NPS-GSDM-20-003

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

MEASUREMENT AND ANALYSIS OF OFFICER OF THE DECK COMPETENCY: NEW FINDINGS FOR FY19-20

by
Jesse Cunha & Vincent Salazar

May 2020

Prepared for: Surface Warfare Officer Schools Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From-To)*
May 19, 2020	Technical Report	10/1/20 – 5/19/20

4. TITLE AND SUBTITLE
Measurement and Analysis of Officer of the Deck Competency: New Findings for FY19-20

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jesse Cunha
Vincent Salazar

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
Graduate School of Defense Management, Naval Postgraduate School

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Surface Warfare Officer Schools Commands Command (SWSC)

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Unlimited release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In response to several high-profile ship collisions in 2017, the Surface Warfare Officers School (SWSCCOLCOM, commonly known as SWSC) implemented a program to assess the proficiency of first-tour U.S. Naval Officers of the Deck (OODs). The program has three components: a simulator exercise assessed by a post-command officer, written exams of rules of the road and seamanship knowledge, and a self-reported survey of OOD's operational experience and background. In a continuation of our study of the first round of data collected in 2018, SWSC asked us to analyze the statistical relationship between proficiency, knowledge, and experience from data collected in 2019. They also asked us to make recommendations for how future assessment data can be collected and analyzed in order to inform optimal training and watchstanding policies.

The 2019 data contains a random sample of 66 OODs who were assessed at the end of their first tour. The experience survey revealed large variation in OODs' operational experience, partly stemming from significant variation in the time spent underway. For example, while the median first-tour OOD had 200 hours of experience, OODs in the 5th and 95th percentiles of the distribution had 18 and 855 hours of experience, respectively. 10% of experience was gained in a simulator, and most OODs had no watchstanding experience in the past 90 days at the time of assessment. Assessment scores were normally distributed around "average" proficiency and had an almost identical distribution as scores from the 2018 data collection round. Knowledge is positively correlated with assessed proficiency, but we found no correlation between experience and proficiency.

Ultimately, the small sample size precludes our ability to make precise recommendations about optimal training policies. However, the Surface Community has developed a plan to assess all OODs at multiple points during their careers starting in 2021 and the resulting data, if consistently collected and stored correctly, will facilitate data-driven policy decisions concerning training, proficiency remediation, and officer detailing.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

Ann E. Rondeau
President

Rob Dell
Provost

The report entitled “Measurement and Analysis of Officer of the Deck Competency: New Findings for FY19-20” was prepared for the Surface Warfare Officer Schools Commands Command and funded by the Naval Research Program.

Further distribution of all or part of this report is authorized.

This report was prepared by:

Jesse Cunha
Associate Professor
Graduate School of Defense Management

Vincent Salazar
LT, U.S. Navy

Reviewed by:

Yu-Chu Shen, Associate Dean of Research
Graduate School of Defense Management

Released by:

Jeffrey D. Paduan
Dean of Research

Measurement and Analysis of Officer of the Deck Competency: New Findings for FY19-20

A Technical Report prepared for the Surface Warfare Officer Schools Command and funded by the Naval Research Program

Dr. Jesse Cunha
Graduate School of Defense Management
Naval Postgraduate School

LT Vincent Salazar
U.S. Navy

May, 2020

Extended abstract

In response to several high-profile ship collisions in 2017, the Surface Warfare Officers Schools Command (SWSC) implemented a program to assess the proficiency of first-tour U.S. Naval Officers of the Deck (OODs). The program has three components: a simulator exercise assessed by a post-command officer, written exams of rules of the road and seamanship knowledge, and a self-reported survey of OOD's operational experience and background. In a continuation of our study of the first round of data collected in 2018, SWSC asked us to analyze the statistical relationship between proficiency, knowledge, and experience from data collected in 2019. They also asked us to make recommendations for how future assessment data can be collected and analyzed in order to inform optimal training and watchstanding policies.

The 2019 data contains a random sample of 66 OODs who were assessed at the end of their first tour. The experience survey revealed large variation in OODs' operational experience, partly stemming from significant variation in the time spent underway. For example, while the median first-tour OOD had 200 hours of experience, OODs in the 5th and 95th percentiles of the distribution had 18 and 855 hours of experience, respectively. 10% of experience was gained in a simulator, and most OODs had no watchstanding experience in the past 90 days at the time of assessment. Assessment scores were normally distributed around "average" proficiency and had an almost identical distribution as scores from the 2018 data collection round. Knowledge is positively correlated with assessed proficiency, but we found no correlation between experience and proficiency.

