policy causes officers to compete with a crowded wardroom of peers for the same watch standing opportunities to earn their qualifications.

The ship's operational schedule can also have a significant impact on the opportunities available for a new officer to acquire enough underway experience to earn qualifications (Cordial, 2017). For example, officers that report to a ship during an extended maintenance availability may not get underway during the first year of their tour. Watch-standing experience is critical to completing advanced qualifications, and special evolutions are a significant component of developing ship-handling skills. Anchoring, mooring to a buoy, and conducting an underway replenishment (UNREP) are examples of these special evolutions. Cordial (2017) found that the optimal point for new officers to check-in to their first ship is between the six- and nine-month mark of the Optimized Fleet Response Plan (O-FRP). The author suggests that this timing gives the officer the best chance of receiving a sufficient amount of underway time during the basic, intermediate, and advanced phases while still providing enough time to complete a full deployment. Conversely, he found that the worst time for a junior officer to arrive on a ship is between the 25th and 27th month of the O-FRP, which is right after a ship returns from deployment. A post-deployment arrival can cause the new officer to miss critical watch-standing opportunities as the ship enters the maintenance phase.

The Navy redesigned the SWO training pipeline following the collision investigations that occurred in 2017 (see LaGrone, 2018 for details). Accident investigators discovered that navigation and ship-handling proficiency, among other factors, were lacking. The redesigned curriculum requires officers to complete an intensive six-week officer of the deck (OOD) course that includes over 100 hours of simulator training along with numerous mariners' skills focused courses in radar operations and charting in addition to the Basic Division Officer Course after initial accession. Rollout of the new curriculum calls for a four-week version of the course to commence in 2019 and the full course will be online in 2021. Officers can expect to report to their first tour for 30 months, where they can focus on OOD qualification followed by SWO qualification. Following this first tour, officers return for additional training to focus on managing the bridge in preparation for a second tour where they can expect to stand a significant proportion of special-evolution details.

B. LITERATURE REVIEW

We divided the literature review into two sub-sections, which separates theory from research. The theory sub-section discusses how the aviation and long-haul trucking industries develop proficiency and describes simulator fundamental principles. The research sub-section delves into the findings that each of these industries employ to improve safety-related outcomes and describes results related to simulator effectiveness.

1. Theory

This Section outlines transportation platforms that share many challenges with the Surface Navy in developing proficiency, maintaining skills, and testing those skills throughout the individual's career. The platforms we researched include aviation and long-haul trucking. Specifically, we reviewed how these communities conduct training during those three distinctive phases of an individual's training continuum, and documented the methods they use to develop, maintain, and assess individual proficiency.

a. Civilian Trucking Proficiency Development

Civilian trucking federal regulations place licensing requirements in the states' hands, which results in a wide range of variance in requirements. In general, drivers must apply for a learner's permit that allows them to practice behind the wheel with a licensed instructor until they gain enough experience to pass the commercial driver's license (CDL) exam (Mayhew & Peterson, 1993). This process is similar to the process for acquiring an individual driver's license, but CDL requirements are more stringent. Lueck and Murray (2011) chronicled the training and orientation of new drivers and characterized it as a mixture of classroom training, safety media, on-the-job training, and tests that require students to demonstrate proficiency-related skills. The authors found that firms use sustainment training to keep safety at the forefront of drivers' focus after initial training is completed. There is little mention of retesting across the literature, excluding CDL replacements for the purpose of inter-state relocation. That said, many firms conduct remedial training for identified deficiencies and crashes (Lueck & Murray, 2011). This allows firms to retrain drivers that are involved in accidents or demonstrate potential risks to the company if left unchecked.

b. Military Trucking Proficiency Development

Military trucking includes vehicles like the five-ton truck, tractor-trailers, armored vehicles, and tanks. These training programs utilize similar methods, and this Section highlights the methods used to develop driver proficiency. Ninety percent of truck-driver-designated soldiers receive vehicle specific training to operate tanks, armored personnel carriers, and five-ton trucks (Mayhew & Peterson, 1993). The authors observed that the remaining 10 percent of truck drivers receive professional training to drive tractor-trailers.

The GAO (2001) reports that a mixture of formal and informal programs are used to train the Army's 88M Motor Transport Operator designator for trucks. Students receive one week of instruction in the classroom and the remainder of the school is dedicated to on-the-job training. Additionally, they found that follow-on commands provide supplemental training to license drivers. The biggest problem the GAO found during its review was that a shortage of instructors and lack of facilities prevent drivers from learning some critical skills such as driving on snow and ice. They noted that simulators could resolve this training gap by providing the necessary skills for drivers to conduct primary missions. Some private sector truck driving schools use simulators to train students, but the Army had not fully adopted the idea as of the time of this report (GAO, 2001).

The Army's professional tractor trailer operators make up only 10 percent of Army soldiers who drive vehicles (Mayhew & Peterson, 1993). Soldiers with prior tractor-trailer experience can earn the Class A Army Commercial Driver's License (ACDL). This program parallels the Commercial Driver's License (CDL) program for civilians that was instituted in 1986 with the Commercial Motor Vehicle Safety Act (Kubiszewski, 1994). Additionally, Mayhew & Petersen (1993) found that the amount of training the non-professional driving soldier receives ranges from two days to several weeks, which excludes the experience gained during unit-level training after initial qualifications.

Simulators are used across trucking communities in various capacities with differing goals and measures of effectiveness (Goode, Salmon, & Lenne, 2012). Additional information can be found for commercial trucking (Mayhew & Peterson, 1993), (Mejza, Barnard, Corsi, & Keane T, 2003), and simulator usage across military combat vehicles can be found across the literature (Oskarsson, Nählinder, & Svensson, 2010), (Mcdade, 1986), and (Lampton, Kraemer, Kolasinski, & Knerr, 1995).

c. Aviation Industry Proficiency Development

Separated and retired military pilots were the primary accession source for the commercial aviation industry until collegiate and on-the-job training pathways stabilized to provide a sufficient supply of airline pilots (Hansen & Oster, 1997). They found that students who graduate from collegiate programs have approximately 250 hours of flight time on average, but regional carriers typically seek candidates with approximately 1200 hours of experience. Additionally, they report that major-airline carriers have a minimum requirement of 1500 hours. Numerous specialized aviation schools have found a niche to help students find financing to earn sufficient experience and then move into positions where they can gain flight time through on-the-job training to overcome experiential shortfalls (Hansen & Oster, 1997). Once a pilot is trained and passes the exam for a Commercial and Airline Transport License (ATP), they are subject to continuous-training requirements as mandated by Code of Federal Regulations (CFR) (Federal Aviation Administration [FAA], 2013).

According to the CFRs, the Federal Aviation Administration (FAA) requires pilots to complete at least one hour of ground training and one hour of flight training in an aircraft covered by the type of license held. Pilots document completion in a logbook entry signed by an authorized instructor within the previous 24 months before serving as the pilot in command. This regulation allows approved simulators at training centers to count toward this requirement. Other requirements, like 14 CFR 61.58, mandate a series of tests that pilots must pass for multi-crew aircraft. These tests include an exam in the aircraft, a series of emergency procedures, written exams, and pilots must complete Crew Resource Management training every six to 12 months in order to remain in good standing (Flight Deck Friend, 2012).

Reis (2000) reports that military aviation training focuses on a lock-step approach whereby candidates commence training with an aviation introductory course, and then they receive simulator and cockpit training sequentially. Specifically, the authors report that the Navy's curriculum requires pilots to complete six weeks of ground school training, which covers flight rules and regulations, water survival, aerodynamics, aircraft engines, navigation, and meteorology. Students report to primary flight training after learning these basics. Once a student completes primary training, they attend intermediate training in one of the four major airframe types, which include jets, maritime propeller, carrier-based propeller aircraft, and helicopters (Reis, 2000). After graduation, student aviators report to a squadron and continue to focus on qualifications while accumulating flight hours.

Pilots with less than 20 years of aviation experience must accumulate 100 annual flight hours and 40 of them must be logged every six months (Department of the Navy, 2016). Commander, Naval Air Forces (CNAF) also requires 12 hours of experience for both nighttime flying and instrument time. Pilots must accomplish half of those hours of experience semiannually. CNAF also requires pilots to complete 50 percent of the minimum hours in an actual aircraft, which leaves the remaining hours available for accomplishment in a simulator. CNAF mandates approximately half the number of hours in each category for pilots with more than 20 years of experience. In both cases, CNAF requires that pilots of multi-crew aircraft and their crews complete Crew Resource Management training annually, which includes an academic portion as well as a team-simulator evaluation.

d. Simulator Training Theory

Fidelity refers to the level of transfer of training (TOT) a trainee has when switching from the simulator to the actual platform of interest (Jacobs et al., 1990). For example, a high-fidelity simulator should allow a trainee to step into the actual platform and perform procedures that they mastered in the simulator. In contrast, a low-fidelity simulator might result in a lower level of TOT and may require the trainee to practice additional sets of procedures in the actual platform to polish the skills developed in a simulator (Jacobs et al., 1990). There is no consensus on the optimal fidelity for simulators to be effective. Some mission-training objectives may be so critical that we should not simulate them, because the risk of negative TOT may be the difference between life and death. Other training evolutions are more conducive to low-fidelity simulation. Policymakers should consider the risks and rewards for leveraging the spectrum of simulation to fill training gaps. Additionally, they should not get hyper-focused on one level of fidelity as the only acceptable type.

2. Long Haul Trucking Industry Research

The commercial trucking industry expends more resources researching the effects of experience as measured by safety than the maritime industry, but still lags far behind the aviation industry. The primary reason for this lag is the historical lack of sufficient data to explore the relationships between driver-level factors and the appropriate safety or proficiency standard, because the industry lacks a national-level database that allowed researchers to study the causal factors for large-truck collisions in the United States (Federal Motor Carrier Safety Administration [FMCSA], 2006). For example, numerous un-linkable databases contain the pertinent information, and there is a lack of accurate collision reporting.

The available data comes from regional sources and is comprised primarily of police officer reports at the scene of the crash (FMCSA, 2006). This presents a problem with identifying causation, because researchers cannot observe many of the factors that could have contributed to the crash. For example, a truck driver that unintentionally departs from his lane on the highway is not necessarily at risk for collision, but if the lane departure happens at the same time as a passenger vehicle that decides to change lanes without signaling in the same location, an accident is likely to occur. Often, it is difficult for investigating officers to discern the root cause of the crash from all the different risk factors they observe during the investigation. The causal factor of a crash is more difficult for an investigator to identify if the true cause is something that occurred pre-crash but does not surface until the culminating point when the crash occurs. For example, investigators routinely cite fatigue factors as a cause when it is clear that the driver logged insufficient downtime given the distances traveled in a given period. Other pre-crash factors are less readily identifiable, for example, driver distraction or lack of situational awareness.

Because of the challenges with data availability, two schools of thought have evolved to answer tough questions about the determinants of safety and driver-related factors. One side has examined the problem based on the corrective actions the firm should introduce while the other side has sought better data collection methods and more accurate reporting to identify driver-related factors. Both firm and driver-focused methods have contributed to the continued improvement of truck-related safety in the industry, which is on the rise since the peak of trucking accidents in the 1970s (Lueck & Murray, 2005). The driver-related approach parallels the maritime perspective on several fronts and the rest of the literature review that covers the commercial truck industry focuses on these driver factors.

The Federal Motor Carrier Safety Administration (FMCSA) reports that driver factors caused 87.2 percent of truck collisions (Cantor et al., 2010). Similarly, researchers consistently attribute 80 percent of maritime accidents to human factors (Macrae, 2009). These statistics suggest that the research goal should focus on improving proficiency and ultimately safety across both industries. Cantor et al. (2010) were one of, if not the first, to utilize FMCSA data to develop a crash-prediction model based on human factors. They argue the importance of focusing attention on the driver because of the percentage of crashes caused by factors within the driver's control. Cantor et al. (2010) found that driver weight, age, gender, height, employment stability, and past driver violations were significant factors in their crash-prediction model.

Driver behavior and violations, among other factors, show a clear link across the literature from studies with different data sets and different primary research questions. Driver violations predict a significant amount of future accidents (Lueck & Murray, 2011). For example, Lueck and Murray (2011) found citations for a failure to signal a lane change raised the likelihood of a future crash by as much as 96 percent. Driver behavior is significant for understanding how driver-related factors compound and create accident risk. Lueck and Murray (2011) quoted Virginia Tech Transportation Institute research regarding the differences in crash risk across the commercial trucking industry and they found 10–15 percent of drivers are responsible for 30–50 percent of the trucking industry's total risk of crashes. A pattern emerges that explains how relatively few individual factors can balloon to become local issues and ultimately fleet-wide problems if left unchecked. Many of these results were first identified by Lueck & Murray, (2005) and were reconfirmed in later studies (Lueck & Murray, 2011). Behavioral links for safe and proficient drivers are significant to policymakers because changes that incentivize positive behavioral traits for OODs can mitigate the risk for some of the problems to fester and trickle downstream to become fleet-wide problems decades later.

Across the trucking literature, measures of experience consistently deliver statistically significant explanatory powers related to proficiency and safety. Monaco and Williams (2000) found results consistent with the rest of the body of literature when they used experience data to assess what driver-level factors influence safety. Additionally, the authors found that the relationship between age and crashes was typically U-shaped, whereby younger drivers tend to report the highest number of crashes, then the middle-aged drivers tend to report fewer crashes, and finally the older drivers appear to match the younger drivers for the number of crashes reported. Driver experience is a better measure than age for predicting safety and numerous studies attribute higher rates of accidents to the age of the drivers involved. But this conclusion is often due to a lack of the appropriate experience data that contributes to age taking the blame (Hallmark et al., 2009). This finding has direct implications for Navy policymakers because OODs tend to be young and relatively inexperienced. Policies that improve opportunities for OODs to spend more time on ships, more time on the bridge, and more time in simulators directly affect the safety of the fleet.

3. Commercial and Military Aviation Research

The aviation community leads other industries in research involving human-factors variables as they relate to proficiency. Human-factor issues are attributed to as high as 90 percent of mishaps and accidents in the aviation community (Diehl, 1991), which is similar to what researchers found in the trucking and maritime domains. The aviation industry began extensive research into human factors and the best training methods to develop pilots between World Wars I and II to meet the rapidly growing demand for quality pilots (Martinussen, 1996). Martinussen (1996) documents studies as early as 1921 that attempted to quantify and qualify determinants of top pilots. For example, researchers once believed that quiet and methodical men would make the best candidates for pilots. As methods improved, researchers examined endless combinations of psychomotor, personality, and intelligence tests to produce better quality pilots, but meta-analysis results from 50 studies suggest the best predictor for pilot performance is previous experience in aviation training (Martinussen, 1996). This holds true across numerous studies and meta-analyses for both pilot-performance prediction before selection as well as performance over the course of a career after initial selection.

An analysis of carrier landings demonstrated a strong link between both the total number of hours a pilot accumulated and the number of training hours in the prior month (Hammon & Horowitz, 1990). The link Hammon and Horowitz (1990) observed extended across multiple platforms as well as several different types of data. They tested a similar hypothesis on a Marine Corps database that contained objective measurements on bombing missions across three different platforms. They found that pilots with greater career experience than newer pilots dropped bombs more accurately as measured by the mean distance to the target. Additionally, pilots with recent experience in the last seven days delivered bombs closer to the target than pilots with less recent experience for manual bomb delivery systems such as the F-4S aircraft, which does not have an automated bomb delivery capability (Hammon & Horowitz, 1990). Lastly, they tested whether overall experience predicts proficiency in an air-to-air combat data set. Total career flight hours proved to be the single greatest factor for determining probability of an air-to-air fighter killing his opponent and experience is also the greatest factor for reducing the probability that he loses a dogfight (Hammon & Horowitz, 1990).

These findings are critical to policymakers because they suggest that there is a quantifiable relationship between a reduction in experience and a reduction in proficiency. The Research and Development (RAND) Corporation searched extensively for similar studies for non-aviation contexts and concluded that such studies either do not exist or were completed and maintained by individual services instead of finding their way into the academic domain (Kavanagh, 2005). This suggests that more research in other contexts could confirm these patterns elsewhere. Policymakers should understand that a relatively small change in training resources or fuel for underway experience can have a direct effect on OOD performance if it is not supplemented with simulator training designed to mitigate the effects of skill atrophy.

Evidence of the impacts from skill accumulation and subsequent degradation is visible across the military and civilian sectors of aviation. For example, research suggests that pilot instrument skill is volatile both in terms of degradation as well as regaining it with deliberate practice and instruction (Seltzer & McBrayer, 1971). Skill-degradation research began as early as 1934 in the Boeing School of Aeronautics. Researchers discovered a correlation between requiring instrument flying prior to contact flying because this change resulted in higher quality pilots, and this method accelerated the learning curve for contact flying. Instrument flying refers to conditions where pilots must use instruments in the cockpit to navigate because they lost visual cues. Contact flying refers to when pilots primarily use visual cues as their means of navigating while referencing instruments as a secondary option.

When the student was introduced to contact flying before instrument flying, the opposite is true (Seltzer & McBrayer, 1971). Researchers expanded their understanding of these correlations throughout the 1950s and by 1961; the FAA mandated a minimum of 10 hours of instruction in instrument flying before an applicant could get a commercial pilot license. The results from Seltzer and McBrayer's experiment (1971) link increased hours of experience on instrument aviation with increased pilot proficiency and this concurs with similar research by McFadden (1997) over two decades later in completely separate aviation contexts.

4. Training and Simulator Effectiveness Research

In 2005, the U.S. Navy maritime patrol aircraft community examined their usage and determined that simulators were sufficient for approximately 50 percent of mission-training exercises and basic-flight missions (Yardley, Thie, Schank, Galegher, & Riposo, 2005). The authors found that the Canadian Navy performed a similar review for shipboard training, which resulted in a change to require at-sea training for only the most challenging exercises like multi-ship maneuvers. The Canadian Navy embraces this training methodology so strongly that they couple ship procurement with simulator construction to improve the long-term quality of their training continuum. This procedure lowers the life cycle cost of training personnel and results in a better-supported trainer throughout the ship's life cycle (Yardley, Thie, Schank, Galegher, & Riposo, 2003). Additionally, they found that the British Royal Navy uses simulators heavily, which allows them to allocate precious underway time for primary mission areas instead of routine training objectives.