Ultimately, the small sample size precludes our ability to make precise recommendations about optimal training policies. However, the Surface Community has developed a plan to assess all OODs at multiple points during their careers starting in 2021 and the resulting data, if consistently collected and stored correctly, will facilitate data-driven policy decisions concerning training, proficiency remediation, and officer detailing.

Keywords: *Surface Warfare, Officer of the Deck, OOD, proficiency, experience, simulator, training*

***** LT Salazar completed his Naval Postgraduate School Master's Thesis as a part of this research, see Salazar (2020)

1. Executive Summary

Background

Historically, the U.S. Navy has not systematically assessed and recorded mariners' ship handling proficiency or operational experience at the individual level. Absent such information, it is difficult for the Surface Community to gauge the proficiency of their OODs, to track their proficiency over time (both growth through the career and the potential atrophy of skills in-between deployments), and to understand the impact of changes to training and policy. Following several high-profile collisions in the summer of 2017, the Surface Community instituted a series of policies aimed to remedy these shortcomings.

We study a policy which instituted an assessment program with three parts: a simulator exercise in which OODs are assessed by a post-command Commander or Captain, written exams on Rules-of-the-Road (RoR) and Navigation, Seamanship, and Ship-handling (NSS), and a self-reported survey of the OOD's operational experience and background. In 2018, SWSC applied this assessment to a random sample of 164 OODs who were in the middle of their first tours of duty aboard ship. Between April of 2019 and January of 2020, SWSC used the same assessment program to evaluate a sample of 66 OODs at the end of their first tours when they arrived in Newport, R.I. for the Advanced Division Officers School (ADOC).

Several other policies were introduced after the ship collisions with the intent of improving OODs ship handling proficiency, including: increasing formal OOD schoolhouse training from 11 to 20 weeks; increasing the length of OOD's first tours; requiring Surface Warfare Officers (SWSC) to track underway and simulator experience in a Mariner Skills Logbook; mandating that OODs have uninterrupted periods for sleep between watches which coincide with our natural circadian rhythms; and instituting 10 formal assessments of ship handling proficiency throughout an OODs career.

Ultimately, we would want to study the impact of these – and future – policies on OOD proficiency, but the current amount and type of data is not sufficient for such analyses. In fact, such analyses are precisely what were called for in a recent U.S. Government Accountability Office (GAO) report (GAO, 2019). The planned future assessment data, if performed consistently and stored correctly, should facilitate analyses that can inform optimal policies.

Data

The self-reported survey collected demographic and career information (commissioning source, ship class, and home port; age, gender, and prior enlisted status; the number of months spent underway and in-port) and detailed information on officer's overall and recent experience as OOD, Junior Officer of the Deck (JOOD), and Conning Officer (CONN), both in a simulator and underway. One of the recommendations from our FY19 analysis, which SWSC implemented, was to collect more detailed information in this survey. Compared to the original survey instrument, the 2019 version: collected more demographic data (age, gender, and time spent in-port/underway); collected experience data as continuous variables (e.g., the number of months or hours) as opposed to categorical bins; and asked OODs to record their overall experience as well as their experience in the past 90 days, which allows us to study how the recency of experience relates to proficiency.

During the scenario, assessors filled in a rubric containing over 70 assessment points which helped them assign grades in four areas (management of bridge team, leadership, performance under stress, decision making) and one overall assessment category. The areas were graded on a 5-point scale and the overall assessment was on a 7-point scale. The inclusion of subjective sub-

categories and the use of 5- and 7-point scales were recommendations from our FY19 report and provide a finer picture of OOD ability. The written RoR and NSS assessments each contained 20 multiple choice questions and covered standard material an OOD is expected to understand; these tests were unchanged from the 2018 assessment process.

Findings

Our first set of findings concern the self-reported survey data on OODs experience during their first tours. Several findings stand out:

- In general, we found large variation across individuals in their operational experience. For example, while the median first-tour OOD had 200 hours of experience (both on the bridge and in a simulator), OODs in the 5th and 95th percentiles of the distribution had 18 and 855 hours of experience, respectively.
- Similarly, there were a significant proportion of OODs who had not completed any special evolutions as OOD on their first tour. For example, the median OOD completed only one straits transit (the scenario tested in the simulator) and 36 percent of the sample had never completed a straits transit as OOD.
- In the 90 days prior to arriving at ADOC and taking the assessment, the majority of OODs had no operational experience – either underway or in a simulator.

These findings can be partly explained by the finding that the majority of OODs spend over half of their first tour either in a shipyard or in-port. As well, the lack of experience within the past 90 days is due largely to the timing of this assessment during a lengthy schooling pipeline in between an officer's first and second Division Officer tours.

Summarizing the assessment data, we found that the distribution of proficiency was normally distributed (i.e., symmetric with tapering tails) around the average score, both for the sub-categories and the overall assessment. When we transformed the overall assessment categories to match the 3-category scale used in 2018, we find an almost identical distribution of proficiency as in 2018 with about two-thirds of the sample being "average," one-sixth of the sample being "unsatisfactory," and one-sixth being "excellent." Finally, the written exam scores showed that a large fraction of students were not proficient: 30 and 58 percent of the sample "failed" the RoR and NSS exams, respectively.

Next, we performed analyses of the statistical relationship between experience and proficiency. Our main analytical tool was a multivariate regression model, where the outcome (the dependent variable) is the performance of OODs on the various measures of proficiency and the explanatory variables (the independent variables) are the demographic and experience-related variables captured in the survey. A multivariate regression framework is crucial in this context because the explanatory variables are likely to be highly correlated with one another - for example, prior-enlisted officers are generally older, or those in high-traffic home-ports will likely have more experience in dense traffic settings.

We find that proficiency is positively correlated with knowledge (RoR and NSS exam scores, and assessment category scores), but we do not find any significant relationship between proficiency and experience. That this lack of significant correlation is in part due to the small sample size (in contrast, the 2018 dataset with 164 observations showed a positive correlation between proficiency and experience), and possibly also do to the inaccuracy of the self-reported data.

Recommendations

Recent changes to SWO training and assessments as codified in the *Surface Warfare Officer Career Manual* (COMNAVSURFOR, 2019) offer opportunities to use future assessment and experience data to inform policy. In particular, the manual (and instructions preceding it) institute:

- the establishment of 10 assessments throughout an OOD's career which will constitute a population-level, longitudinal database of ship-handling proficiency, and
- the requirement to track experience in the Mariner Skills Logbook which will contain a complete, validated record of OODs' operational experience throughout their careers.

However, in order for this data to be useful for analysis, several issues must be addressed:

- The experience survey should be continually assessed to make sure that it is collecting as detailed data as possible while minimizing reporting errors that result from survey fatigue. In particular, we suggest removing some of the sub-categories in hours of experience and numbers of special evolutions completed.
- Experience data is a necessary component of any future research effort. While we can continue to collect data from surveys at the time of assessment, it may be easier and less subject to recall bias to create a system by which the Mariner Skills Logbook data is routinely entered into an electronic format. Until logbooks are made fully electronic, this function could be performed by having SWSC assessors "audit" the logbooks of officers undergoing OOD assessments and having the "auditor" enter the required experience data.
- The assessment scenario will need to be updated regularly so that officers always see a novel scenario in each assessment. When it is, we must ensure that assessments are reflecting a consistent measure of "proficiency" across cohorts and across different career milestones.
- The assessment scenario should also be evaluated to ensure it is testing the correct set of skills. Would an easier, a harder, a more complex, or a simpler scenario better assess the proficiency of an OOD?
- Assessors must be trained to ensure consistency and comparability across the population. We suggest creating a detailed "Assessment Guide" which codifies how to assign OODs to various proficiency scores, which will be especially useful to prevent inconsistencies when assessment personnel rotate positions.
- Future research and analysis can only be as good as the data that is available. We strongly recommend the creation of a data repository for all future assessments and experience surveys (or Mariners Skills Logbook data) which can track individuals and cohorts over time.