The literature is consistent regarding how powerful simulators are for training (McLean, 2012). Researchers routinely cite Transfer Effectiveness Ratio (TER) as proof of simulator effectiveness, but these authors argue that it is methodologically questionable at best. They describe TER as a ratio, which compares hours saved in an aircraft to the hours used in a simulator. Their research concludes that most studies provide a certain number of hours in an aircraft to the control group in order to obtain a base level of competence but offer the treatment group additional time in a simulator. This method creates a disparity between the hours of training for the two groups and researchers should report these conditions to prevent the abuse of the TER statistic. This provides strong evidence for the value of experience as a main determinant of proficiency, but it also unfairly inflates the value of the simulator-training time. At that point, it is impossible to separate the value of the simulator from the value of the extra training (McLean, 2012).

A meta-analysis of jet pilot training effectiveness revealed simulator training, in addition to aircraft training, consistently produced superior results over aircraft-only training (Jacobs, Prince, Hays, & Salas, 1990). Additionally, their research suggests that trainees should reach a certain level of proficiency before advancing instead of the lock-step modules that force all trainees through the program together. In other words, the individual trainee's proficiency attainment should be the primary trigger to advance through the training pipeline instead of group proficiency or time-based measures.

The U.S. Navy aviation community leveraged their adoption of simulator usage to review training requirements line-by-line, which allowed them to separate items that required trainees to perform those evolutions in an aircraft (Judy, 2018). This process allows decision makers to push more tasks from aircraft training schedules onto simulator training schedules to save costs and time. Some policymakers have determined that simulators are insufficient for some training objectives, which limits the scope of such transfers. For example, Yardley et al. (2005) noted that the Navy's review of U.S. fighter strike mission training requirements determined that only a small percentage could utilize simulators due to the nature of the training objectives.

Earlier research by Schank et al. (2002), studied the optimal balance between simulator and actual platform training. They found that F/A-18 pilots average one hour of simulator time per month due to a lack of perceived fidelity and poor accessibility for the simulators then in use. Finally, they concluded that there is insufficient evidence in the literature to proclaim the magnitude of the simulator's effect toward F/A-18 pilot proficiency, despite plenty of evidence that suggests that simulators are effective. Moving forward, the authors challenged the Navy to reconsider the process used to measure readiness. Under then-current guidance, they found that the Navy measures readiness by units attaining a minimum level of demonstrated proficiency.

U.S. Navy aviation leadership continually analyzes their readiness policies and commissions studies to understand the implications of future-policy changes. Judy (2018) reports that the Navy estimated 61 percent of naval aviation training was accomplished in an aircraft in 2010. The Navy was interested in reducing that number to as low as 44 percent if they could maintain levels of training and measures of pilot effectiveness. This is a major concern for the aviation community, because although simulator technology advances have produced positive results on average, there is concern that the fidelity has not reached the requisite level to mirror certain critical flying procedures. With this knowledge in mind, Judy's (2018) primary question was whether flight simulation time or actual time in the aircraft were more predictive of trainee's Naval Standard Scores at the end of intermediate and advanced flight training. Unsurprisingly, he found actual aircraft time did a better job of predicting the relative performance built into the trainee's Naval Standard Score, but he did not address the potential of reverse causality that biased his estimate on simulator training's effectiveness. The author of this study did not address the fact that low-performing students were more likely to seek additional simulator training than high performers, which introduced the potential for reverse causality. Judy (2018) also concluded that simulator training was more efficient, effective, and utilized more in the early stages of aviation training than it was in the advanced phases. This is important for policymakers to consider because there may be an upper bound on when simulators are no longer effective for teaching certain skills, but researchers have not tested the existence of such a limit.

Yardley et al. (2005) analyzed how the Navy allocates training time and supports decisions with policy. Their study focused on ways to reduce underway training across warfare areas by expanding the use of simulators in port for DDG-51 class ships. The authors concluded that the Surface Navy could execute a large number of training exercises in port, but they were not doing so due to policy restrictions or a lack of identified equivalencies. Their report defines an equivalency as at-sea training that ships could accomplish in port through drills or simulation under certain conditions. Yardley et al. (2003) observed that high-frequency evolutions had the fewest equivalencies that allowed units to accomplish evolutions in port despite having the advantage of the greatest potential savings, such as piloting by gyro and piloting during low-visibility conditions. They reported that low-frequency evolutions had the highest proportion of equivalencies, but since these exercises often had a biennial-periodicity, they offered the lowest savings for training, such as choke point transits for anti-submarine warfare training. Lastly, they determined that Navy culture might need to change to reap the benefits of optimizing the use of simulation in the future. Judy (2018) reported similar cultural challenges in fighter-pilot communities, which led to low-adoption rates for simulator usage in certain aspects of training.

5. Summary

The literature demonstrates the importance that skills, knowledge, and experience—both cumulative and current—plays in the safe operation of vehicles in the trucking and aviation industries. Researchers have found that proficiency development is the result of short and long-term training processes and currency reflects how recently trainees acquired experiences. Policymakers use on-the-job training mixed with simulators to build robust training programs that provide the necessary experience and mitigate the degradation of skills that occur when skills atrophy from non-use. Unfortunately, the literature offers little insight about how a reduction in opportunities for OODs to develop these functional inputs at sea affects mariners' proficiency. Our research analyzes the performance of first tour OODs to correlate skills, knowledge, experience, and currency to proficiency. Our approach provides insight into how underway and simulated experiences are related and provides policymakers with tools to assess proficiency at the individual level.

III. OOD PROFICIENCY EXERCISE

The reader should recall that there are three distinct periods associated with the evolution of ship-handling proficiency assessments that we refer to throughout our research. This chapter focuses primarily on the pilot study that Surface Warfare Officers School (SWOS) developed as a first step to understanding the correlates of proficiency. The reader should also recall that SWOS continues to conduct proficiency assessments as we develop better protocols. The purpose of this chapter is to frame the pilot study so that we can offer recommendations designed to enhance the process for future assessments.

A. OVERVIEW

In order to examine the correlates of OOD proficiency, we use data collected by SWOS as directed by Congress, the Secretary of the Navy, the Chief of Naval Operations, and United States Fleet Forces in response to several ship collisions during the summer of 2017. Leadership selected a random sample of 164 first tour OODs from 61 ships across seven homeports. Officers filled out a demographic survey and completed a 45-minute OOD proficiency check in a moderate traffic-density simulation. Additionally, officers completed a rules of the road (ROR) exam and a naval seamanship and ship-handling (NSS) exam. This is an individual-level data set and includes several measures of OOD proficiency as demonstrated during the exam, including: bridge team management (focused on managing the personnel), use of bridge resource management (focused on managing bridge equipment), leadership and presence, performance under stress, application of ROR, ROR exam score, and NSS exam score. Senior assessors graded OODs on all of these skills, except the two-written exams. SWOS scored the ROR and NSS written exams as a percentage of total correct answers for each test respectively.

1. Pilot Study Assessment

The OOD proficiency exam began when an officer met with SWOS staff at the fleet testing sites. SWOS staff members were dressed in civilian clothes to mitigate concerns that OODs, who are O-1s and O-2s, might have when they commence the scenario with a SWOS-assigned junior officer of the deck (JOOD) and conning officer (CONN) wearing

O-3 rank insignia. The senior assessor wore an appropriate uniform to simulate a Commanding Officer (CO) throughout the assessment.

After arriving, SWOS staff members followed a checklist designed to standardize the experience for each assessment and provided officers with a Non-Disclosure Agreement.[3] The agreement highlights the objective of the assessment, which is to assess overall performance as a qualified OOD in a written exam and simulated underway scenario. The agreement requires the officer to acknowledge that they may not share examination questions or details about the simulated scenario with individuals outside of the assessment team unless an appropriate representative authorizes them to do so in writing later.

Once signed, the OODs completed a questionnaire that recorded basic demographic and experience information that they accumulated during their operational tour.[4] Then, officers moved onto the written portion of the exam. The exam has a disclaimer that states that SWOS adapted the questions from pertinent Navy Personnel Qualification Standard(s) (PQS). This document is not included as an appendix so as to prevent future cohorts from training to the test. Interested parties should contact the authors with questions related to this document. The first half of the exam covers ROR and the second half covers NSS questions.

After completing the exam, SWOS briefed the officer with relevant information for the simulation.[5] The brief contains information similar to what an underway OOD might expect during a pre-watch brief with the Combat Information Center and off-going OOD. The brief included a prior engineering casualty, speed limitations, course, speed, location, next planned event, and the average speed of advance required to conduct the follow-on evolution as scheduled. The brief concluded with visibility, weather, traffic patterns, and CO’s standing order requirements. The senior assessor introduced him/herself and

---

[3] The OOD Competency Check Checklist and Non-Disclosure Agreement are included in Appendices A and B.

[4] This questionnaire is included in Appendix C.

[5] This document is not included as an appendix to prevent future cohorts from training to the test. Interested parties should contact the authors with questions related to this document.

provided information on how to contact them for reports. SWOS provided an opportunity to clear up any questions before moving into the simulator.

After arriving in the simulator, a SWOS instructor (representing the off-going OOD) briefed the status of contacts within visual range and specified which contact reports the CO had previously received. The officer was offered a familiarization session for the simulator layout, if desired. Then, the CONN and JOOD introduced themselves. Additional SWOS staff members, in civilian clothes, filled these roles. Both watch standers described their experience levels, which included the bridge equipment they were familiar with and proficient. SWOS provided the OOD time to set up the Automatic Radar Plotting Aid (ARPA) and Voyage Management System (VMS). During the familiarization time, the JOOD and senior assessor graded changes the OOD made to ARPA and VMS from the default configuration. Lastly, SWOS provided a generic contact report script that the JOOD would be prepared to fill out to facilitate contact reports to the CO.[6]

Once the officer verbally took responsibility for the deck, the senior assessor and simulator operator started the scenario. The senior assessor and JOOD followed a script and rubric that outlined the decision points and hazards to navigation throughout the scenario.[7] Each line in the script corresponded to an approximate estimated time into the scenario. Additionally, the script listed the anticipated decision the OOD would make as well as the actions the senior assessor would take and questions they would ask. The senior assessor's responses varied from directing radio communications with other vessels to asking questions about ROR situations related to the OOD's maneuvering decisions. As the officer negotiated each hazard along the planned navigational track, the senior assessor annotated a score for communications with the CO, communications with the vessel, correctly applying the ROR, and executing maneuvers as briefed to or directed by the CO. Every line item listed under each checkpoint was graded as "unsatisfactory, requires improvement, meets standards, or exceeds standards."

---

[6] A copy of the contact report is included in Appendix D.

[7] The script and rubric are not included as an appendix to prevent future cohorts from training to the test. Interested parties should contact the authors with questions related to this document.

During the scenario, a SWOS assessor graded the written exam results and recorded them along with the survey results in an excel document. After the scenario, the senior assessor and JOOD debriefed the OOD, addressed areas of concern, and reemphasized the importance of the Non-Disclosure Agreement. The JOOD prepared the simulator with identical starting configurations for the next officer and the senior assessor emailed an assessment to the OOD's CO.

2. Pilot Study Survey Questionnaire

The survey included 23 questions that collected information about each OOD's background, time required to qualify, and ship-handling experiences. Commissioning source offered three alternatives, which were United States Naval Academy (USNA), Reserve Officer Training Corps (ROTC), and Officer Candidate School (OCS), but one candidate did not fall under these categories and recorded "limited duty officer (LDO)" instead. The number of months assigned aboard current ship provided options for six months or less, seven to 12 months, 13 to 18 months, and greater than 18 months. The number of months on deployment provided options for three months or less, four to six months, seven to 11 months, 12 to 17 months, and greater than 18 months. The number of months to qualify as an OOD provided options for six months or less, seven to 12 months, 13 to 18 months, and greater than 18 months. The number of months to qualify SWO provided options for not qualified, one to six months, seven to 12 months, 13 to 18 months, and greater than 18 months.

The questionnaire then asked a series of yes or no questions that covered equipment-related responses. This question asked whether the officer felt comfortable operating ARPA in congested waterways. Additionally, the question asked whether they understood where VMS inputs come from, how the system determines the ship's position, and whether officers felt confident using VMS during transits. Additional questions listed on the questionnaire are available in the appendix.

The questionnaire then asked officers to report the length of time since they last attended Bridge Resource Management (BRM) training, which is a course conducted at the Navigation, Seamanship, and Ship handling Training (NSST) centers in fleet concentration areas. These options included never, one to three months, four to 11 months, 12 to 18 months, and greater than 19 months.

Lastly, the questionnaire solicited information on the ship-handling experiences of each officer as a CONN and as an OOD. These questions focused on the number of special evolutions that the officer conducted at each watch station and the total number of watches at each station. These were in two separate sections in the survey, but we combined them here to show the parallels between the collection processes for both watch stations. The question for getting underway and mooring to a pier provided options for none, one to two, three to four, five to six, and greater than six total evolutions as the CONN or OOD respectively. The number of strait transits and high-density traffic environments provided options for none, one to four, five to eight, nine to 12, and greater than 12 times for CONN and OOD respectively. The approximate number of days the officer stood watch underway provided options for less than 20 days, 21 to 99 days, 100 to 200 days, and greater than 200 days as the CONN or OOD respectively.

3. Pilot Study Sampling Methodology

Officers represented six fleet concentration areas: Norfolk, VA; Jacksonville, FL; Everett, WA; San Diego, CA; Pearl Harbor, HI; and Sasebo/Yokosuka, Japan. One senior assessor graded officers in Norfolk, Jacksonville, and Everett, while three other senior assessors were solely responsible for grading officers in one of the three remaining ports (San Diego, Pearl Harbor, and Sasebo/Yokosuka). Unfortunately, this assignment of assessors to ports does not allow us to distinguish between the impact of a port and the assessor without assuming that every senior assessor graded officers identically. For example, if one homeport outperforms another homeport with different graders, there is no way to determine whether this differential is due to something about the performance of officers in a particular location or if it is the result of the senior assessor’s grade assignment. This is important for the future for data collection opportunities because SWOS should randomly assign assessors to students. This design change will allow future analysts to isolate the homeport effects, and this information will allow us to understand if homeport assignment is a correlate of proficiency.

B. CRITICISMS OF PILOT STUDY DATA COLLECTION

We have identified a few shortcomings of the data collection process and the plan that SWOS intends to use for future data collection. These range from the questions solicited on the survey to the rubric for assessing the proficiency simulation. This Section serves as a repository of identified concerns and potential solutions that would make future data-collection opportunities more effective in leveraging data to optimize the development of proficient mariners and minimize skill degradation.

1. Survey Questions

The survey asks the respondents to indicate whether they commissioned through USNA, ROTC, or OCS. Some ROTC schools have ship-handling simulators, and their students accumulate significant amounts of experience prior to commissioning. Anecdotal evidence from SWOS suggests that ROTC graduates from schools with simulators performed well compared to OODs from ROTC schools without simulators. Despite the survey including the school names for ROTC graduates, SWOS did not provide this information with the data included for our research. We recommend including this information in future data sets by breaking the survey response into two categories (ROTC with simulator and ROTC without simulator). Additionally, we recommend adding a separate question that asks whether the officer had ship-handling experience through simulators prior to commissioning. This information would allow us to answer whether these pre-commissioning experiences are correlates of proficiency.

Many of the survey questions only allow categorical responses to data that is inherently non-categorical. For example, officers recorded the number of months assigned to a ship, the number of months spent on deployment, and the number of days the respondent stood watch as the OOD in bins, such as “0-6 months” or “21-99 days.” While it may seem that collecting data in pre-defined bins is simpler and less prone to error, the benefit to collecting the true underlying data vastly outweighs any added cost or complexity. Quite simply, researchers or policymakers using this data can easily categorize the data, as necessary, but only if that raw data exists. We suggest that SWOS collect all data in the most disaggregated manner possible. In particular:

1. Record the number of months for time spent aboard ship, underway, in port, in a maintenance availability, time since attending BRM course, to qualify OOD and SWO.

2. Record the number of hours for time spent on watch bills as OOD, JOOD, and CONN for both underway time and simulator training.

3. Record the number of evolutions for special sea and anchor details such as anchoring, pier work, underway replenishments, straits transits, and TSS transits for both underway time and simulator training.

The watch station specific questions are broken down into categories that cover experiences as CONN and OOD. The survey fails to address experiences as a JOOD and does not address whether respondents should include simulated experiences in the data collected. These oversights likely resulted in a mixture of how respondents answered all questions on the survey. The most likely response may have been to discount these valuable experiences entirely since they did not fit into any category on the survey.

The survey asks the respondent to provide an estimate of how long it has been since they last attended a formal BRM course at a NSST center. This question provides the only insight into the currency and simulator experience of each individual but does so in a way that minimizes the potential interpretations of this variable. This variable is a poor proxy for the real currency information we would prefer to utilize, and this constraint is the primary factor limiting the interpretations we can expect as a result.

The pilot-study survey partially addresses experience's contribution but fails to address currency's contribution to develop proficiency in a direct manner. In November 2018, after the OOD pilot study, Commander, Naval Surfaces Forces (COMNAVSURFOR) issued COMNAVSURFORINST 1412.6, which mandates "proficiency requirements." COMNAVSURFOR refers to the concept that we describe as "currency" as "proficiency" for OODs, among other watch standers. COMNAVSURFORINST 1412.6 identifies the minimum hours of experience required to remain qualified, and it provides solutions to resolve an inability to maintain the prescribed level of "proficiency." The instruction references three times to distinguish levels of "proficiency" and actions to attain these levels. The first period is from one to 45 days, the second is from 45 to 90 days, and the third is for watch standers that have not stood a particular watch for greater than 90 days. If qualified watch standers continue to stand the prescribed amount of watch within the 45-day window, they maintain their "proficiency" and no additional actions are required. If they fall into the 45 to 90-day window since their last watch, the instruction provides guidance and additional actions to regain "proficiency." Once watch standers fall outside of the 90-day window, the instruction identifies additional requirements to reestablish "proficiency."

Researchers in the aviation industry generally consider training to be "current" if it was accomplished in the last four to six weeks, and they have found that currency is responsible for as much as 25 percent of a pilot's proficiency (Hammon & Horowitz, 1990). Additionally, they found that career hours make up the other 75 percent. Details on the lapsed time since each officer last accumulated ship-handling experience could have explained the role currency plays in ship-handling proficiency through mitigating skill degradation. This information would have provided a better proxy for currency than what the "time since BRM course" variable provides.

Within the 164 assessments, some OODs with little underway experience exhibited strong proficiency, while other OODs accumulated significant amounts of underway time but performed poorly. The choice of survey questions, lack of relevant questions, and the categorical nature of the responses to these questions limit the conclusions we can draw in our analysis. SWOS has since resolved many of these issues through the implementation of the Mariner Skills Logbook. We recommend that SWOS continue to focus on improving the data-collection method and types of data collected, because these variables provide the foundation that will support the future-policy effectiveness questions that SWOS may want to analyze.

2. Proficiency Simulation Assessment

The SWOS-developed OOD proficiency assessment check sheet, as it presently exists, may provide a reasonably accurate snapshot of OOD proficiency in a qualitative sense, but it may be too subjective to be useful in future quantitative analysis. SWOS could incorporate more objectivity in future assessments by modifying the check sheet that the senior assessor uses to evaluate OOD proficiency. SWOS could develop “leveling” training for the senior assessors, which would lessen the subjective differences between assessors. For example, SWOS could develop a video-recorded scenario where the assessor explains the intricacies of the more subjective areas of the assessment and explain why the OOD’s decision may have missed the mark of acceptable performance. This process would help to ensure that senior assessors grade all OODs more consistently.

SWOS designed the proficiency check sheet to accomplish several objectives. The first objective was the need to assess the proficiency of the OODs, which was a first step to understand the proficiency level of the Force. The second objective was to provide training and feedback to the individual OOD, and the third was to provide feedback to the OOD’s CO via email after the simulation. Assessing an OOD is subjective by nature due to factors such as style and the interpretation of the Rules of the Road that mariners must observe, and “leveling” training for senior assessors would help to mitigate this factor.

The senior assessor uses the proficiency check sheet’s assessment points to grade OODs on a zero to five scale with only four categories. The categories include “UNSAT” (zero points), “Requires Improvement” (one point), “Meets Standards” (three points), and “Exceeds Standards” (five points). A fifth category could provide the senior assessor with a more normally distributed grading option. The middle category would become the “average” score and there would be options for two standard deviations above and below the “average.” Alternatively, SWOS could alter each assessment point to reflect a binary outcome. This design would lend itself to a more objective format at the cost of lower-fidelity feedback to the OODs and their COs.

Additionally, a binary outcome would translate to a more objective overall score for the dependent outcomes. Some post-assessment check sheets have annotations, which may have come from senior assessors or SWOS personnel adding up individual component scores from the assessment points. The number of assessment scores that senior assessors summed using this method is unclear because many of them were unavailable for analysis. If you assume every decision point is exactly evenly weighted, this is likely the most accurate method for deriving a final score. Other senior assessors calculated these scores mentally without writing them down or simply estimated each of these final scores without calculating any of the individual component assessment-point scores. There does not appear to be any standardization for this part of the grading. Training assessors to one common standard and altering the format of the assessment check sheet could facilitate a more objective outcome. More objectivity would translate into better fidelity in the data for future analysis.

The proficiency final assessment was limited to three alternatives. Those alternatives included "Complete – No Concerns," "Complete – Concerns in the Following Areas," and "Significant Concerns." This format appears to approximate a normal distribution using the middle category as an "average" and the alternatives covering one standard deviation above and below. The interpretation of the category titles for proficiency likely thwarted this design feature. Two of the three grades seem to depict an OOD that successfully passed while the third category depicts a failure. Alternatively, one could argue that only the OODs that earned the highest mark is considered "proficient" while the remaining OODs were "not proficient" because both lower categories were marked with "concerns" about performance. The labels that SWOS chose may aid with the feedback for individuals and their COs, but it likely creates an unnecessary distinction for the senior assessor and the analyst. SWOS could make this distinction more clearly by making the overall proficiency a binary outcome. For example, either an OOD passes and is proficient or fails and needs remediation.

A better alternative is expanding the assessment's rubric to allow a wider variation of scores, which would induce more variation between individual OODs. Then, a cut off could be established that differentiates proficient from non-proficient OODs similar to the binary outcome proposal. The advantage under this proposal is that more information is usually better for analysis. For example, a proficiency score on a scale from zero to 100 would allow us to analyze the variation between an OOD who scored a 90 from an OOD who scored a 70. Additionally, we could establish a cut off between those two scores, which would allow us to divide the sample between those who passed and those who failed for additional analysis. This method does not work in reverse. For example, if we assess the sample as a binary outcome during grading, we are not able to convert those scores into continuous variables that hold relevant meaning because of a lack of variation in that type of sample.

IV. EMPIRICAL FRAMEWORK AND RESULTS

We analyzed the data that SWOS collected during their pilot study, and we provided recommendations for future data collection and analysis. To guide our analysis, we first identified a theoretical functional relationship between proficiency and its determinants, drawing on insights gained through discussion with subject matter experts at SWOS and supplemented with a review of the literature on proficiency development. We define proficiency as a function of the skills, knowledge, experience, and currency of that experience. Some inputs in our theoretical model are not directly observable to both researchers and policymakers yet we may find proxies through observed variables. For example, motivation is a likely determinant of proficiency; we do not observe a direct measure of motivation, but we do understand that the development of skill and accelerated accumulation of knowledge are often due in part to individual motivation.

There are three primary policy goals of this research. The first of these goals is to identify the current level of proficiency for the stock of OODs in the Navy. Secondly, we need to understand how we can leverage the knowledge of these inputs in our theoretical model to improve proficiency at both the individual and fleet levels. Improved training, retention, detailing of proficient OODs, and potentially higher-quality selection or qualification standards for new OODs may allow us to accomplish these objectives. Lastly, we must be able to track the mariners' skills proficiency of SWOs throughout their career, so that we can measure their development and ensure their readiness to progress to higher levels of shipboard responsibility.

These policy goals force us to consider the definition of proficiency. Specifically, we should consider whether we should measure proficiency subjectively or objectively. We must also consider whether overall proficiency is more important or if it is more relevant in specific skills, for example BRM, performance under stress, or application of the ROR. Lastly, we must understand whether the big-picture objective is to maximize individual proficiency or to ensure the entire fleet body of OODs pass a minimum standard, because this consideration will drive solutions for follow-on issues like resource allocation.

We decomposed our theoretical model's inputs into constituent parts so we could classify and organize the different information that SWOS collected. The skill component is comprised of the individual assessment categories, which include management of the bridge team (BTM), BRM, formality/presence/leadership, application of ROR, and performance under stress. We divided knowledge into the ROR and NSS written exam components, which demonstrate the theoretical ship-handling information the officer has learned. We divided experience into several smaller sections, which included time spent on a ship, time spent in various bridge-watch positions, the characterization of their watches by traffic, as well as the number of special evolutions that officers reported on the survey. Lastly, we recognized that currency is a representation of the rate of decay for critical OOD skills. The existing data set did not account for currency directly. Instead, the proxy used for currency is limited to the "time since last attending the BRM course." We recommend that future collection opportunities focus on this critical input by utilizing the SURFOR definition of currency (termed "proficiency" in their instruction), which they define as experience in the last 90 days.

We use multivariate regression models, which are the empirical analog of our theoretical model, to estimate the partial correlations between inputs to proficiency and assessed proficiency measures. In particular, we regressed the assessed proficiency score for each OOD on the functional inputs categorized by skills, knowledge, experience, and currency. Multivariate regression allows us to estimate the partial correlations for each of the functional inputs while holding all other inputs fixed. For example, we can determine what happens to an OOD's proficiency due to increasing their overall experience without changing their skills, knowledge, or currency.

A. SUMMARY OF DATA

Table 1 contains the means for the demographic variables collected during the survey. SWOS collected the survey data in categorical bins, and we converted these bins to a continuous variable for ease of summary, imputing with the midpoint of the range in each category.[8] All officers in the sample were on their first division officer tour, and there is no missing data on either the tests or the surveys. There is a broad sampling of ship classes, but the pilot-study sample does not represent all ship types. For example, the pilot study did not include aircraft carrier and patrol craft OODs. Second tour OODs or non-SWO officers typically fill these billets. Amongst the 61 ships, 10 were Cruisers (CG), 35 were Destroyers (DDG), five were Amphibious Landing Helicopter Dock ships (LHD), five were Amphibious Landing Platform Dock ships (LPD), four were Amphibious Landing Ship Docks (LSD), and two were Mine Countermeasure ships (MCM).

For commissioning source, USNA commissioned 51 officers, the ROTC commissioned 68 officers, and Officer Training Command, which includes both Officer Development School (ODS) and OCS, commissioned 45 officers in the sample.

Table 1. Demographic Data Summary Statistics for OOD Pilot Study Proficiency Assessment

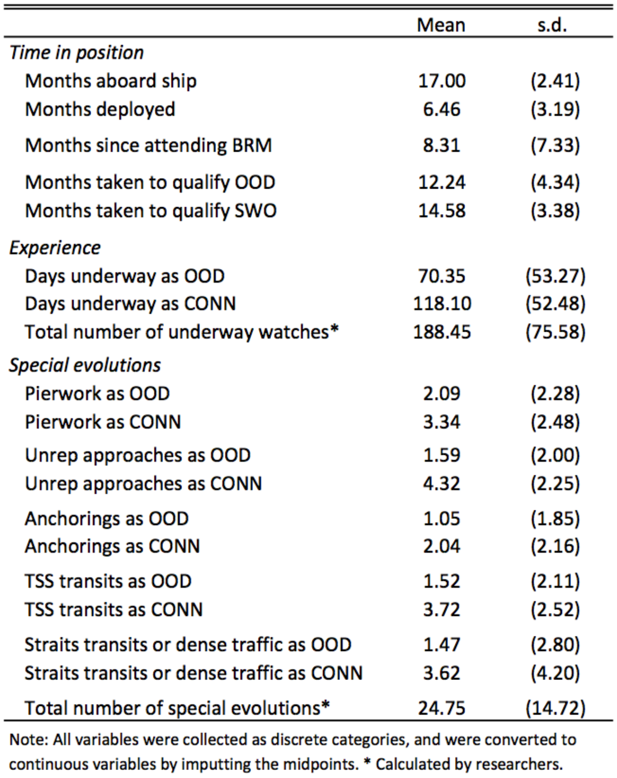
	Mean
Commissioning source	
Officer Candidate School	0.27
US Naval Academy	0.30
ROTC	0.43
Ship class	
Cruiser	0.21
Destroyer	0.52
Mine Countermeasures (MCM)	0.01
Landing Helicopter Dock (LHD)	0.07
Landing Platform Dock (LPD)	0.07
Landing Ship Dock (LSD)	0.11
Home port	
Everett, WA	0.08
Mayport, FL	0.11
Norfolk, VA	0.21
Pearl Harbor, HI	0.14
San Diego, CA	0.24
Sasebo, Japan	0.08
Yokosuka, Japan	0.14
Observations	164

---

[8] For example, the survey asked the respondents to report the number of months assigned to their ship, and we converted observations from zero to six months into three months, seven to 12 months into nine and a half months, 13 to 18 months converted into 15 and a half months, and greater than 18 months converted to 19 months.

Table 2 contains the mean and standard deviation for the time aboard ship and experience-related variables collected during the survey. The average OOD completed six months of deployment, qualified OOD at the one-year mark, and qualified SWO at 15 months, which is roughly on par with the expected times for these major milestones during an OOD's first tour. It is evident that this sample of OODs had much more experience as a CONN than OOD. This result in turn explains why the sample, on average, reported more special evolutions as CONN than OOD. Looking deeper in to the data, it stands out how many OODs reported no experience in each of these special evolutions. Second-tour OODs usually carry the burden for many of the special evolutions that a ship conducts. This may be one area where additional research could determine what proportion of special evolutions each cohort of first- and second-tour OODs are conducting and if that burden sharing is appropriate to maintain a healthy community of OODs.

Table 2. Input Summary Statistics for OOD Pilot Study Proficiency Assessment



	Mean	s.d.
Time in position		
Months aboard ship	17.00	(2.41)
Months deployed	6.46	(3.19)
Months since attending BRM	8.31	(7.33)
Months taken to qualify OOD	12.24	(4.34)
Months taken to qualify SWO	14.58	(3.38)
Experience		
Days underway as OOD	70.35	(53.27)
Days underway as CONN	118.10	(52.48)
Total number of underway watches*	188.45	(75.58)
Special evolutions		
Pierwork as OOD	2.09	(2.28)
Pierwork as CONN	3.34	(2.48)
Unrep approaches as OOD	1.59	(2.00)
Unrep approaches as CONN	4.32	(2.25)
Anchorings as OOD	1.05	(1.85)
Anchorings as CONN	2.04	(2.16)
TSS transits as OOD	1.52	(2.11)
TSS transits as CONN	3.72	(2.52)
Straits transits or dense traffic as OOD	1.47	(2.80)
Straits transits or dense traffic as CONN	3.62	(4.20)
Total number of special evolutions*	24.75	(14.72)

Note: All variables were collected as discrete categories, and were converted to continuous variables by imputting the midpoints. * Calculated by researchers.

Table 3 shows the percentage of the sample that were graded in each of the four proficiency categories "unsatisfactory" through "exceeds standards" for each of the proficiency sub-categories. It is clear that the vast majority of OODs earned a score of "requires improvement" or better. It is clear from the category titles that "unsatisfactory" is a failure, but the distinction between pass and fail becomes vague when the senior assessor deliberates between grading an OOD as "requires improvement" instead of "meets standard." The titles suggest that "requires improvement" is sufficient but in reality, "meets standard" seems to be a more appropriate minimum standard for passing the individual skill assessments.

Table 3. Measurements of Individual Sub-category for Pilot Study Proficiency Assessment

	% Unsatisfactory	% Requires improvement	% Meets standard	% Exceeds standard
Management of bridge team	1.2%	7.9%	72.0%	18.9%
Effectively used bridge resources	1.2%	38.4%	48.8%	11.6%
Formality / Presence / Leadership	1.8%	16.5%	64.0%	17.7%
Practical application of Rules of the Road	7.9%	64.6%	20.1%	7.3%
Performance under stress	4.3%	29.3%	50.6%	15.9%

Table 4 contains the overall measure of performance graded by the senior assessor.

There is no clear indication whether senior assessors calculated these results via a mathematical aggregation of the sub-category scores, or whether senior assessors used this category as a subjective measure of proficiency. If SWOS designed the assessment to measure JOOD proficiency, we might interpret the 82 percent pass rate as acceptable. The problem is that COs authorized these OODs to stand a senior watch dedicated to maintaining the safety of the ship and not simply a subordinate watch stander in training. As a result, the 18 percent of OODs assessed as "significant concerns" may cast doubt on the training programs that allowed these OODs to qualify. Of the 61 ships, 20 of them had at least one OOD that a senior assessor graded as "significant concerns." Eight of those 20 ships also had at least one OOD assessed as "complete with no concerns." This suggests that the training program may not be at fault, and that some other factor may cause this disparity. For example, maybe some OODs are qualified due to the lack of mechanisms to de-select officers that are low-quality matches with the SWO community. Alternatively, these low performing officers may not have had as many opportunities to stand watch as an OOD because they were competing for time within a crowded wardroom.

Table 4. Overall Measurements of OOD Proficiency for Pilot Study

	% Significant concerns	% Complete with concerns	% Complete no concerns
Overall performance	18%	66%	16%

Table 5 contains the results of the ROR and NSS exams that each OOD completed prior to commencing the simulated scenario. Fleet OODs must earn a 90 percent or higher on a similar ROR exam monthly. The NSS exam encompasses general knowledge an OOD should know as part of qualifying through the OOD PQS. Both exams had low pass rates.

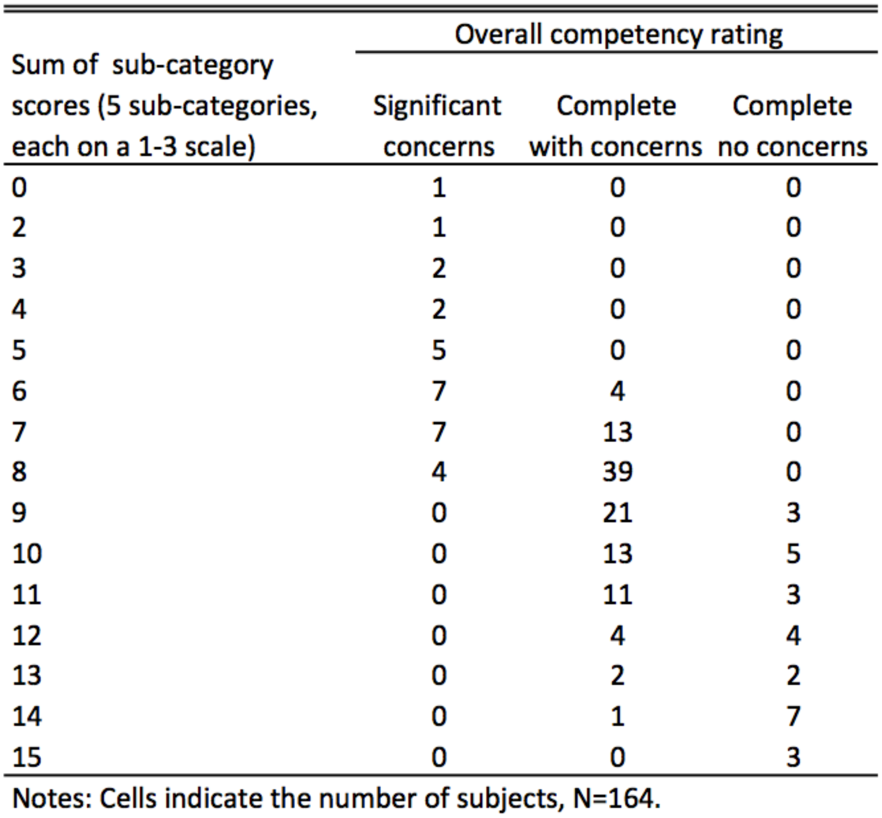
Table 5. Measurements of OOD Written Knowledge for Pilot Study Proficiency Assessment

	Mean	s.d.	min	max	% pass
Rules of the Road test	0.91	0.08	0.60	1	72.0%
NSS test	0.80	0.10	0.50	1	65.2%

Note: Passing RoR and NSS scores defined as 90% and 80%, respectively.