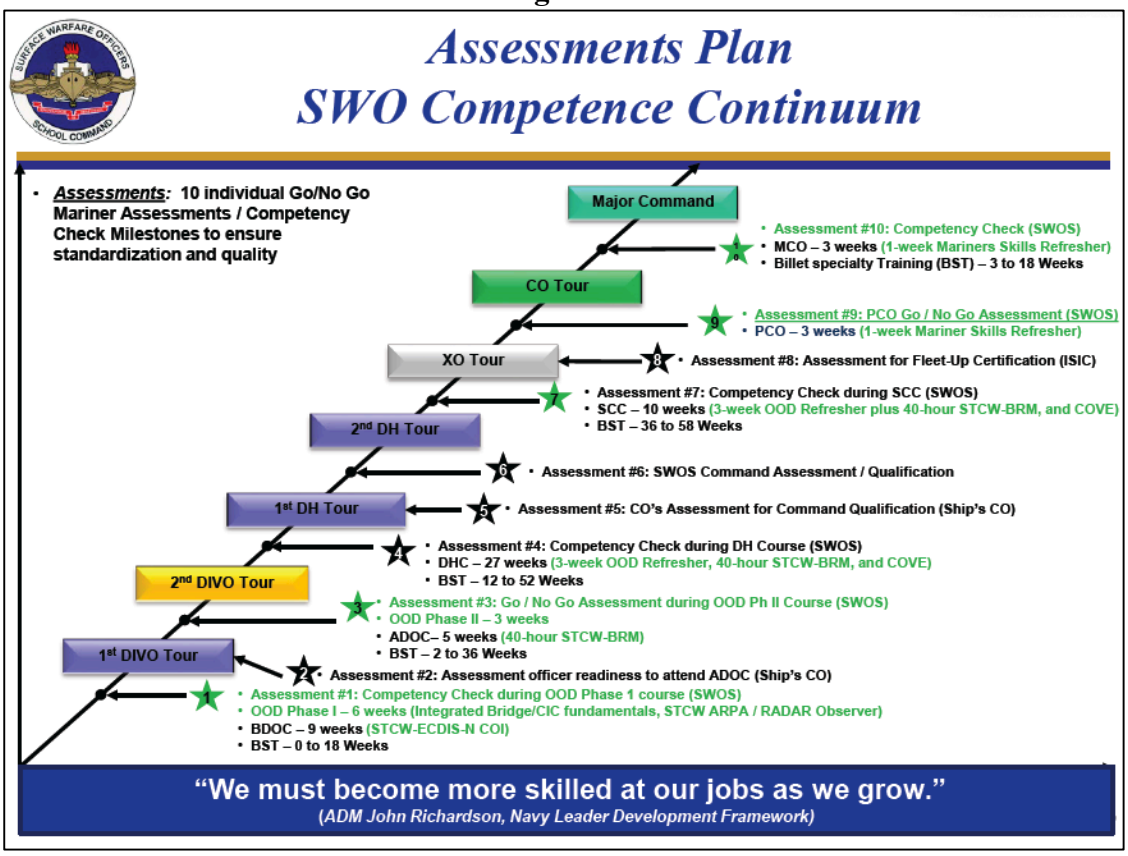
2. Background of recent policy changes

In the summer of 2017, the U.S. Naval Surface Force suffered two major collisions within a short timeframe. The USS *Fitzgerald* and USS *McCain,* both from 7th Fleet, collided with separate

commercial transport ships resulting in severe damage and the deaths of 17 sailors. These incidents sent shockwaves throughout the Navy and resulted in a public call for thorough investigations into the causes of the collisions of these highly capable warships during routine operations at sea.

The resulting investigations led to numerous changes to both the Surface Warfare Community's training and operations and to its assessments of the proficiency of Officers of the Deck. On December 16th, 2019, COMNAVSURFOR issued his newest instruction titled *Surface Warfare Officer Career Manual* (COMNAVSURFOR, 2019) which cancelled five previous instructions which had been issued since the 2017 collisions with the intent to "establish the single SWO community governing document, providing the requirement and milestones from accession through major command" (p. 1). Below, we summarize the major policy changes instituted by this instruction.

Figure 1



This chart identifies the timing of competency assessments and trainings throughout a Surface Warfare Officer's career. Green text identifies new assessments/trainings, black text identifies existing assessments/trainings.

2.1 Increased training

Beginning in 2021, SWSC will receive nine extra weeks of formal training in addition to the 11 weeks of training they received in the past:

- Ensigns will now attend two training courses prior to reporting to their ship: the 9-week Basic Division Officer Course (BDOC) and the 6-week OOD Phase I Course. Previously, ensigns only attended BDOC, and this course itself was slightly shorter.
- Upon completion of their first tour (or prior to fleeting up on the same ship) OODs will now attend a new 3-week OOD Phase II Course in addition to the existing 5-week Advanced Division Officer Course (ADOC) and other second-tour "pipeline" training.

2.2 New proficiency assessments and experience tracking

Officers will have their mariner's skills assessed more frequently during their career and they are now required to record their operational experience as ship handlers.

- Four additional assessments will be made throughout a SWO's career on top of the six existing assessments (see Figure 1). Assessments allow the community to prohibit unfit officers from standing watch, and they also provide the data necessary to track changes in proficiency over time and across the fleet. The first sets of these added assessments will begin in 2021.
- As of 2019, all SWSC are required to record the details of each underway watch, special evolution, and simulator training they have performed in a Mariner Skills Logbook. This physical record facilitates quarterly verifications and end-of-tour summaries by Commanding Officers.

2.3 Changes to tour length and type

Numerous tour changes have been introduced that should allow officers to further improve their skills as ship handlers. These changes will be phased in by commissioning year group:

- The First Division Officer tour has been extended from 24 to 30 months, with an option for a single tour of 48 months instead of two separate tours.
- The Second Division Officer tour has been reduced from 24 to 18 months, and it can no longer be served at an afloat staff.
- Afloat staff tours will now be conducted only by officers who have completed their Division Officer tours.
- Afloat Staff Department Head staff tours will now be conducted by officers who have completed their two shipboard Department Head tours.
- The time between the second Department Head and the Executive Officer tour will be reduced from 5.3 to 4.5 years.