Table 6 contains a cross tabulation of the overall performance and the sum of the scores from the individual skill sub-categories. This table shows how the individual skills (BRM, BTM, leadership, application of ROR, and performance under stress) map into the senior assessor’s overall assessment of proficiency. For example, no OOD earned a score below nine that completed the scenario with “no concerns.” Similarly, no OOD earned a score above eight that completed the scenario with “significant concerns.” One challenge in understanding how assessments in the sub-categories map into the overall competency assessment of an OOD is understanding what is different about OODs that earned different overall proficiency ratings while having the same score on the summation of the sub-category skills. It is likely that each assessor used a different set of criteria to map sub-category performance into overall proficiency, and it is worth considering whether SWOS should use a more formal mapping to ensure consistency across assessors.

Table 6. Cross Tabulation Sub-category Scores and Overall Performance for Pilot Study Proficiency Assessment



Sum of sub-category scores (5 sub-categories, each on a 1-3 scale)	Overall competency rating: Significant concerns	Overall competency rating: Complete with concerns	Overall competency rating: Complete no concerns
0	1	0	0
2	1	0	0
3	2	0	0
4	2	0	0
5	5	0	0
6	7	4	0
7	7	13	0
8	4	39	0
9	0	21	3
10	0	13	5
11	0	11	3
12	0	4	4
13	0	2	2
14	0	1	7
15	0	0	3

Notes: Cells indicate the number of subjects, N=164.

B. MODEL SELECTION AND CONSIDERATIONS

We considered several functional forms and alternative methods for describing experience. The data included measures of both OOD and CONN experience. OOD experience did a better job of explaining the variation in performance than CONN experience. This may be because poorer performing OODs are more likely to spend greater amounts of time as a CONN or JOOD, which may come at the cost of experience as an OOD. Since the goal of the assessment was OOD proficiency, the logical choice was to focus on OOD experience as the functional input. We also considered whether to use the continuous variables derived from categorical responses or binary indicators of high versus low experience. While both functional forms show qualitatively similar relationships, we focus our analysis mainly on the binary indicators, as they are easier to interpret. One disadvantage of binary indicators of experience is that we must choose where the sample size division occurs. For example, we split the sample size for overall experience into officers that had more than 60 days on the watch bill as OOD and those who had had less than 60 days of experience. Sixty days of experience as an OOD does not provide an intuitive frame of reference for interpretation, but this sample-division point is a limit due to the categorical nature that SWOS used to collect survey data for this pilot study. In the future, examination of OOD logbooks will enable SWOS to collect the underlying, continuous data, such as hours, days, or months of experience and the number of special evolutions completed.

The proposed approach for future assessments, which utilizes continuous variables from logbooks instead of broad categories, enables us to divide future samples into a binary variable for analysis. This approach would allow us to estimate the effect of policies designed to improve OOD proficiency. For example, this proposal enables future analysts to assign officers to a binary variable that differentiates OODs that have acquired 300 hours of OOD experience from officers who have less than 300 hours of experience. We can then use this binary variable to determine whether this is the appropriate “break point” in experience by analyzing the differences between how the two groups of OODs perform. If we find that there is no difference between the performances of officers with more than 300 hours of OOD experience as compared to those with less than 300 hours of experience, then we may have evidence that suggests we need to reassess the “break point” of experience. Alternatively, if we find that there is a difference between how the two groups of OODs perform, then we should consider implementing a policy that establishes this level of experience as a minimum prior to OODs transferring to their second tour.

We could repeat this process with other appropriate policies as long as we have the foresight and appetite to modify the Mariner Skills Logbook to include the relevant data. If we do not deliberately think about the types of policies we want to assess, we may not have the right data to support such analysis under the current version of the logbook. These types of data-driven policy assessments may help SWOS identify the best training alternatives. With that knowledge, SWOS can advocate for additional resources based on the expected return as a function of proficiency.

Table 7 provides estimates for the preferred models where we regress two functional forms of proficiency on the inputs measured as skills, knowledge, currency, and experience. All four regressions include commissioning source. Homeport is included to serve as a way to capture fixed effects due to the assignment of graders.

The interpretation of these estimates is simpler than their continuous variable counterparts, which we included in the appendix as part of our robustness checks. For example, in Table 7 column 1, we find that an officer that reports more than 60 days of experience as an OOD is 15 percentage points more likely to pass the proficiency assessment simulation than an officer that reports less than 60 days of experience. Similarly, from Table 7 column 2, we find that an officer that passes the "performance under stress" portion of the skills assessment is 42 percentage points more likely to pass the proficiency assessment simulation than an officer that earns a failing score on "performance under stress." We interpret the "platform" variable similarly. For example, from Table 7 column 3, an officer that serves on an amphibious platform is five percentage points more likely to pass the proficiency assessment simulation than an officer from a non-amphibious platform, however, this estimate is not statistically significant.

The reader should not interpret non-statistically significant estimates as "no effect." Instead, we should interpret such estimates as "no evidence of an effect." The difference is that we might have a strong effect, for some officers, which is approximately equally as strong in the opposite direction for other officers. The result is that we do not see any evidence of an effect, which is not the same as "no effect." We can resolve these types of ambiguities through more accurate testing, reducing measurement error in the data we collect, or through collecting larger sample sizes, amongst other alternatives.

Table 7. Main Regression Results for Pilot Study Proficiency Assessment

	Complete with no concerns or Complete with concerns	Complete with no concerns or Complete with concerns	Complete with no concerns	Complete with no concerns
	(1)	(2)	(3)	(4)
Skills Assessment				
Passed bridge team management		0.10		-0.04
		(0.09)		(0.08)
Passed bridge resource management		0.07		0.16***
		(0.06)		(0.06)
Passed leadership		0.31***		-0.01
		-0.07		-0.06
Passed application of Rules of the Road		0.01		0.50***
		(0.05)		(0.05)
Passed performance under stress		0.42***		0.05
		(0.06)		(0.05)
Knowledge				
Passed Rules of the Road exam		0.09*		-0.02
		(0.05)		(0.05)
Passed navigation, seamanship, and ship handling exam		-0.02		-0.02
		(0.05)		(0.05)
Currency				
Attended BRM course within 7.5 months	0.07	0.03	0.07	0.06
	(0.08)	(0.06)	(0.07)	(0.06)
Experience				
More than 60 days experience as OOD	0.15*	0.10*	0.12	0.07
	(0.08)	(0.06)	(0.07)	(0.05)
More than 5 months on deployment	-0.00	0.03	0.08	0.08
	(0.07)	(0.05)	(0.06)	(0.05)
More than 16 months aboard ship	0.02	0.01	0.00	0.04
	(0.06)	(0.04)	(0.06)	(0.04)
Completed more than 2 pier work evolutions as OOD	0.06	-0.07	0.15**	0.06
	(0.08)	(0.06)	(0.08)	(0.06)

Completed any UNREP approach as OOD	-0.09	-0.05	-0.11	-0.10
	(0.09)	(0.07)	(0.09)	(0.07)
Has any TSS experience as OOD	0.12	0.05	0.21**	0.15**
	(0.11)	(0.08)	(0.10)	(0.08)
Has any experience in dense traffic/straits transits as OOD	-0.03	-0.05	-0.15*	-0.13*
	(0.10)	(0.07)	(0.09)	(0.07)
Has any experience anchoring as OOD	-0.11	0.08	-0.05	0.03
	(0.09)	(0.07)	(0.08)	(0.06)
Platform				
Serves on amphibious ship	0.06	0.06	0.05	0.02
	(0.09)	(0.07)	(0.08)	(0.06)
Commissioning source				
USNA	0.06	-0.07	0.02	-0.07
	(0.07)	(0.05)	(0.07)	(0.05)
OCS	0.13*	0.01	0.12*	0.02
	(0.08)	(0.06)	(0.07)	(0.05)
Home port				
Everett, WA	0.05	0.23**	-0.18	-0.05
	(0.14)	(0.11)	(0.13)	(0.10)
Mayport, FL	0.14	0.20**	0.10	0.13
	(0.13)	(0.09)	(0.12)	(0.09)
Pearl Harbor, HI	-0.05	0.14	0.02	-0.05
	(0.12)	(0.09)	(0.11)	(0.09)
San Diego, CA	-0.05	0.05	0.01	-0.02
	(0.10)	(0.08)	(0.09)	(0.08)
Sasebo, Japan	0.01	-0.12	0.28**	-0.00
	(0.14)	(0.11)	(0.13)	(0.11)
Yokosuka, Japan	0.08	0.04	0.14	-0.07
	(0.12)	(0.09)	(0.11)	(0.09)
Observations	164	164	164	164
R-squared	0.11	0.58	0.21	0.58
Mean of outcome	0.82	0.82	0.16	0.16

*** p<0.01, ** p<0.05, * p<0.1 N=164. Reference groups: Norfolk, VA; commissioning via ROTC. Non-amphibious ships include destroyer cruisers, and mine counter measure ships.

C. EFFECTS OF FUNCTIONAL INPUTS ON PROFICIENCY

In general, across the various models we analyzed, the assessed skills, experience, and knowledge are correlates of proficiency.[9] There does not appear to be evidence of an effect for the contribution that currency plays in predicting proficiency. Again, this is likely due to the poor proxy for this variable, because it is based solely on "time since last attending the BRM course." This Section breaks down each of the functional inputs by differentiating which variables are correlates of proficiency from those that do not appear to be correlates of proficiency. The results are specific to the pilot study that we analyzed. Future data sets will benefit from some of the recommendations that we advocate throughout our analysis. As a result, the correlations we outline may change with better data and assessment methodologies.

The functional input for skills is composed of BRM, BTM, leadership, performance under stress, and application of ROR. Future studies may provide clear evidence of a better way to define skills but results from the pilot study suggest that these factors proxy the information we intend to capture. Across the various robustness models, there is evidence that BRM, leadership, performance under stress, and the application of ROR are strong correlates of proficiency. BTM is not a correlate of proficiency under any permutation. This result suggests that we may need to define the boundaries between BRM and BTM more clearly, because some assessors may have unintentionally credited the wrong category during assessments. Alternatively, maybe we need to design a more objective way for assessing each OOD's BTM score.

The functional inputs for experience included the overall number of days of OOD watch, experience throughout various special evolutions, months on deployment, and months aboard ship. Overall experience as OOD, TSS experience, and pier-work experience are strong correlates of proficiency. This result is somewhat sensitive to whether the functional input of skills is included in the same regression, which is likely due to the nature of how experience may develop the skills a proficient OOD masters

---

[9] Tables 8-10 represent the robustness models used during our analysis and they are available in the appendix.

through repetition. There is no evidence of an effect that indicates months on deployment and months aboard ship are correlates of proficiency. These results are good in the sense that these factors are not easy to change at the individual level. For example, if an OOD is not proficient, we cannot give more deployment time or advance their time aboard ship through a policy change. Additionally, this result suggests that OODs on ships in extended-maintenance availabilities are not completely disadvantaged because they can utilize simulators or other cross-deck opportunities to mitigate ship-handling skill degradation. There is weak evidence that UNREP and anchoring evolutions are correlates of proficiency.[10] The lack of evidence of a relationship for these two special evolutions may be due to the difference in skills required as compared to navigating a moderately dense TSS.

The correlation between proficiency and experience in dense traffic is counter intuitive across all the robustness models. These estimates are negative correlates of proficiency in every permutation. This result is likely due to the relative differences in the subjective definition of “dense traffic” between OODs. For example, an OOD that is used to a low-traffic homeport might have a lower threshold for what defines “high traffic” than an OOD that earned experience in the most traffic-dense environments. As a result, OODs that claimed additional experience in dense-traffic environments performed worse, on average, when holding all other factors constant, than OODs that reported less dense-traffic experience. The development of more objective measures for traffic density may serve to better clarify the relationships between traffic-density experience and proficiency.

The functional input of knowledge is composed of written exams covering ROR and NSS subject matter. There is some evidence that ROR knowledge is a correlate of proficiency, which makes sense because understanding these rules are essential to applying

---

[10] It is worth considering whether we should pursue a more-holistic approach to assessing OOD proficiency. For example, a series of special evolutions, as opposed to the moderate-density TSS transit, may provide a clearer picture of the skills that are necessary for a proficient OOD. This type of assessment may come with some disadvantages. For example, if the series of assessments included an anchoring, UNREP, harbor transit, and a moderate-density TSS transit, the time required to assess each OOD may become prohibitive when scaled to support a sample size large enough to represent the entire fleet. Policymakers should consider whether this type of more-rigorous testing is worth the extra man-hour costs to assess. Additionally, there are opportunity costs associated with simulator unavailability, which ships could otherwise use to schedule training.

them when encountering other vessels and hazards to navigation. There is no evidence of an effect that indicates NSS knowledge is a correlate of proficiency throughout any of the robustness models. After examining the types of questions each OOD answered for this exam, this result might indicate that SWOS should tailor the exam specific to the knowledge critical for passing this type of simulator assessment.

Many of the questions for both written exams focused on broad knowledge, which may be representative of minimum qualification requirements. For example, one question on the NSS exam asked the OOD to select one of four multiple-choice answers that best described the force that would most affect a low-freeboard vessel with a deep draft. If the purpose of the written exams is to provide SWOS with a general idea of the knowledge retention and understanding an OOD has, these types of questions may be ideally suited. Alternatively, if SWOS wants to leverage these written exams to understand each OODs practical application of knowledge throughout the simulation, then we should tailor the exams to meet this objective. SWOS could accomplish this by revising all questions on both exams to focus on the decisions that OODs must make during the simulation. For example, instead of the question about forces on a low-freeboard vessel, SWOS could ask questions that we could objectively analyze during the scenario such as equipment failure reports to the CO. Since equipment failure occurs during the scenario, we could identify the relationship between their actions, when the failure occurs, and whether they answered the written exam question correctly.

The disadvantage to this change is that it might require a smaller bank of quasi-fixed questions on both the ROR and NSS exams, which might be easier for officers to anticipate. The additional value created through the analysis would likely offset the shallow depth of questions, because we could analyze exam questions as a functional input and the actual application of the knowledge in the simulation as the outcome. The results of this analysis could inform risk-management models by providing estimates of the value of each type of question as well as focus areas that schoolhouse training should emphasize. For example, if we assess a large sample of officers and find that OODs that pass low-visibility condition ROR questions are 2 percentage points more likely to apply this knowledge correctly in the simulator, then maybe we should allocate training resources to more significant concerns. Alternatively, if those same OODs are 50 percentage points more likely to apply the ROR correctly, then we may have clear evidence that the resources we are leveraging are appropriate and effective.

Within this data set, platform does not affect proficiency no matter which functional form is considered. This finding is important because it suggests that OODs are not at a disadvantage based on the type of ship where they earn their OOD qualification. This is likely because SWOS designed the scenario to test OODs on a similar platform to the one where they earned their experience.

Currency does not function as a significant correlate to proficiency in most of the models we analyzed. This is likely due to the low-quality nature of the only currency proxy available in the data set, which is the variable capturing the amount of time since last attending the BRM course. The fact that the currency variable is also the only variable associated with simulator experience, as opposed to actual experience on the bridge of a ship, may also explain this finding. A future data set, as proposed as part of the Mariner Skills Logbook recommendations in Chapter V, could provide a better opportunity to analyze the contribution that currency plays in predicting proficiency.

D. ROBUSTNESS CHECKS

Table 8's format (Appendix E) is identical to Table 7 except that we increased the requirement to pass the individual-skills assessments to match the highest available score of "exceeds standards." This change allows us to compare estimates for different definitions of "pass" for the functional input of skill toward understanding correlations with proficiency. A better alternative might have included lowering the passing threshold to include OODs assessed as "requires improvement" but there are insufficient observations for this model to provide meaningful results. In general, the results from Table 8 are qualitatively similar to our main results in Table 7.

Table 9's format (Appendix F) is identical to Table 7 except that we utilized the raw values from the assessment of the functional inputs of skills and knowledge instead of using the binary versions that we created for the main analysis. This change allows us to explore whether the correlation between proficiency and both knowledge and skills are real or are simply a function of the mathematical form we chose to utilize. Although the interpretation is different, the magnitude of the effect and the variables that are correlates of proficiency are qualitatively similar to our results from Table 7.

Table 10’s format (Appendix G) is identical to Table 9 except that we utilized the raw values given by the senior assessors to categorize the OODs assessed proficiency score on a scale from zero to two. For example, OODs assessed as “significant concerns” earn a zero, OODs assessed as “completed with concerns” earn a one, and OODs assessed as “completed no concerns” earn a two. We regressed these values on the same variables that we utilized in Table 9. Although the interpretation is different from both Tables 8 and 10, the magnitude of the effect and the variables that are correlates of proficiency are qualitatively similar to our results from previous estimates. The one exception is that the estimate for ship class is statistically significant, where it was previously insignificant. This result is likely due to a statistical anomaly because this is the first time that ship class has displayed any evidence of an effect over the course of several model revisions.

E. SUMMARY

Through our research, we have provided the first empirical estimates of the proficiency returns that knowledge, skills, experience and currency provide. We examined the currency component and noted that our proxy of time since attending the BRM course is insignificant, which suggests that we must get more accurate data in future studies, by leveraging the OOD logbooks instead of self-reported survey data. We also found that ship class is not a correlate of proficiency. We did find evidence that experience, skills, and knowledge are correlates of proficiency. Approximately 80 percent of the variation of proficiency is unaccounted for across many of the models we analyzed. This may be evidence that the pilot data set did not account for many of the issues that we have identified throughout our analysis. Despite this limitation, our research illustrates the value of the functional inputs to explain proficiency outcomes. The first edition of the Mariner Skills Logbook is a strong first step to provide data that is more accurate. Accurate data enables us to understand the correlates of proficiency.

V. CONTINUOUS DATA COLLECTION AND TESTING MECHANISMS

This chapter discusses the value of data collection and alternative collection mechanisms. Additionally, we describe suggestions for testing OODs, and explain how SWOS can use the data to improve the Fleet. We close this chapter with suggestions for SWOS to revise the Mariner Skills Logbook, which would provide analysts appropriate data to study additional training policies in the future.

A. VALUE OF FUTURE COLLECTION

SWOS issued the first Mariner Skills Logbooks in 2018, and they plan to issue one to every SWO.[11] Using data collected from the newly developed logbook for future studies provides a broader horizon for the types of questions we could answer during follow-on iterations of the proficiency assessment. The current data set is limited to answering questions about the determinants of proficiency and the percentage of OODs that were proficient in the pilot study. In the future, a more detailed data set could provide insight that allows us to analyze an individual's experience and prescribe a tailored training plan to address those shortfalls while at SWOS for advanced training. For example, an officer with high hours of OOD experience due to an extended deployment would likely have developed sufficient UNREP skills. These evolutions may come at the expense of experience in other evolutions like mooring to a buoy. In the future, a SWOS instructor could review the OOD's logbook data, recognize these experience gaps, and write a training plan for these observed experience deficits.