2.4 Instituting minimum requirements to stand key watches

Starting in November of 2018, COMNAVSURFOR put into place minimum currency requirements needed to stand, and requalify for, key watches:

- Officers of the Deck, Tactical Action Officers, and Combat Information Center Watch Officers must now stand a minimum of one watch every 45 days.
- Engineering Officers of the Watch and Combat Systems Officers of the Watch must now stand a minimum of one watch every 30 days.

Such "proficiency watches" may be conducted either underway, in-port, or using a simulator, and the Mariners Skills Logbook will now contain the information needed to verify whether these requirements have been met. If watch standing proficiency lapses, an officer must conduct a "refresh" watch under the supervision of a proficient watchstander or an officer of a higher position.

2.5 Introducing mandatory sleep periods between watches

Reviews of the 2017 incidents identified fatigue - and the culture that surrounds it - as a key contributing factor in the collisions. A November 2017 instruction directed units to implement fixed watches and protected sleep periods which are integral to human natural circadian rhythms. In particular, Sailors must be allowed either an uninterrupted 7-hour period for sleep or a 5-hour period for sleep with a 2-hour nap between watches.

2.6 Summary - The need for more and better data

The goal of these policy changes is to reduce the likelihood of future ship-handling mishaps. While mishaps are clearly observable and measurable events, it is much more difficult to understand how policies are changing mariner's skills and behavior in ways which improve ship-handling.

The analysis we present below demonstrates how data on proficiency and experience can be used to inform policy. However, as we make clear throughout this document, the current set of data is insufficient to address the most pressing policy questions, such as:

- Which training programs work the best to increase proficiency?
- What is the optimal amount of underway versus simulator experience?
- How fast do mariner's skills degrade over time when they are not used?
- What is the functional form of the relationship between experience and proficiency?

A December 2019 Government Accountability Office (GAO) report states: *"The Navy has relied on added skill checks conducted throughout a SWO's career to ensure that each SWO has basic ship-driving skills, but has not put key processes and assessments in place to evaluate comprehensively the effectiveness of its changes to ship-driving training"* (GAO, 2019). More data which tracks the proficiency and experience of OODs is needed, and that data must be collected and stored in a manner that facilitates policy evaluation. This recommendation, and others which concern the specific type of data that is collected, directly address the issue identified by the GAO and will provide the long-term framework from which data-driven policy analysis can improve our Navy's ship-handling capability.

3. OOD proficiency assessments: Methods

In this section, we describe and summarize the OOD proficiency assessment data that was collected by SWSC between April 2019 and January 2020. The assessments had three components: a survey of OODs' experience during their first tour, a test of their skills in a simulator scenario, and written exams on Rules of the Road and Navigation, Seamanship, and Shiphandling. For ease of discussion, we refer to this as the 2019 dataset. Doing so also distinguishes this data from the first round of assessments that were administered in 2018. The 2018 data was analyzed by Cunha and Dearth (2019) and below we describe important changes in the data collection

methods between 2018 and 2019. These changes highlight both differences which are crucial for comparing the databases and issues which should be addressed in future iterations of the assessment scenario, the assessment rubric, and the survey instrument.

The 2019 assessments were conducted on a random sample of 69 end-of-first-tour OODs at the beginning of their ADOC training in Newport, R.I. In contrast, the first round of data contained 164 assessments that were performed on a cross-section of first-tour OODs while on their ships stationed around the globe. Thus, the 2019 data reflects OODs who, on average, have more experience than those assessed in 2018 but also reflects OODs who have been away from their ships for some period of time before being assessed.

3.1 Assessment scenario and grade sheet

The assessment consisted of a simulator exercise in which the officer under evaluation played the role of the OOD and SWSC staff played scripted roles of the Junior Officer of the Deck (JOOD), the Conning Officer (CONN), and the Commanding Officer. The scenario was the same high-traffic straits transit that was used in the 2018 assessments, and we omit details of the scenario in this report in order to prevent future OODs from seeing the assessment.

OODs were graded on their performance by a post-command Commander or Captain who assigned categorical scores in five specific areas (management of bridge team, bridge resource management, leadership, application of rules of the road, and performance under stress) and in overall performance. During the scenario, the assessor filled in a rubric containing 76 assessment questions covering specific actions an officer should or should not have taken. This rubric was intended to aid the assessor in assigning area and overall scores. While almost all of rubric questions are objective (e.g., did the OOD take an action), the final and sub-category scores are purely subjective (i.e., how well did the OOD perform). Fully objective scoring is not ideal for this type of assessment which requires a holistic evaluation of an OOD's ability to safely navigate a ship, but subjective scoring raises concerns about whether different assessors are evaluating OODs consistently. We discuss strategies for mitigating such concerns in our discussion below.

Following the recommendations of Cunha and Dearth (2019), the grade sheet used in the 2019 assessments was updated to allow for more response options. In particular, the scores for the specific categories now allow a five-point response (*excellent, above average, average, below average, unsatisfactory*) up from a four-point response (*exceeds standards, meets standards, requires improvement, unsatisfactory*); and the overall performance now allows a seven-point response (*exceptional, excellent, above average, average, below average, marginal, unsatisfactory*) up from a three-point response (*completed with no concerns, completed with concerns, significant problems*). Increasing the range of possible responses allows evaluators to more finely distinguish between the performance of OODs, and we advocate either the continued use of these scales or the adoption of even finer scales.

3.2 Experience survey

Just prior to taking the simulator assessment, OODs were asked to fill in a paper-based survey of their demographics and operational experience (see Appendix Figure 1). The demographic data includes commissioning source, ship class, ship type, homeport, gender, and age. For operational experience, subjects were asked questions about the number of months spend in-port, in a shipyard, or deployed, and then they were asked to record the hours of operational experience and the

number of special evolutions completed (such as underway replenishment, anchoring, or navigating a Traffic Separation Scheme).

Cunha and Dearth (2019) recommended several changes which were incorporated into the 2019 survey instrument. First, they recommended to collect simulator experience in addition to underway experience, and to collect both overall experience during the first tour and experience in the past 90 days. The literature (see the review in Cunha and Dearth (2019)) has shown that operators' skills in other industries can be improved with simulators but also decay with time; we suspect the same may be true for Navy shiphandlers. Second, they recommended collecting experience data in three separate roles: CONN, JOOD, and OOD. While CONN and JOOD are not in charge of the bridge, experience in those positions, observing the OOD, may be just as valuable as actually being in charge.

Finally, as recommended by Cunha and Dearth (2019) the survey now captures experience variables (months deployed, hours on watch, and numbers of special evolutions) as continuous variables (e.g., the number of months) as opposed to categorical bins (e.g., 0 months, 1-3 months, 3-5 months, etc.). As with the assessment grades, more detailed data allows for a richer summary of the variables and a more precise evaluation of the correlates of proficiency. However, we suspect respondents suffered survey fatigue from the length of the questionnaire, with some skipping experience questions or writing down one number at the top of a column and putting a line indicating "the same" down the column. There is a tension between collecting more data which allows us to study more nuanced questions and collecting less data which is possibly of better quality. We recommend several ways in which we can reduce the number of survey questions without sacrificing the integrity of the dataset.

3.3 Written exams

The third and final dataset consists of scores from two written exams. These exams cover standard material that OODs are expected to know. Both exams evaluate knowledge that has been formally taught and assessed in BDOC and subsequently reinforced in the fleet. SWSC are expected to continue studying RoR on the ship and they are periodically evaluated on them by senior officers. Both exams contained 20 questions, and passing scores were 90% and 80% for the RoR and NSS exams, respectively.

3.4 Summary of the sample

We removed three observations due to data quality issues, leaving a working sample of 66 observations.[1] Table 1 summarizes the demographic variables and confirms that the sample is similar to what we would expect from the population of end-of-first-tour OODs. In particular, 20% commissioned via Officer Candidate School, 48% commissioned via the Reserve Officer Training Corps (ROTC), 30% attended the U.S. Naval Academy, and 2% attended the U.S. Merchant Marine Academy. Officers served on seven different ship classes which were homeported in nine locations around the world. 41% of the sample are female and regardless of gender Division Officers are overwhelmingly young, with only 8% being over the age of 29.

In order to underscore an important point, Table 1 also includes the number of observations in each of the demographic categories: if we are interested in certain categories of officers (say,

[1] One observation was removed because the number of reported hours standing OOD, JOOD and CONN were unrealistically large, another was removed because the officer failed to answer a significant portion of the experience survey questions, and a third observation was removed due to glaring inconsistencies across survey questions in addition to incomplete responses.

Officer Candidate School graduates or those homeported in Japan), the already low sample size of 66 effectively becomes even smaller. In general, a small sample size impedes our ability to confidently extrapolate summary statistics and correlations from this sample to the population of similar OODs. While the ultimate solution to this statistical problem is to perform more assessments, we are able to strengthen our statistical analysis (that is, increase the degrees of freedom) by collapsing variables to fewer categories when we estimate regressions below.[2]

[2] In particular, we code the one officer that attended the Merchant Marine Academy as attending the Naval Academy, we collapse homeports to three categories (East coast U.S., West coast U.S., and overseas), and we collapse ship class to two categories (amphibious ships versus cruisers or destroyers).