The next evolution of the potential for proficiency assessment data analysis is the ability to leverage experience markers when detailing prospective COs. In the future, prospective COs will have a full history of their ship-handling experiences documented in logbooks. Detailers will be able to leverage this data to ensure we optimally place our officers aboard ships. For example, a detailer may review an officer's logbook and discover that they

---

[11] An abbreviated version of the Surface Warfare Mariner Skills Logbook is included in Appendix H.

completed two division officer tours in Pearl Harbor with low levels of underway experience and subsequent tours as a Department Head in extensive shipyards in San Diego. If the detailer has a billet to fill for a DDG in Japan and a second billet for a DDG expecting to head into the yards, the experience levels found during the review could be critical to getting the right person into the job where they have the strongest chances of success.

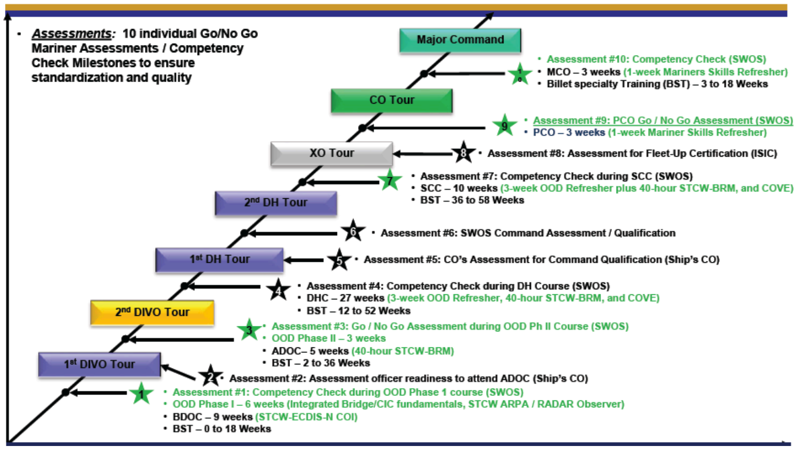
B. ALTERNATIVE COLLECTION METHODS

Several data collection mechanisms could prove viable in the long term for OOD experience-related data collection. This Section introduces some of these options from the most basic to the more complex. The initial decision for the long-term solution is whether officers should handwrite or electronically record them. If the handwritten option is the preferred solution, the next consideration is when to digitize them, so that SWOS can leverage this information for policy decisions and proficiency-check analysis. If an electronic format is the preferred solution, the next consideration is whether a new database is required, or whether we can leverage spreadsheets or an electronic survey option to transmit relevant information. Each potential solution comes with separate costs and benefits that affect the quality of the return on investment related to the collection, analysis, and dissemination of OOD proficiency research. We recommend that a follow-on analyst conduct research aimed at clearly articulating the various options, the costs, and the constraints associated with each, so that policymakers can make an informed decision.

1. Status Quo

SWOS has directed that individual OODs maintain sole possession of their logbooks until they transfer commands. OODs are required to fill out entries as they stand watch, carry data forward from page to page, and present them to their CO during quarterly reviews. On the occasion an officer transfers, the CO is required to summarize the data collected and forward it to Navy Personnel Command (PERS-41) in the format provided by COMNAVSURFPAC/COMNAVSURFLANTINST 1412.9 released on September 6, 2018. The instruction assigns PERS-41 as the organization responsible to document and maintain a digital file for each officer based on the CO's review. Then, periodically, PERS-41 shall provide the relevant data to SWOS for analysis. The instruction provides an enclosure as an example of what data COs are expected to send to PERS-41. This example lacks many critical pieces of information for future data analytics (for example, DoD identification number [DoD ID]), which would allow researchers to append demographic information to the data. Commissioning source, age, and years of service are some examples of pertinent information that researchers may need in future analysis and including DoD ID on the provided template could facilitate that process. Those variables are a small sub-set of the information that is not included in the current revision of the Mariner Skills Logbook.

The process outlined in COMNAVSURFPAC/COMNAVSURFLANTINST 1412.9 does not support the data necessary to analyze information acquired through the Assessments Plan SWO Competence Continuum. Figure 1 outlines the 10 mariner assessments that OODs must complete throughout their career. Under current guidance, PERS-41 does not provide the first report to SWOS until after the first assessment point at least, and the data might not even arrive until after the third assessment point. This process forces SWOS to collect the appropriate data manually during each assessment point. We recommend that a more coordinated process be implemented, which may necessitate a digital approach to recording and transferring experience-related data from logbooks to SWOS.



Source: CAPT S. Robertson, SWO mentoring session at Naval Postgraduate School, July 2, 2018.

Figure 1. SWO Competence Continuum Assessment Plan

Alternatively, SWOS could collect demographic and experience-related information through an electronic survey prior to proficiency checks. For example, when an officer receives orders for advanced training, SWOS could provide a web link that directs the OOD to a survey. The survey could include all the experience and demographic information necessary for future analysis. This process could serve as one low cost solution to resolve the data collection issues the community faces. An electronic survey could be a quarterly requirement for OODs to share experience data and it would alleviate the reporting delay and man-hour intensive process to transfer data from individual commands to SWOS via PERS-41, manually. For oversight, SWOS could develop a mechanism to forward the experience data to the officer's Senior Watch Officer to certify the accuracy, or SWOS could spot check the officer's logbook upon arrival for training.

2. Electronic Logbooks

A digital solution may bridge the gap where handwritten logbooks fall short. After OODs manually track experience throughout a tour, they must duplicate this effort on a survey at SWOS for it to be useful in a proficiency assessment. If officers digitized data initially, instead of handwriting them in a written logbook, SWOS would have instant access to a trove of data that they could use to monitor experience trends in the fleet. Such data could provide valuable insight at a moment's notice as new SWO personnel and training policies are considered. For example, if a policy is under review that reduces funding for contractors critical to operating the NSST centers, SWOS could leverage the digital data from the most recent 2,000 simulator hours utilized to determine what effect the policy might have on fleet readiness.

While SWOS would gain from the benefits of electronic logbooks, the individual, Senior Watch Officer, and ship's administrative office assumes the majority of the workload for recording, compiling, and manually transferring the written logbook data into an electronic or memorandum format. If the record does not arrive in a digital format, then PERS-41 must develop a solution that converts a typed memorandum into a spreadsheet format or pass the responsibility on to SWOS. This may still prove too cumbersome for SWOS to utilize in the short term. Thousands of records would flow in; require some level of quality control, compilation into a usable format, storage, and protection due to the sensitive nature that may be inherent for the records. An electronic solution would prove more useful if the record began life digitally, but this has its own shortfalls to overcome such as formatting, version control, and privacy controls. Leveraging database technology could simplify the process as outlined in the following Section.

3. Database Technology

A database solution could reduce the numerous steps required to record watch-standing experience into a physical logbook initially, digitize it, and transmit it through appropriate channels to SWOS for analysis. Instead of handwriting pertinent watch-related details, access to a database on the bridge could allow the OOD's Common Access Card (CAC) to create a new record at watch turnover and record system time and other appropriate unclassified parameters. A networked solution could append ship's AIS data, weather message data, or other significant details to the OOD's database entry. Granular data like traffic density from AIS could dramatically improve the snapshot for each entry by depicting a more accurate representation of the quality and quantity of experience earned during a watch as measured by traffic density.

Existing database options could suffice in the short term to facilitate the most basic functions of this process until the community identifies a permanent solution. For example, at-sea commands develop watch bills in the Relational Administrative Data Management (RADM) application. Similar to how watch station PQS requirements function as a pre-requisite to stand a watch, overall experience and currency standards that are set by fleet decision makers could ensure compliance with relevant policies. The ship's Senior Watch Officer could manage these pre-requisites, which OODs would record at watch turnover. This process would mimic entering training in RADM, which is something most division officers are already familiar.

RADM already communicates with the Fleet Management and Planning System (FLTMPS) by replicating data off-ship during scheduled uploads. This design could convert a multi-step process into a condensed version. This process would still rely on the individual entering the data at watch turnover.

The problem with RADM as a solution is that it is cumbersome to operate, and it lacks funding to implement useful changes that would support a database approach to share OOD experience-related data. The Surface Community should consider one of three options to identify a long-term solution. If there is sufficient funding, we should develop a ground-up approach that focuses on several key aspects. This solution must present ease of use for ships underway, reliable replication for data transfer to shore, and a solid framework that captures the factors that are correlates of proficiency. Additionally, it is important that we incorporate the ability for the community to change data fields as we learn more about the functional inputs to proficiency. If there is insufficient funding for an ideal approach, we should consider a system that is already in operational use by another community. If there is insufficient funding for this type of an approach, we should invest money into a new generation of RADM that is dual-purposed to track the traditional training information as well as the experience-related information that is critical to proficiency assessment research.

C. SUGGESTIONS FOR TESTING OODS

First-tour officers were the only OODs tested during the pilot study. These officers represent one link in the larger picture of the OOD training continuum. Future proficiency-assessment opportunities could create critical data sets to answer tough questions that we are currently unable to answer due to the limited scope of the data collected in the pilot study. Four main groups of officers could each bring different pieces of the proficiency picture into focus if we choose to utilize them appropriately. These groups are comprised of first tour, second tour, Department Head, prospective Executive officers (XO) and prospective COs.

First-tour officers could provide an excellent snapshot of how effective the Basic Division Officer Course (BDOC) curriculum is by testing a random sample before and after each convening. Similarly, we could examine the effectiveness of the OOD Phase I curriculum, using first-tour officers. This analysis could ensure that SWOS recognizes relevant trends in sufficient time to continue to send a top-quality product to the fleet. Additionally, they represent one sub-set of the total ship-handling population at any given time. Since these officers must return to SWOS for advanced training, we can leverage this opportunity to maintain the pulse of the fleet's proficiency as they transition to commence their second tour. Downward trends may alert policymakers that an operational pause may be necessary to prevent a catastrophic accident. Second-tour officers provide the other main sub-set of the total OOD population of the fleet. The assessment prior to the commencement and at the conclusion of ADOC could provide evidence of the quality of instruction and a representative sample of the quality of officer that is reporting to the fleet to train prospective OODs. Prospective Department Head officer testing provides critical information on skill degradation resulting from time on shore duty. Finally, testing XO/COs could explain what influences the long-term mastery of maritime skills and where the focus should be starting from initial-accession training. Once the logbook data has caught up with these senior leaders, analysts could finally understand the mechanisms through which overall experience and currency shape individual proficiency.

The type of testing to which these officers are exposed could answer different types of questions in the future as well. For example, proficiency testing may provide a snapshot of how each officer acts on the bridge of a ship in a generic situation. In contrast, a test that puts OODs in extremis might paint a better picture of their gut reactions, which should be the culmination of their understanding of how their ship responds to controllable and uncontrollable forces. This is especially true for prospective COs, because their OOD may not request assistance in sufficient time for the CO to direct the OOD to take the appropriate action in accordance with the ROR and with due regard to good seamanship. In-extremis extraction assessments may provide the best predictor of how a prospective CO may react under the highest-pressure situations involving ship-handling skills. This type of analysis may challenge assumptions that decision makers fail to consider during curriculum development and revision.

D. HOW TESTING AND EXPERIENCE DATA CAN BE USED TO IMPROVE THE FLEET

Testing and the relationship to experience data can improve the Fleet's qualification process, training prescriptions, officer rotation guidance, retention of top performers, detailing of appropriate officers based on proficiency, and selection of the best officers for

Command at sea. An optimally designed data collection mechanism ensures we have the right information to answer questions about the minimum experience necessary to qualify a watch station. This knowledge allows us to apply minimum standards to currency requirements or as a training prescription for experience deficits. Expanding on this research, we can analyze the point where the SWO community gets the best return on investment from an OOD, so we do not rotate them too early, which could deprive them of valuable at-sea experience. OOD proficiency testing builds on this evaluation by identifying the top performers that the community should incentivize toward retention.

E. SUGGESTIONS FOR MARINER SKILLS LOGBOOK REVISION

The reader should recall that SWOS designed logbooks for OODs to record data related to watch-standing experience while underway and in simulators. These logbooks are similar to the traditional logbooks that pilots use to record flight hours. Logbook data provides a significant advantage over the estimated data that SWOS collected during the 164 assessments for the pilot study, because the logbook solution reduces the measurement error inherent to the survey estimates. Measurement error tends to bias the estimated effect toward no effect in regression analysis. As a result, a factor that is a correlate of proficiency, in actuality, may not demonstrate the true relationship during data analysis. By utilizing logbook data exclusively for future data sets, we avoid the measurement error of estimated data and we will observe more of the true relationships between experience and proficiency.

In addition to the variables that the current Mariner Skills Logbooks direct OODs to record, several other variables could be useful in order to understand the correlates of good ship handlers. We categorize these variables, state where they could be included in the logbook, and describe why they are important to collect from research and policy points of view. If new elements are added, the notes section of the logbook could be expanded and serve as a reference for OODs to resolve any ambiguities. These revisions could be critical to future-data collection because the logbooks should be the sole source of experience-related data for proficiency assessments instead of the subjective surveys, which rely on individual memory. Including these changes in a future logbook revision should offer more insights during future data analysis.

1. Demographics

SWOS could add the following variables at the beginning of the logbook, after "Ships Employment," and prior to the Table of Contents.

Commissioning Source (USNA/ROTC/OCS): An OOD's ship-handling experience may differ based on commissioning source. For example, USNA midshipmen may train on Yard Patrol craft (YP), while there is no such opportunity for OCS students. Additionally, some ROTC units have ship-handling simulators, while other ROTC units do not have this simulation capability, a distinction that SWOS should collect. It is important for us to understand whether these factors are correlates of proficiency. In fact, the proficiency-check data that SWOS collected between in the pilot study suggests that commissioning source has statistically significant explanatory powers.

Age: The trucking industry crash prediction models (Monaco & Williams, 2000) and (Cantor et al., 2010) suggest that there is a strong statistical relationship between driver age and the likelihood of collisions. The evidence suggests that maturity is a significant component in the quantity and quality of individual risk management decisions while operating heavy equipment. There may be evidence of maturity in the data collected in the pilot study, for example, age is likely a correlate with prior service. These factors may explain some of the differences in performance observed due to commissioning source.

Years of Service and Prior Enlisted (Y/N): This variable builds on the maturity component concept. Years of service and whether an OOD has prior enlisted service may explain more of the relative differences between OOD performance across commissioning sources and may offer evidence that any differences observed are more accurately representative of differing maturity levels. For example, an aviation student with three years of service that enters the SWO pipeline may outperform a new SWO candidate with zero years of service despite having the same amount of experience on the bridge of a ship purely due to differences in maturity, brain development, and skill transfer.

Gender (Male/Female): While gender as a predictor of proficiency is often considered controversial, the evidence from the trucking industry suggests that male drivers are more likely to exhibit risky driving behaviors than their female counterparts (Cantor et al., 2010). This variable could confirm or disprove whether similar patterns exist in the maritime environment and inform future operational risk management models.

Home Port: The proficiency check data from the pilot study suggests that homeport has statistically significant explanatory powers in seven of seven outcome variables assessed. Since there is no mathematical way to separate the effects of homeport from the effect of the senior assessors in the current data set, future data collection should include homeport information.

2. TAB A: Individual Watch Log

SWOS could add the following variables in the Individual Watch Log.

Traffic Density: Currently, this variable is logged as "LOW MED HIGH" with an appropriately placed "x" in the watch-standing log in TAB A. This method provides an opportunity for too much ambiguity. A definition would clarify some of the ambiguity, for example, "LOW" could signify averaging one or less contact report per hour. Similar definitions scaled for "MED" and "HIGH" would clarify this logbook entry for bridge-watch standers. Alternatively, an objective data point extracted from the ship's AIS would provide better fidelity to facilitate uniform standards for OOD logbook entries. This capability does not exist with our present equipment configuration. The ideal situation for a future AIS configuration would provide the OOD with an objective measurement of the number of vessels encountered within CO contact reporting requirements. Analysts could convert this information into a "traffic-density function" by incorporating a time element for the entire watch or broken down by sub-sections of the watch. This type of traffic-density metric would allow analysts to determine precisely how traffic density affects an OOD's development and whether there is an optimal-traffic density to build experience through training.

Weather: Currently, the logbook collects this variable through manually written notes, for example "Environmentals: Low visibility, Heavy weather/seas." This variable provides limited capability to leverage it in future data analysis because of its subjectivity. The Beaufort scale for example, would provide a more objective measurement than information collected in the notes section. One constraint is the lack of details from other components of poor weather conditions like fog or rain that such a scale includes. SWOS could resolve this by adding a binary variable for "Fog/Hazy," where the OOD would mark "yes" if any portion of the watch were complicated with either factor. Similarly, another variable for "rain" could be included for the OOD to mark "yes" if any portion of the watch is complicated with the presence of rain. SWOS could provide a metric to remove ambiguity, which could assist the OOD with answering this question. For example, a note could instruct the OOD to mark "yes" for sustained conditions lasting longer than fifteen-to-thirty minutes. If experience operating in traffic-dense environments produces more proficient ship handlers during busy straits transits, then mariners experienced in deteriorated weather conditions should outperform their fair-weather experienced counterparts during poor environmental conditions.

Low-Light Conditions: Research suggests that low-light conditions exist in a disproportionately large number of maritime collisions and groundings (Macrae, 2009). An appropriately worded note could instruct the OOD to log this variable as "yes" if more than fifteen minutes of the watch warrant the use of navigation lights, excluding situations where they are used for fog.

Hours of Sleep since Last Watch or within 24 Hours: The literature for maritime (Macrae, 2009) and trucking collisions (Cantor et al., 2010) suggests that upwards of 87 percent of industrial crashes are the product of driver-related factors. These researchers cite fatigue, misjudgment, and carelessness as significant predictors and contributors to driver-related factors, and this variable could serve as a proxy to measure these effects.

CIC Support to Bridge Grade: Although this variable is subjective in nature, it could inform the CO of trends and illustrate the effectiveness of communication for different CIC and Bridge watch teams. This would be a scale from one to 10 and would describe the quality and timeliness of requested support from CIC.

Near Misses: Research suggests that a large proportion of accidents and collisions follow near misses or smaller-scale incidents shortly before major accidents. Researchers in the trucking industry (Cantor et al., 2010), (Lueck & Murray, 2005), (Lueck & Murray, 2011), and (Monaco & Williams, 2000) observed that prior driver or vehicle violations are statistically significant factors that increase the likelihood of a crash occurrence. Near misses could serve as a significant proxy for risk of collision prediction. OODs that determine a situation existed where they narrowly avoided a larger disaster due to luck, a last-minute change of course, or speed that resulted in a "near miss" instead, could annotate this in the logbook. As opposed to a negative experience, this record might facilitate opportunities to develop lessons learned and train other watch teams to improve proficiency.

3. TAB B: Special Evolutions Log

We recommend maintaining this section in its current form.

4. TAB C: Simulator Training Log

SWOS could add the following variables in the Simulator Training Log.

Instructor Provided Numerical Grade: This variable would be on a scale from one to 10 and would capture the overall performance demonstrated by the OOD. For example, if an OOD spent four hours in the simulator and divided that time evenly between pier work and transiting a TSS, OODs could document two separate entries, and the instructor would provide a separate grade for the overall performance of each evolution. This could serve as an outcome variable for researchers and provide quantitative feedback to the OOD in addition to the qualitative information provided by the instructor to focus on future training and ultimately the next proficiency check.

Weather: OODs would document this variable the same way as we recommended for TAB A.

Simulated Location: This variable would describe where the simulated event took place and serve as a control variable to describe the locations where an OOD receives training. For example, training conducted during open-ocean steaming with no traffic should yield less benefit than a challenging coastal location where the OOD may have to consider effects like strong currents or shallow water in addition to traffic.

Traffic Density: OODs could document this variable the same way as we recommended for TAB A.

Page Totals/Carried Forward/New Totals: This formatting could resemble TAB A to facilitate the transfer to the CO Quarterly Endorsement Log.

5. TAB D: CO Quarterly Endorsement

SWOS could add the following variables in the CO Quarterly Endorsement.

Numerical Grade: This variable would be on a scale from one to 10 and would allow the CO to convey his/her perception of the overall performance demonstrated by the OODs. Additionally, CO could omit this section if they had insufficient opportunity to observe the officers performance as a CONN or JOOD. It could serve as a mini-proficiency check to let officers know where they stand to facilitate productive conversations regarding how they can improve during CO quarterly reviews.

We also recommend including the following variables in the CO Quarterly Endorsement Log to consolidate pertinent information for the CO to provide feedback to the OOD during the quarterly review. These variables may also shed light on the tone the OOD establishes with his watch team, identify some of the OODs strengths/weaknesses, and may provide a visual representation of critical information concerning the amount of risk the CO assumes for each OOD standing watch. For example, if a CO is concerned about an OOD's performance and notes a mediocre RADAR Proficiency Grade, low ROR exam grades, and poor communication with Combat Information Center watch standers, it may prompt the CO to consider substituting a stronger JOOD to mitigate some of the risks observed. We recommend adding the following variables:

- Quarterly RADAR Proficiency Assessment Grade (as observed during quarterly assessment aboard ship)
- Monthly ROR exam grade (or quarterly average)

- Average hours of rest (self-reported by OOD)
- Average Combat Information Center-Bridge support grade
- Near misses

6. Proposed TAB E: Record of Assessments

This additional section of the log enables OODs to record all ship-handling related assessments in one location. For example, this section could include all SWOS proficiency checks, simulator training evaluations, RADAR proficiency assessments, ROR exams, etc. This record could facilitate collecting information for the CO Quarterly Endorsement and provide a one-stop shop to confirm that the OOD meets pertinent watch-standing requirements quickly, beyond the OOD Letter to stand OOD underway. We recommend that SWOS include the following information at a minimum:

- Date
- Type of exam
- Evaluation/Grade (X out of X, Passed, or 91%)

VI. CONCLUSION

The Navy must establish a procedure that objectively measures OOD navigation and ship-handling skills. SWOS tackled this problem through a pilot study where they assessed OODs through a moderate-density simulator, written exam, and collected experience-related data. Our research focused on each of these processes to establish what metrics are correlates of proficiency from the data available. Additionally, we leveraged aviation and trucking industry research to consider additional factors that may explain the variation in OOD performance. We found evidence that skills, knowledge, and overall experience are correlates of proficiency. We found no evidence of an effect that indicates that currency is a correlate of proficiency, which is experience gained in the most recent three-month period; however, the poor proxy for currency in this data set indicates a reexamination is required. Demographics may provide additional information about proficiency. Since this information is limited to ship class and commissioning source, we default to the literature, which suggests that age, gender, and years of service information are likely keys to understanding the maturity-related components of proficiency.

The current data set in the pilot study represents first tour OODs and is not representative of the entire stock of ship handlers in the fleet. Future research should study (1) the entire cross-section of underway OODs and (2) OODs in a shipyard environment. The entire cross-section of underway OODs include both first and second-tour OODs instead of the first-tour OODs assessed in the pilot study of 164 assessments. This cross-sectional sample would provide a clearer picture of the state of the underway fleet's proficiency. OODs that reported extensive periods in a shipyard environment would allow researchers to study how navigation and ship-handling skills decay over time.

In the future, Mariner Skills Logbook data will become the standard for OOD-related experience records. As a result, entire cohorts of Department Heads and prospective COs will have a career's worth of experience documented. This end-state will provide additional opportunities for researchers to investigate the relationship of long-term mariners' skill mastery and allow them to recommend policies that leverage these insights toward further development and retention of our best ship handlers.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. OOD COMPETENCY CHECKLIST

_____1. O3 Assessor shall ensure participant signs Non-Disclosure Agreement.

_____2. O3 Assessor shall ensure participant completes Experience Survey (two-sided)

_____3. O3 Assessor shall collect all material prior to the start of the Exam.

_____4. During the test, the O3 JOOD will configure the Bridge Equipment IAW the approved script.

_____5. O3 Assessor will proctor a 40 question RoR and NSS Exam and ensure the following:

  (a) Inform the participant: "Cheating is an unacceptable behavior which is contrary to the Navy Core Values and Sound Shipboard Operating Principles. A participant caught cheating will be awarded a test score of zero and recommend for disciplinary actions."
  (b) Collect the exam.

_____6. O3 JOOD Assessor will give the approved scene setter brief to the participant.

_____7. O5 Assessor will give a brief to the participant.

_____8. O3 Assessor will walk the participant into the Bridge for a familiarization session.

_____9. During the familiarization session, the O3 JOOD and O5 Assessor will grade the participant on the VMS/APRA Set-Up changes.

_____10. Following the Set-Up Changes, the O5 Assessor and NSST Operator will put the Problem in Start.

_____11. The O5 Assessor and O3 JOOD, will execute the problem IAW the script and the rubric.

_____12. During the scenario, the O3 Assessor will grade the Exam and record the results. He or she will then record the results of the survey and input the survey results in the excel document provided.

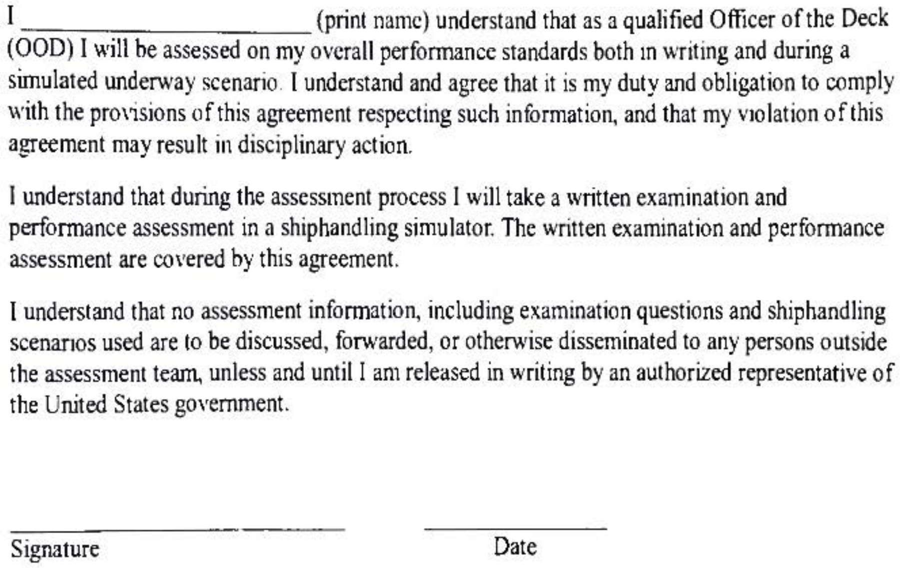
_____13. Upon FINEX, the O5 Assessor and O3 JOOD will give a quick de-brief to the participant to address any areas of concerns. They will reinforce the Non-Disclosure Agreement.

_____14. While de-brief is being conducted, the O3 Assessor will prepare for the next participant.

_____15. The O5 Assessor will fill out the PDF Drop Down based on his/her assessment of the individual and email it to the Commanding Officer of the participant's ship.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. NON-DISCLOSURE AGREEMENT



I ________________________ (print name) understand that as a qualified Officer of the Deck (OOD) I will be assessed on my overall performance standards both in writing and during a simulated underway scenario. I understand and agree that it is my duty and obligation to comply with the provisions of this agreement respecting such information, and that my violation of this agreement may result in disciplinary action.

I understand that during the assessment process I will take a written examination and performance assessment in a shiphandling simulator. The written examination and performance assessment are covered by this agreement.

I understand that no assessment information, including examination questions and shiphandling scenarios used are to be discussed, forwarded, or otherwise disseminated to any persons outside the assessment team, unless and until I am released in writing by an authorized representative of the United States government.

____________________________ ________________

Signature Date

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. OOD EXPERIENCE SURVEY

Name: ________________________ **Ship**: ____________________

Commissioning Source:

○ Naval Academy
○ ROTC: School: ______________________________
○ OCS:

Fill in the circle that closely applies:

How many months have you been assigned to your current ship?

0-6	7-12	13-18	>18
○	○	○	○

Number of months you were on deployment?

0-3	4-6	7-11	12 -17	>18
○	○	○	○	○

Number of months it took you to qualify as an OOD?

0-6	7-12	13-18	>18
○	○	○	○

If SWO Qualified, the number of months it took you to qualify?

Not Qualified	1-6	7-12	13-18	>18
○	○	○	○	○

I feel comfortable operating the RADAR/ARPA to its fullest extent in congested waterways?

Yes	No
○	○

When using the VMS system on your ship, do you understand from where the system gets its inputs from and how the system determines your position?

Yes	No
○	○

Do you feel confident when making Bridge to Bridge radio calls?

Yes	No
○	○

When was the last formal BRM training you received from the Navigation, Seamanship, and Ship handling Training (NSST) center?

Never	1–3 months	4–11 months	12–18 month	> 19 months
○	○	○	○	○

Watch station specific questions

Conning Officer: The number of evolutions that you have conducted.

Getting underway / Mooring to a pier:

None	1-2	3-4	5–6	> 6
○	○	○	○	○

Underway replenishment (approach):

None	1-2	3-4	5–6	> 6
○	○	○	○	○

Anchoring:

None	1-2	3-4	5–6	> 6
○	○	○	○	○

Transit a designated Traffic Separation Scheme (TSS):

None	1-2	3-4	5–6	> 6
○	○	○	○	○

Number of straits transits conducted:

None	1-4	5-8	9–12	> 12
○	○	○	○	○

Approximate number of days underway where you were assigned Conning Officer on the watch bill:

< 20	21-99	100-200	>200
○	○	○	○

As the Officer of the Deck (OOD): The number of evolutions that you have conducted.

Getting underway / Mooring to a pier:

None	1-2	3-4	5–6	> 6
○	○	○	○	○

Underway replenishment (approach):

None	1-2	3-4	5–6	> 6
○	○	○	○	○

Anchoring:

None	1-2	3-4	5–6	> 6
○	○	○	○	○

Transit a designated Traffic Separation Scheme (TSS):

None	1-2	3-4	5–6	> 6
○	○	○	○	○

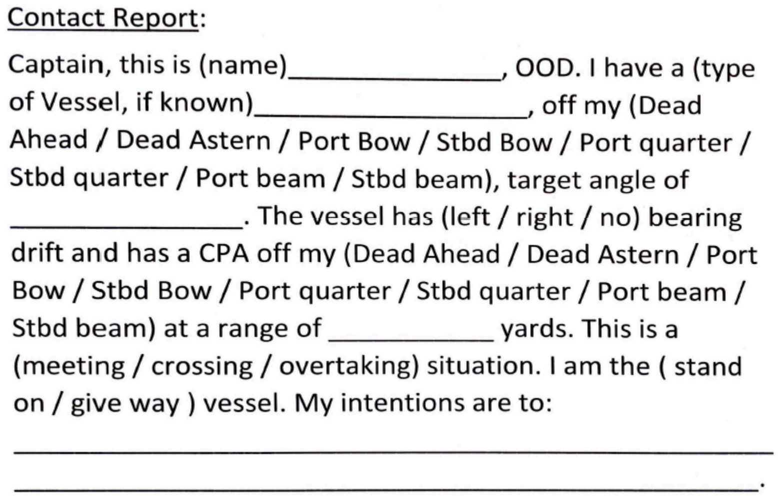
Number of straits transits conducted:

None	1-4	5-8	9–12	> 12
○	○	○	○	○

Approximate number of days underway where you were assigned OOD on the watch bill:

< 20	21-99	100-200	>200
○	○	○	○

APPENDIX D. CONTACT REPORT TEMPLATE



Contact Report:

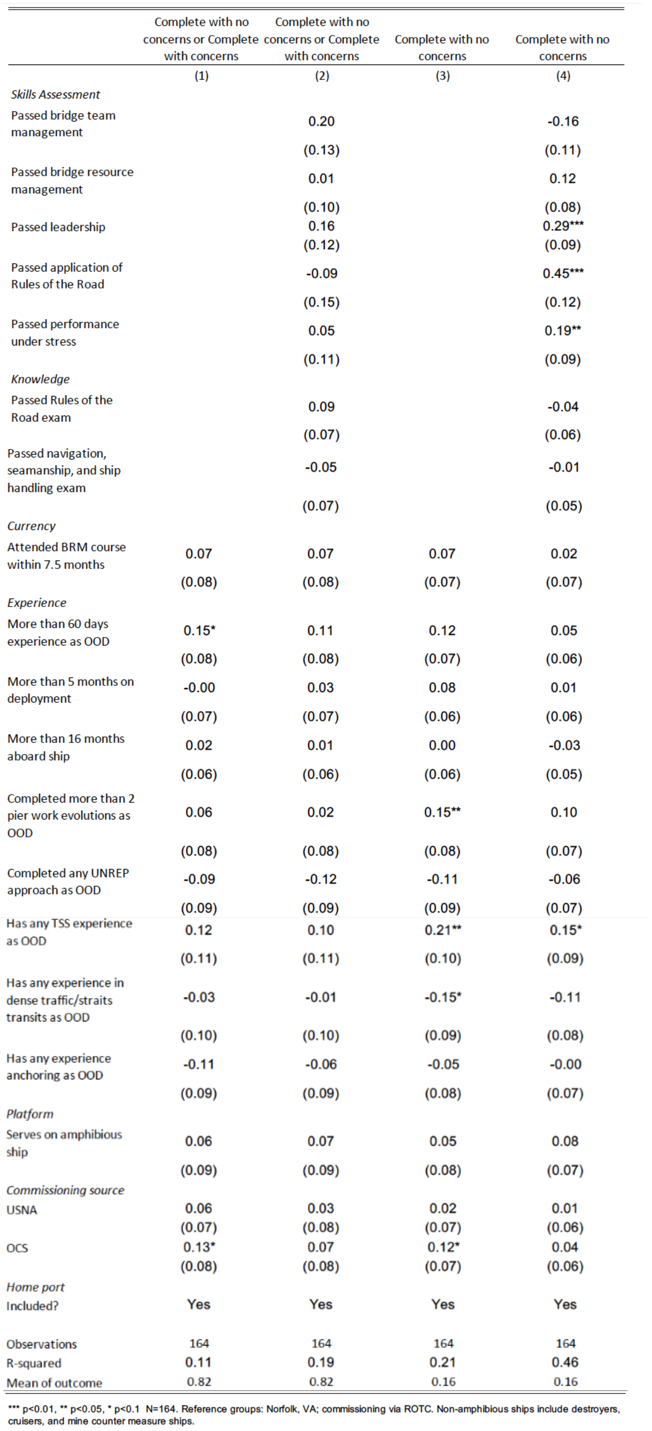
Captain, this is (name)_______________, OOD. I have a (type of Vessel, if known)___________________, off my (Dead Ahead / Dead Astern / Port Bow / Stbd Bow / Port quarter / Stbd quarter / Port beam / Stbd beam), target angle of ________________. The vessel has (left / right / no) bearing drift and has a CPA off my (Dead Ahead / Dead Astern / Port Bow / Stbd Bow / Port quarter / Stbd quarter / Port beam / Stbd beam) at a range of ____________ yards. This is a (meeting / crossing / overtaking) situation. I am the (stand on / give way) vessel. My intentions are to:

____

____.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. ROBUSTNESS TABLE 8



	Complete with no concerns or Complete with concerns	Complete with no concerns or Complete with concerns	Complete with no concerns	Complete with no concerns
	(1)	(2)	(3)	(4)
Skills Assessment				
Passed bridge team management		0.20		-0.16
		(0.13)		(0.11)
Passed bridge resource management		0.01		0.12
		(0.10)		(0.08)
Passed leadership		0.16		0.29***
		(0.12)		(0.09)
Passed application of Rules of the Road		-0.09		0.45***
		(0.15)		(0.12)
Passed performance under stress		0.05		0.19**
		(0.11)		(0.09)
Knowledge				
Passed Rules of the Road exam		0.09		-0.04
		(0.07)		(0.06)
Passed navigation, seamanship, and ship handling exam		-0.05		-0.01
		(0.07)		(0.05)
Currency				
Attended BRM course within 7.5 months	0.07	0.07	0.07	0.02
	(0.08)	(0.08)	(0.07)	(0.07)
Experience				
More than 60 days experience as OOD	0.15*	0.11	0.12	0.05
	(0.08)	(0.08)	(0.07)	(0.06)
More than 5 months on deployment	-0.00	0.03	0.08	0.01
	(0.07)	(0.07)	(0.06)	(0.06)
More than 16 months aboard ship	0.02	0.01	0.00	-0.03
	(0.06)	(0.06)	(0.06)	(0.05)
Completed more than 2 pier work evolutions as OOD	0.06	0.02	0.15**	0.10
	(0.08)	(0.08)	(0.08)	(0.07)
Completed any UNREP approach as OOD	-0.09	-0.12	-0.11	-0.06
	(0.09)	(0.09)	(0.09)	(0.07)
Has any TSS experience as OOD	0.12	0.10	0.21**	0.15*
	(0.11)	(0.11)	(0.10)	(0.09)
Has any experience in dense traffic/straits transits as OOD	-0.03	-0.01	-0.15*	-0.11
	(0.10)	(0.10)	(0.09)	(0.08)
Has any experience anchoring as OOD	-0.11	-0.06	-0.05	-0.00
	(0.09)	(0.09)	(0.08)	(0.07)
Platform				
Serves on amphibious ship	0.06	0.07	0.05	0.08
	(0.09)	(0.09)	(0.08)	(0.07)
Commissioning source				
USNA	0.06	0.03	0.02	0.01
	(0.07)	(0.08)	(0.07)	(0.06)
OCS	0.13*	0.07	0.12*	0.04
	(0.08)	(0.08)	(0.07)	(0.06)
Home port				
Included?	Yes	Yes	Yes	Yes
Observations	164	164	164	164
R-squared	0.11	0.19	0.21	0.46
Mean of outcome	0.82	0.82	0.16	0.16

*** p<0.01, ** p<0.05, * p<0.1 N=164. Reference groups: Norfolk, VA; commissioning via ROTC. Non-amphibious ships include destroyers, cruisers, and mine counter measure ships.