Table 1. Summary statistics of the 2019 assessment sample.

	Percentage	Observations
Commissioning source		
Officer Candidate School	20%	13
ROTC	48%	32
US Naval Academy	30%	20
US Merchant Marine Academy	2%	1
Ship class		
Cruiser	15%	10
Destroyer	48%	32
Landing Helicopter Dock (LHD)	12%	8
Landing Helicopter Assault (LHA)	3%	2
Landing Platform Dock (LPD)	8%	5
Amphibious Command Ship (LCC)	2%	1
Landing Ship Dock (LSD)	12%	8
Ship type		
Amphibious ship	36%	24
Crusiser or Destroyer	64%	42
Home port		
Everett, WA	3%	2
Mayport, FL	18%	12
Norfolk, VA	30%	20
Pearl Harbor, HI	5%	3
San Diego, CA	26%	17
Manama, Bahrain	3%	2
Rota, Spain	2%	1
Sasebo, Japan	6%	4
Yokosuka, Japan	8%	5
Home port location		
West Coast U.S.	33%	22
East Coast U.S.	48%	32
Overseas	18%	12
Gender		
Male	59%	39
Female	41%	27
Age (years)		
Less than 25	36%	24
25 through 29	56%	37
Greater than 29	8%	5
Observations		66

Note: Data are from self-reports surveys administered during the 2019/20 round of assessments at the end of OODs' first tours.

4. OOD proficiency assessments: Results

There are three parts to our analysis. First, we present a detailed summary of the experience data that was collected on the survey. This data identifies the variation in experience across OODs and serves as the explanatory, or independent, data in our regression models. Second, we summarize the assessment and exam scores, which are the dependent variables, or outcomes, in our regression analysis. Finally, we estimate multivariate regression models (regressing assessed proficiency on experience) to learn how experience is correlated with proficiency.

All three of these analyses – the variation in experience, the variation in proficiency, and the correlations between them – provide crucial information about the current readiness of the OOD community, and they outline the framework under which we can evaluate the effectiveness of training and manning policy changes once more data is available.

4.1 Summary of OOD experience

Overall experience

Table 2 contains summary statistics of the distributions of overall experience variables (mean, standard deviation, median, 5th percentile, 95th percentile). On average, officers spent 27.8 months aboard their ships during the first tour, of which 11.5 months were spent in a shipyard and just 5 months were spent deployed. The mean of all underway time (underway-not-deployed and deployed) is 10.3 months, which implies that officers could not physically stand an underway bridge watch for over half of their first tour. In addition, Figure 2 and the other summary statistics in Table 1 show that the distribution of underway time is very unequal across officers, with some having no underway time at all and some having considerable underway experience. This distribution in ship status (underway or in-port) is reflected in most of other the experience variables: if a ship is not at sea, it is hard for an officer to gain experience.

Figure 2. Time spend in-port and underway during the first tour.

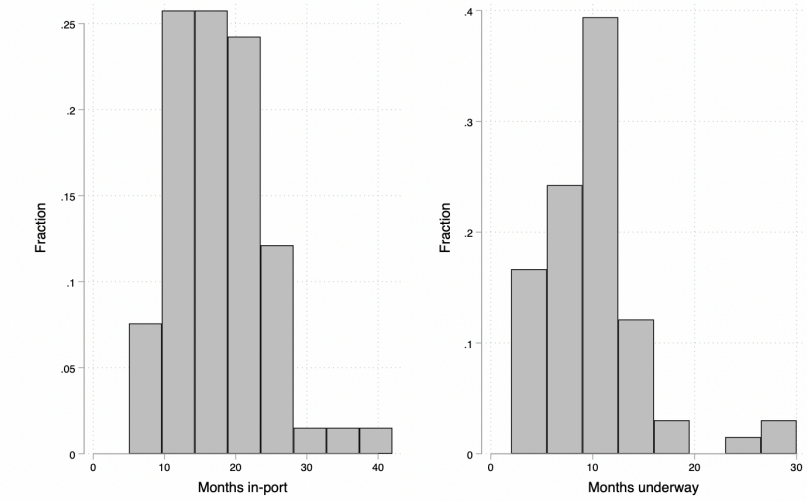


Table 2. OOD watchstanding experience during the first tour.