THIS PAGE INTENTIONALLY LEFT BLANK

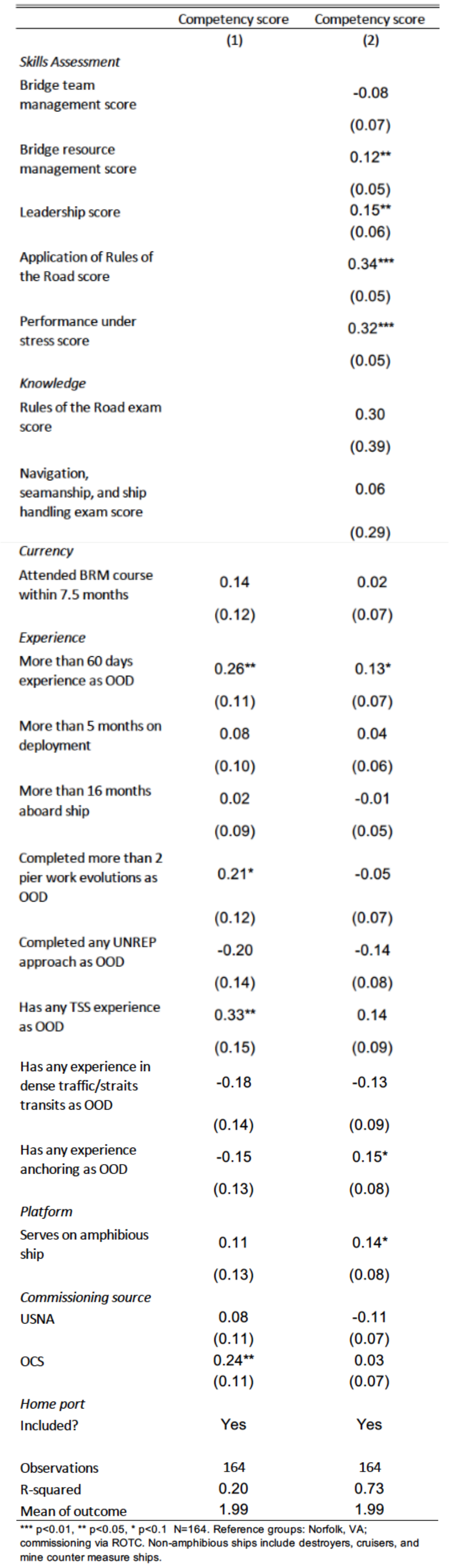
APPENDIX F. ROBUSTNESS TABLE 9

	Complete with no concerns or Complete with concerns	Complete with no concerns or Complete with concerns	Complete with no concerns	Complete with no concerns
	(1)	(2)	(3)	(4)
Skills Assessment				
Bridge team management score		0.01		-0.10
		(0.06)		(0.06)
Bridge resource management score		0.02		0.10**
		(0.04)		(0.04)
Leadership score		0.14**		0.01
		(0.05)		(0.05)
Application of Rules of the Road score		0.05		0.29***
		(0.04)		(0.04)
Performance under stress score		0.25***		0.06
		(0.04)		(0.04)
Knowledge				
Rules of the Road exam score		0.29		0.01
		(0.34)		(0.33)
Navigation, seamanship, and ship handling exam score		0.17		-0.11
		(0.25)		(0.24)
Currency				
Attended BRM course within 7.5 months	0.07	-0.00	0.07	0.02
	(0.08)	(0.06)	(0.07)	(0.06)
Experience				
More than 60 days experience as OOD	0.15*	0.08	0.12	0.04
	(0.08)	(0.06)	(0.07)	(0.06)
More than 5 months on deployment	-0.00	-0.00	0.08	0.05
	(0.07)	(0.05)	(0.06)	(0.05)
More than 16 months aboard ship	0.02	-0.01	0.00	-0.01
	(0.06)	(0.05)	(0.06)	(0.05)
Completed more than 2 pier work evolutions as OOD	0.06	-0.07	0.15**	0.02
	(0.08)	(0.06)	(0.08)	(0.06)
Completed any UNREP approach as OOD	-0.09	-0.09	-0.11	-0.05
	(0.09)	(0.07)	(0.09)	(0.07)
Has any TSS experience as OOD	0.12	0.02	0.21**	0.12
	(0.11)	(0.08)	(0.10)	(0.08)
Has any experience in dense traffic/straits transits as OOD	-0.03	-0.01	-0.15*	-0.12*
	(0.10)	(0.08)	(0.09)	(0.07)
Has any experience anchoring as OOD	-0.11	0.09	-0.05	0.06
	(0.09)	(0.07)	(0.08)	(0.07)
Platform				
Serves on amphibious ship	0.06	0.07	0.05	0.07
	(0.09)	(0.07)	(0.08)	(0.07)
Commissioning source				
USNA	0.06	-0.05	0.02	-0.06
	(0.07)	(0.06)	(0.07)	(0.06)
OCS	0.13*	-0.01	0.12*	0.03
	(0.08)	(0.06)	(0.07)	(0.06)
Home port				
Included?	Yes	Yes	Yes	Yes
Observations	164	164	164	164
R-squared	0.11	0.52	0.21	0.53
Mean of outcome	0.82	0.82	0.16	0.16

*** $p<0.01$, ** $p<0.05$, * $p<0.1$ N=164. Reference groups: Norfolk, VA; commissioning via ROTC. Non-amphibious ships include destroyers, cruisers, and mine counter measure ships.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX G. ROBUSTNESS TABLE 10



	Competency score	Competency score
	(1)	(2)
Skills Assessment		
Bridge team management score		-0.08
		(0.07)
Bridge resource management score		0.12**
		(0.05)
Leadership score		0.15**
		(0.06)
Application of Rules of the Road score		0.34***
		(0.05)
Performance under stress score		0.32***
		(0.05)
Knowledge		
Rules of the Road exam score		0.30
		(0.39)
Navigation, seamanship, and ship handling exam score		0.06
		(0.29)
Currency		
Attended BRM course within 7.5 months	0.14	0.02
	(0.12)	(0.07)
Experience		
More than 60 days experience as OOD	0.26**	0.13*
	(0.11)	(0.07)
More than 5 months on deployment	0.08	0.04
	(0.10)	(0.06)
More than 16 months aboard ship	0.02	-0.01
	(0.09)	(0.05)
Completed more than 2 pier work evolutions as OOD	0.21*	-0.05
	(0.12)	(0.07)
Completed any UNREP approach as OOD	-0.20	-0.14
	(0.14)	(0.08)
Has any TSS experience as OOD	0.33**	0.14
	(0.15)	(0.09)
Has any experience in dense traffic/straits transits as OOD	-0.18	-0.13
	(0.14)	(0.09)
Has any experience anchoring as OOD	-0.15	0.15*
	(0.13)	(0.08)
Platform		
Serves on amphibious ship	0.11	0.14*
	(0.13)	(0.08)
Commissioning source		
USNA	0.08	-0.11
	(0.11)	(0.07)
OCS	0.24**	0.03
	(0.11)	(0.07)
Home port		
Included?	Yes	Yes
Observations	164	164
R-squared	0.20	0.73
Mean of outcome	1.99	1.99

*** p<0.01, ** p<0.05, * p<0.1 N=164. Reference groups: Norfolk, VA; commissioning via ROTC. Non-amphibious ships include destroyers, cruisers, and mine counter measure ships.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX H. MARINER SKILLS LOGBOOK EXAMPLE

Surface Warfare Mariner
Skills Logbook

NAME: ________________________________

Ship assigned: Date Reported Date Detached

Crossdeck Ships: Date Reported Date Detached

Ships Employment

Table of Contents

Quick Reference Abbreviations Guide

Mark 'X'; as appropriate

D	Daytime
N	Nighttime

Special Evolution Codes

Code	Evolution		
S&A	Sea and Anchor Transit	ANC	Anchoring Evolution
RAS	Replenishment at Sea	PLG	Planeguard
PUW	Pier (Getting Underway)	HRT	Harbor Transit
PMO	Pier (Mooring)	MOB	Man Overboard
HDT	High Density Traffic	MTB	Moor to a buoy
DIV	Division Tactical Maneuverings	TBT	Tow Be Towed

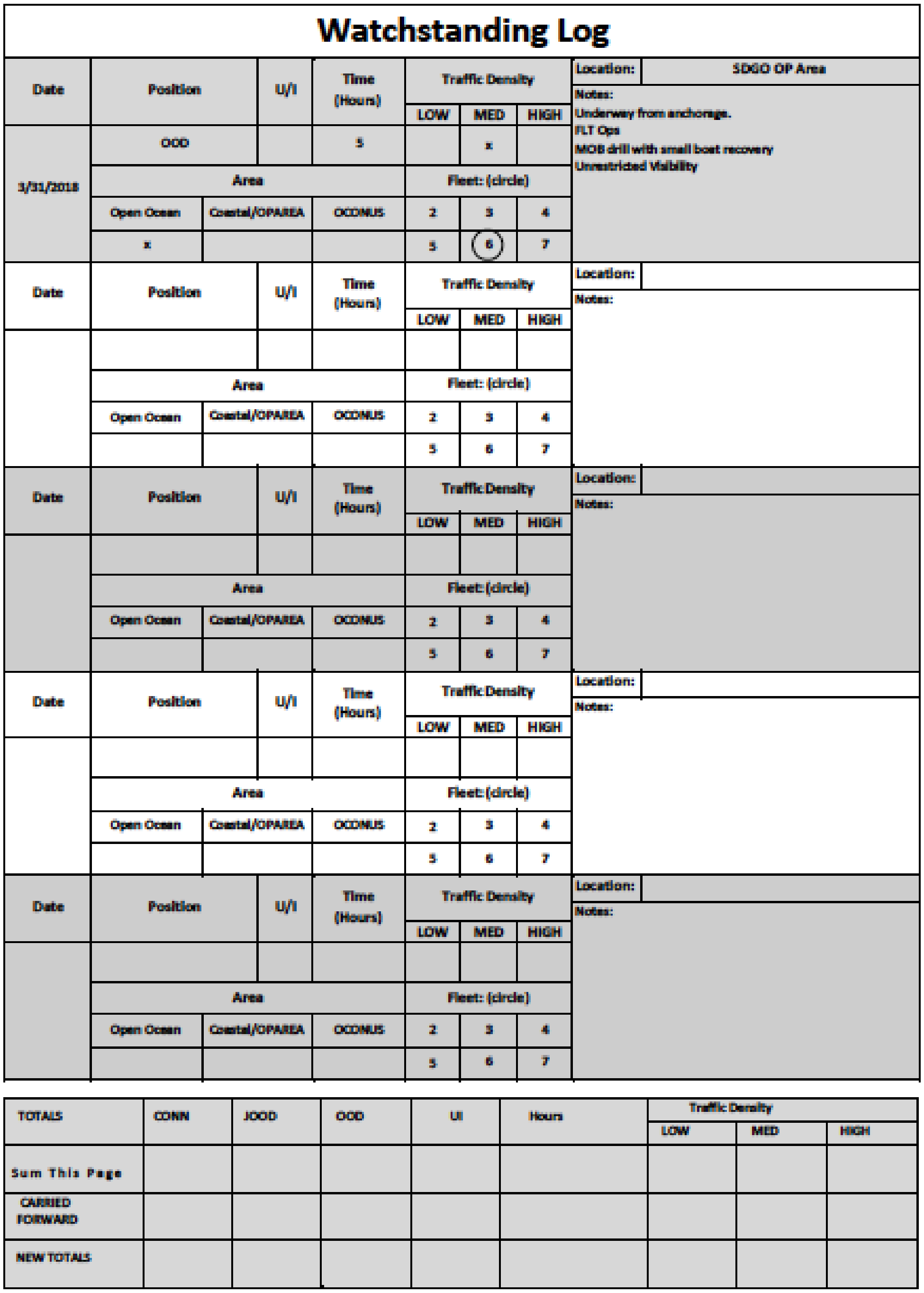
U/W Written Notes examples:

Environmentals:
Low Visibility
Heavy weather/seas

Ship considerations:
Shipboard drills conducted (Ex: GQ, Small Boat Ops, Engineering Drills)
Non Routine Engineering or Combat Systems Configurations
Night Steam Box
Exercises: (Ex: COMPTUEX, PASSEX)
Training Phase

TAB A

Individual Watch Log



Watchstanding Log

<table>
<tr><td rowspan="2">Date</td><td rowspan="2" colspan="2">Position</td><td rowspan="2">U/I</td><td rowspan="2">Time (Hours)</td><td colspan="3">Traffic Density</td><td>Location:</td><td>SDGO OP Area</td></tr>
<tr><td>LOW</td><td>MED</td><td>HIGH</td><td colspan="2" rowspan="5">Notes:
Underway from anchorage.
FLT Ops
MOB drill with small boat recovery
Unrestricted Visibility</td></tr>
<tr><td rowspan="4">3/31/2018</td><td colspan="2">OOD</td><td></td><td>5</td><td></td><td>x</td><td></td></tr>
<tr><td colspan="4">Area</td><td colspan="3">Fleet: (circle)</td></tr>
<tr><td>Open Ocean</td><td colspan="2">Coastal/OPAREA</td><td>OCONUS</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>x</td><td colspan="2"></td><td></td><td>5</td><td>[✓ 6]</td><td>7</td></tr>
</table>

<table>
<tr><td rowspan="2">Date</td><td rowspan="2" colspan="2">Position</td><td rowspan="2">U/I</td><td rowspan="2">Time (Hours)</td><td colspan="3">Traffic Density</td><td>Location:</td><td></td></tr>
<tr><td>LOW</td><td>MED</td><td>HIGH</td><td colspan="2" rowspan="5">Notes:</td></tr>
<tr><td rowspan="4"></td><td colspan="2"></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="4">Area</td><td colspan="3">Fleet: (circle)</td></tr>
<tr><td>Open Ocean</td><td colspan="2">Coastal/OPAREA</td><td>OCONUS</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td></td><td colspan="2"></td><td></td><td>5</td><td>6</td><td>7</td></tr>
</table>

<table>
<tr><td rowspan="2">Date</td><td rowspan="2" colspan="2">Position</td><td rowspan="2">U/I</td><td rowspan="2">Time (Hours)</td><td colspan="3">Traffic Density</td><td>Location:</td><td></td></tr>
<tr><td>LOW</td><td>MED</td><td>HIGH</td><td colspan="2" rowspan="5">Notes:</td></tr>
<tr><td rowspan="4"></td><td colspan="2"></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="4">Area</td><td colspan="3">Fleet: (circle)</td></tr>
<tr><td>Open Ocean</td><td colspan="2">Coastal/OPAREA</td><td>OCONUS</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td></td><td colspan="2"></td><td></td><td>5</td><td>6</td><td>7</td></tr>
</table>

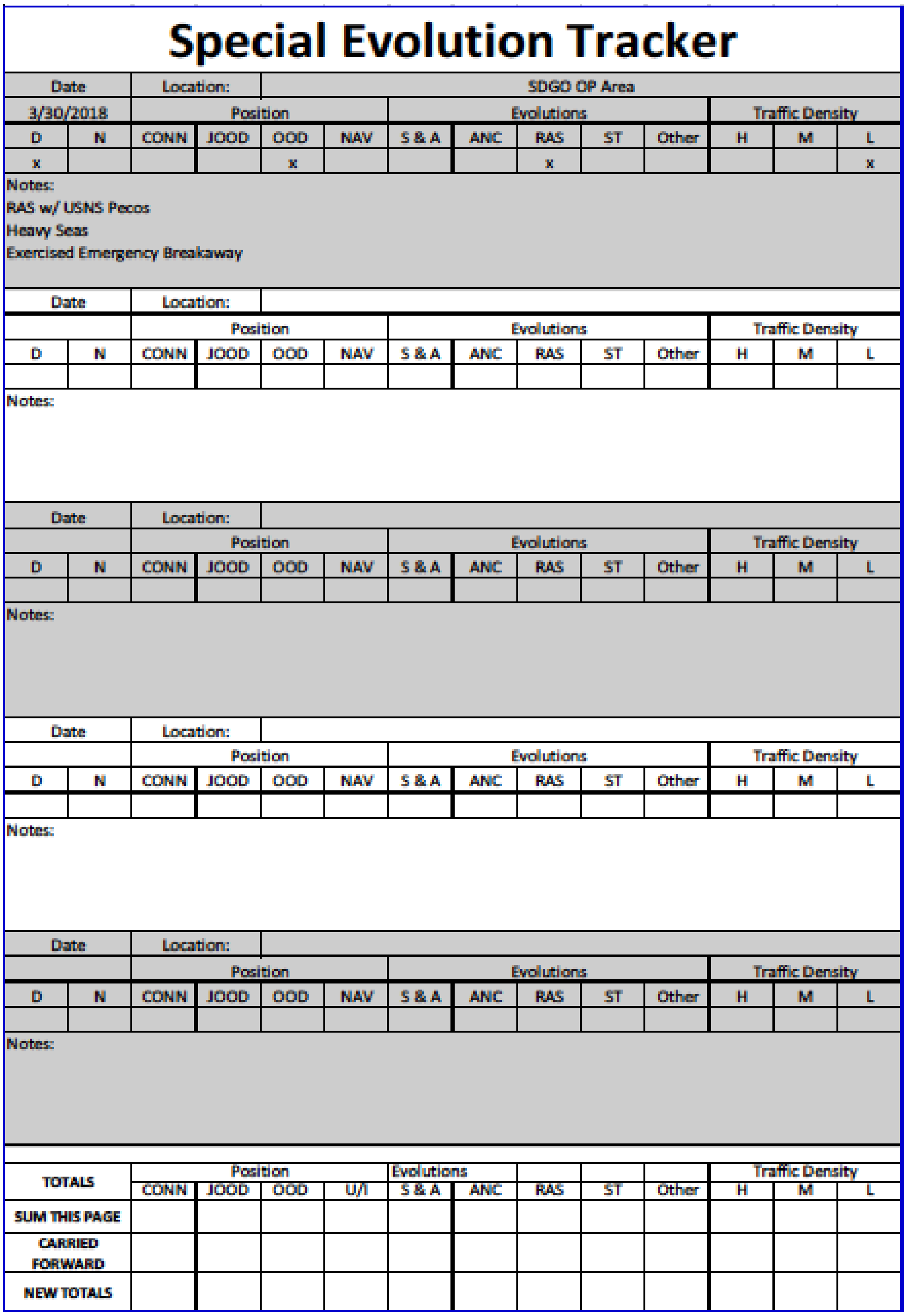
<table>
<tr><td rowspan="2">Date</td><td rowspan="2" colspan="2">Position</td><td rowspan="2">U/I</td><td rowspan="2">Time (Hours)</td><td colspan="3">Traffic Density</td><td>Location:</td><td></td></tr>
<tr><td>LOW</td><td>MED</td><td>HIGH</td><td colspan="2" rowspan="5">Notes:</td></tr>
<tr><td rowspan="4"></td><td colspan="2"></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="4">Area</td><td colspan="3">Fleet: (circle)</td></tr>
<tr><td>Open Ocean</td><td colspan="2">Coastal/OPAREA</td><td>OCONUS</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td></td><td colspan="2"></td><td></td><td>5</td><td>6</td><td>7</td></tr>
</table>

<table>
<tr><td rowspan="2">Date</td><td rowspan="2" colspan="2">Position</td><td rowspan="2">U/I</td><td rowspan="2">Time (Hours)</td><td colspan="3">Traffic Density</td><td>Location:</td><td></td></tr>
<tr><td>LOW</td><td>MED</td><td>HIGH</td><td colspan="2" rowspan="5">Notes:</td></tr>
<tr><td rowspan="4"></td><td colspan="2"></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="4">Area</td><td colspan="3">Fleet: (circle)</td></tr>
<tr><td>Open Ocean</td><td colspan="2">Coastal/OPAREA</td><td>OCONUS</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td></td><td colspan="2"></td><td></td><td>5</td><td>6</td><td>7</td></tr>
</table>

TOTALS	CONN	JOOD	OOD	UI	Hours	Traffic Density		
						LOW	MED	HIGH
Sum This Page								
CARRIED FORWARD								
NEW TOTALS								

TAB B

Special Evolutions Log



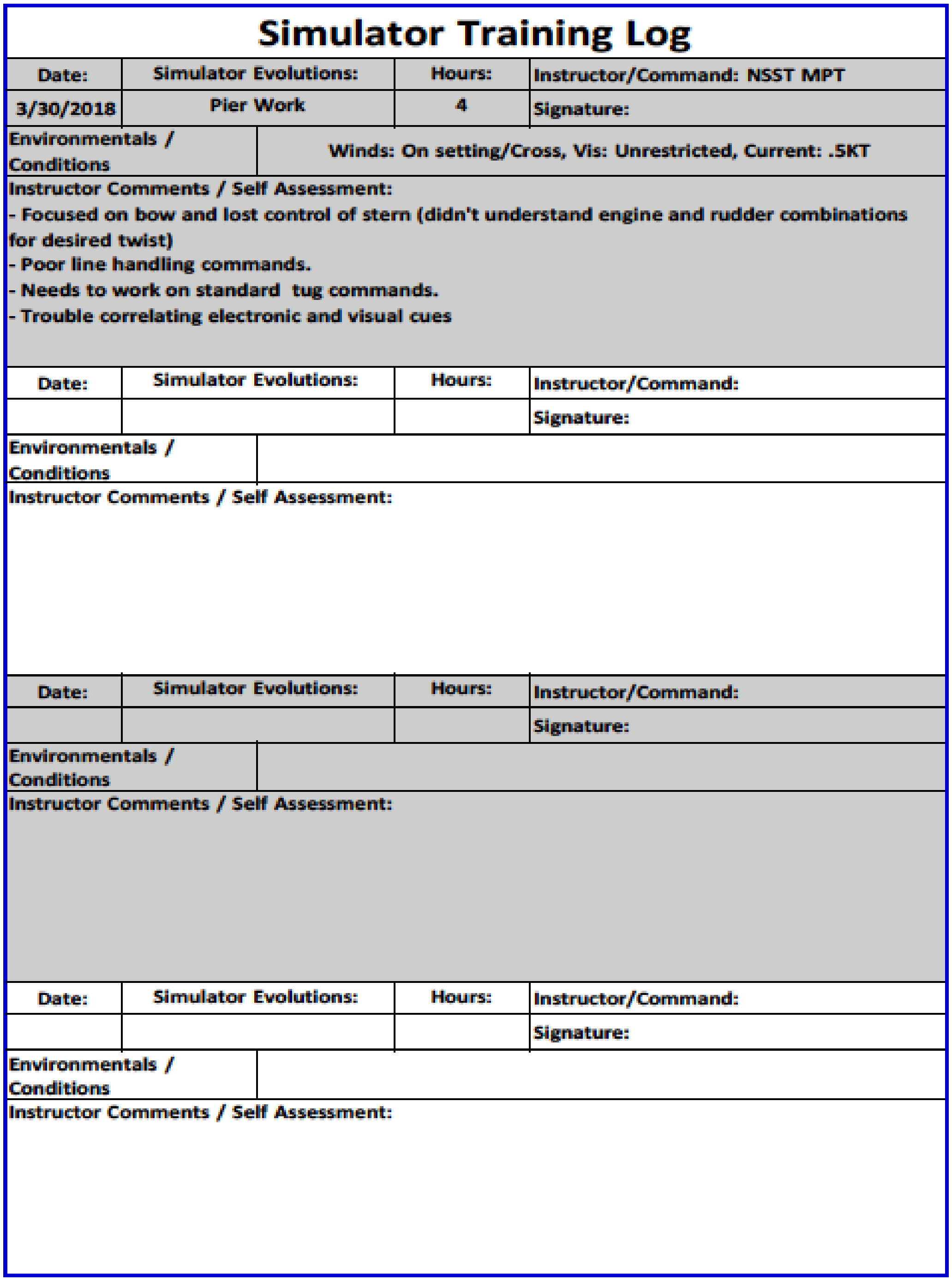
Special Evolution Tracker

Date		Location:		SDGO OP Area									
3/30/2018		Position				Evolutions					Traffic Density		
D	N	CONN	JOOD	OOD	NAV	S & A	ANC	RAS	ST	Other	H	M	L
x				x				x					x
Notes: RAS w/ USNS Pecos Heavy Seas Exercised Emergency Breakaway													
Date		Location:											
		Position				Evolutions					Traffic Density		
D	N	CONN	JOOD	OOD	NAV	S & A	ANC	RAS	ST	Other	H	M	L
Notes:													
Date		Location:											
		Position				Evolutions					Traffic Density		
D	N	CONN	JOOD	OOD	NAV	S & A	ANC	RAS	ST	Other	H	M	L
Notes:													
Date		Location:											
		Position				Evolutions					Traffic Density		
D	N	CONN	JOOD	OOD	NAV	S & A	ANC	RAS	ST	Other	H	M	L
Notes:													
Date		Location:											
		Position				Evolutions					Traffic Density		
D	N	CONN	JOOD	OOD	NAV	S & A	ANC	RAS	ST	Other	H	M	L
Notes:													

TOTALS	Position				Evolutions					Traffic Density		
	CONN	JOOD	OOD	U/I	S & A	ANC	RAS	ST	Other	H	M	L
SUM THIS PAGE												
CARRIED FORWARD												
NEW TOTALS												

TAB C

Simulator Training Log



Simulator Training Log

Date:	Simulator Evolutions:	Hours:	Instructor/Command: NSST MPT
3/30/2018	Pier Work	4	Signature:

Environmentals / Conditions: Winds: On setting/Cross, Vis: Unrestricted, Current: .5KT

Instructor Comments / Self Assessment:

- Focused on bow and lost control of stern (didn't understand engine and rudder combinations for desired twist)
- Poor line handling commands.
- Needs to work on standard tug commands.
- Trouble correlating electronic and visual cues

Date:	Simulator Evolutions:	Hours:	Instructor/Command:
			Signature:

Environmentals / Conditions:

Instructor Comments / Self Assessment:

Date:	Simulator Evolutions:	Hours:	Instructor/Command:
			Signature:

Environmentals / Conditions:

Instructor Comments / Self Assessment:

Date:	Simulator Evolutions:	Hours:	Instructor/Command:
			Signature:

Environmentals / Conditions:

Instructor Comments / Self Assessment:

TAB D

CO Quarterly Endorsement

Commanding Officer Quarterly Endorsement Page

I certify completion of ______Underway hours, _____Simulator hours, _____ Sea and Anchor Detail, _____ Replenishments at Sea, ____ Anchoring Evolutions, and ____ Straits Transits during the period from ___________ to___________.

____

Commanding Officer

Comments:

I certify completion of ______Underway hours, _____Simulator hours, _____ Sea and Anchor Detail, _____ Replenishments at Sea, ____ Anchoring Evolutions, and ____ Straits Transits during the period from ___________ to___________.

____

Commanding Officer

Comments:

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Blower, D., & Campbell, K. (2002). The large truck crash causation study. Retrieved from https://www.researchgate.net/publication/30817521_The_large_truck_crash_causation_study

Cantor, D. E., Corsi, T. M., Grimm, C. M., & Ozpolat, K. (2010). A driver focused truck crash prediction model. *Transportation Research: Part E: Logistics and Transportation Review, 46*(5), 683–692.

Cordial, B. (2017, March). Too many SWOs per ship. *Proceedings, 143*(3). Retrieved from https://www.usni.org/node/90091

Dahlström, N. (2008). Pilot training in our time: Use of flight training devices and simulators. *Aviation, 12*(1), 22–27. https://doi.org/10.3846/1648-7788.2008.12.22-27

Department of the Navy. (2016). *NATOPS general flight and operating instructions manual* (CNAF M-3710.7). Washington, DC: Author. Retrieved from https://www.public.navy.mil/airfor/vaw120/Documents/CNAF%20M-3710.7_WEB.PDF

Department of the Navy. (2017). Comprehensive review of recent surface force incidents. Retrieved from https://www.public.navy.mil/usff/Pages/usff-comprehensive-review.aspx

Diehl, A. (1991). Human performance and systems safety considerations in aviation mishaps. *International Journal of Aviation Psychology, 1*(2), 97–106.

Federal Motor Carrier Safety Administration. (2006). *Report to Congress on the large truck crash causation study* (Report No. MC-R/MC-RRA). Retrieved from https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/ltccs-2006.pdf

Flight Deck Friend. (2012). Yearly Training Requirements for Airline Pilots. Retrieved December 19, 2018, from https://www.flightdeckfriend.com/yearly-training-requirements-for-airline-pilots/

Flight Review, 14 C.F.R. § 61.56. (2013). Retrieved from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgFAR.nsf/0/AB8B895828E3309486257C24005B6A14?OpenDocument

Goode, N., Salmon, P., & Lenne, M. (2012). Simulation-based driver and vehicle crew training: Applications, efficacy and future directions. *Applied Ergonomics, 44*(3), 435–444. https://doi.org/10.1016/j.apergo.2012.10.007

Government Accountability Office. (2001). *Improvements are needed in 5-ton truck driver training and supervision* (Army Training No. GAO-01-436) Washington, DC: Government Accountability Office.

Hallmark, S., Hsu, Y.-Y., Maze, T., McDonald, T., & Fitzsimmons, E. (2009). Investigating factors contributing to large truck lane departure crashes using the FMCSA's LTCCS database. Retrieved from https://intrans.iastate.edu/app/uploads/2018/03/lg_truck_lane_departure.pdf

Hammon, C. P., & Horowitz, S. A. (1990). Flying hours and aircrew performance. Retrieved from Defense Technical Information Center website: https://doi.org/10.21236/ADA228582

Hansen, J., & Oster, C. (1997). *Taking flight: Education and training for aviation careers.* Washington, DC: National Academy Press.

Jacobs, J. W., Prince, C., Hays, R. T., & Salas, E. (1990). A meta-analysis of the flight simulator training research. Fort Belvoir, VA: Defense Technical Information Center. https://doi.org/10.21236/ADA228733

Judy, A. D. (2018). *A study of flight simulation training time, aircraft training time, and pilot competence as measured by the naval standard score* (Doctoral dissertation). Retrieved from https://firescholars.seu.edu/coe/22

Kavanagh, J. (2005). *Determinants of productivity for military personnel: A review of findings on the contribution of experience, training, and aptitude to military performance*. Santa Monica, CA: RAND.

Kubiszewski, R. (1994). *Standardized driver's licensing policy—Yes or no?* (Executive Research Project S42 No. NDU-ICAF-93-842). Washington, DC: National Defense University.

LaGrone, S. (2018). New surface warfare officer career path stresses fundamentals; More training before first ship, more time at sea. *U.S. Naval Institute News.* Retrieved from https://news.usni.org/2018/06/28/new-career-path-surface-warfare-officers-stresses-fundamentals-training-first-ship-time-sea

Lampton, D., Kraemer, R., Kolasinski, E., & Knerr, B. (1995). *An investigation of simulator sickness in a tank driver trainer* (Research Report No. 1684). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lueck, M.D., & Murray, D. C. (2005). *Predicting truck crash involvement: Developing a commercial driver behavior-based model and recommended countermeasures*. Retrieved from ATRI website: https://truckingresearch.org/2005/10/18/predicting-truck-crash-involvement-2005/#.XIA-ES3MwWo

Lueck, M. D., & Murray, D. C. (2011). Predicting truck crash involvement: Linking driver behaviors to crash probability. *Journal of Transportation Law, Logistics, and Policy, 78*(2), 109–128. Retrieved from https://web.b.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10785906&AN=63698384&h=OXo5sXcyuTIk4AeOealgJLkoNkYd3bhKWTe%2bmIj5HB45SlRsSG%2bNwyfh4NLn3jjDqChGWvx5nzUzCLurC4nZcA%3d%3d&crl=f&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authtype%3dcrawler%26jrnl%3d10785906%26AN%3d63698384

Macrae, C. (2009). *Human factors at sea: Common patterns of error in groundings and collisions. Maritime Policy and Management, 36*(1), 21–38.

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology, 6*(1), 1–20. https://doi.org/10.1207/s15327108ijap0601_1

Mayhew, S., & Peterson, J. (1993). *A benchmark of tractor trailer operator training between the United States Army's 37th transportation command and a selected civilian industry leader* (Master's thesis). Air Force Institute of Technology Air University, Wright-Patterson Air Force Base, OH.

Mcdade, M. (1986). *Driver trainer training developments study for M113 family of vehicles*. Retrieved from Defense Technical Information Center website: https://apps.dtic.mil/dtic/tr/fulltext/u2/a174218.pdf

McFadden, K. L. (1997). Predicting pilot-error incidents of U.S. airline pilots using logistic regression. *Applied Ergonomics, 28*(3), 209–212. https://doi.org/10.1016/S0003-6870(96)00062-2

McLean, G. M. T. (2012). Flight simulation: An expensive placebo? *Defense Science & Technology Organization.* Retrieved from http://www.simulationaustralasia.com/files/upload/pdf/research/Flight_Simulation_An_Expensive_Placebo_G_McLean.pdf

Mejza, M., Barnard, R., Corsi, T., & Keane T. (2003). Driver management practices of motor carriers with high compliance and safety performance. *Transportation Journal, 42*(4), 16–29. Retrieved from http://libproxy.nps.edu/login?url=https://search-proquest-com.libproxy.nps.edu/docview/204587504?accountid=12702

Monaco, K., & Williams, E. (2000). Assessing the determinants of safety in the trucking industry. *Journal of Transportation and Statistics, 3*(1).

Oskarsson, P., Nählinder, S., & Svensson, E. (2010). A meta study of transfer of training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(28), 2422–2426. https://doi.org/10.1177/154193121005402813

Reis, P. (2000). *Determinants of flight training performance: An analysis of the impact of undergraduate academic background* (Master's thesis). Naval Postgraduate School, Monterey, CA.

Schank, J. F., Thie, H. J, Graf III, C. M., Beel, J, & Sollinger, J. M. (2002). *Finding the right balance: simulator and live training for Navy units* (No. MR-1441-NAVY). Santa Monica, CA. Retrieved from https://www.rand.org/pubs/monograph_reports/MR1441.html

Seltzer, L., & McBrayer, J. (1971). *A study of the effect of time on the instrument skill of the private and commercial pilot* (Report No. FAA-DS-70-12). Cahokia, IL: Parks College of Aeronautical Technology of Saint Louis University.

Siem, F. M., & Murray, M. W. (1997). *Personality factors affecting pilot combat performance: A preliminary investigation.* Fort Belvoir, VA: Defense Technical Information Center. https://doi.org/10.21236/ADA459823

Yardley, R., Thie, H., Schank, J., Galegher, J., & Riposo, J. (2003). Use of simulation for training in the U.S. Navy surface force (Report No. MR-1770-NAVY). Santa Monica, CA. Retrieved from https://www.rand.org/pubs/monograph_reports/MR1770.html

Yardley, R., Thie, H., Schank, J., Galegher, J., & Riposo, J. (2005). *Can under way training be reduced? The use of simulation for training in the U.S. Navy surface force* (Report No. RB-7567-NAVY). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/research_briefs/RB7567.html

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California