	Mean	s.d.	Median	5th percentile	95th percentile
Months in position					
aboard ship	27.8	6.2	27	18	34
underway, not deployed	5.3	4.6	5	0	12
deployed	5.0	3.2	6	0	8
in-port, not underway	6.3	4.8	5	0	15
in shipyard	11.5	7.2	10	2	24
Months since attending BRM	10.8	7.0	10.5	1	24
Months taken to qualify OOD	19.5	6.6	19.5	10	30
Months taken to qualify SWO	22.9	6.6	23	14	33
Underway hours in 1st tour					
as CONN	514.7	388.1	450	25	1200
as JOOD	342.0	257.0	300	15	833
as OOD UI	135.5	167.6	100	0	500
as OOD	252.8	255.0	180	5	850
in any position	1245.9	653.6	1212	383	2318
Simulator hours in 1st tour					
as CONN	60.4	163.6	20	0	250
as JOOD	44.6	139.4	5.5	0	200
as OOD	30.0	52.1	10	0	192
in any position	135.0	332.2	40	6	400
Overall hours (underway and simulator)					
as CONN	575.1	450.1	477.5	103	1540
as JOOD	386.5	337.7	300	21	1100
as OOD	283.8	252.8	200	18	855
in any position	1380.9	816.3	1340	414	3319

Note: Data were self-reported on surveys administered during the 2019/20 round of assessments at the end of OODs' first tours.

As expected given a natural career progression, officers report spending more time standing CONN watches, then JOOD watches, and finally OOD watches. The mean total hours of experience on the bridge is 1245, or roughly 31 weeks of full-time (8 hours/day, 5 days/week) experience. Of note – and concern from an operational point of view – the distribution in hours of experience is very wide, with those in the left-tail having very few hours of experience. For example, the 5th percentile of experience as OOD is just 5 hours over the first tour.

Patterns in simulator use are similar to underway experience (most experience as CONN, wide variation across individuals). On average, the 135 hours of simulator experience represents about 10% of the total (simulator plus underway) experience. Considering the large amount of time that OODs spend in-port or in a shipyard, there appears to be scope to significantly increase simulator hours.

Commanding Officers understand that first-tour OODs need to gain watchstanding experience and they likely substitute simulator time for underway time when the latter is not possible. However, even when we look at overall hours both underway and in a simulator, there is a large disparity across officers. This is most easily seen in Figures 3 and 4 which presents histograms of total experience as an OOD and in any position. For example, over 30% of OODs had less than 100 hours of experience, and over 20% had less than 400 hours of total experience.

Figure 3. Total hours of watchstanding experience as OOD in the first tour.

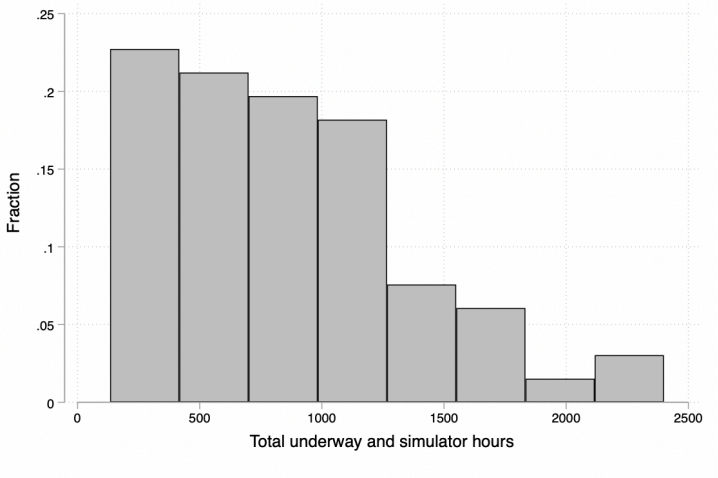
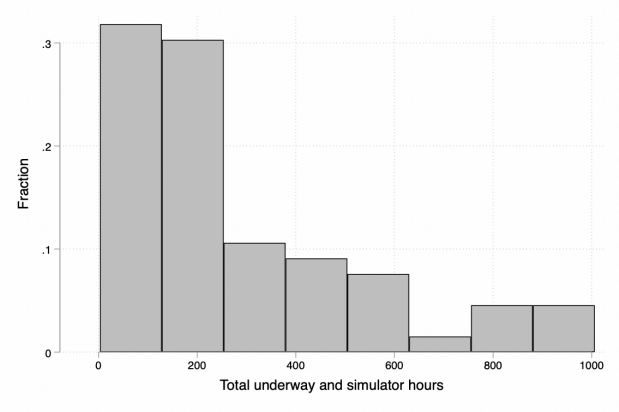


Figure 4. Total hours of watchstanding experience in any position in the first tour.



Recent experience

Table 3 summarizes OODs' watchstanding experience in the past 90 days. On average, officers spent 41 hours underway and 2 hours in a simulator as an OOD in the past 90 days. However, as with overall experience, the average is masking considerable heterogeneity across the sample, with the vast majority of officers having no recent experience whatsoever. Figure 4 demonstrates this point in more detail, showing the full distribution of recent experience (underway and simulator) in any position: almost 80% of officers have 0 hours.

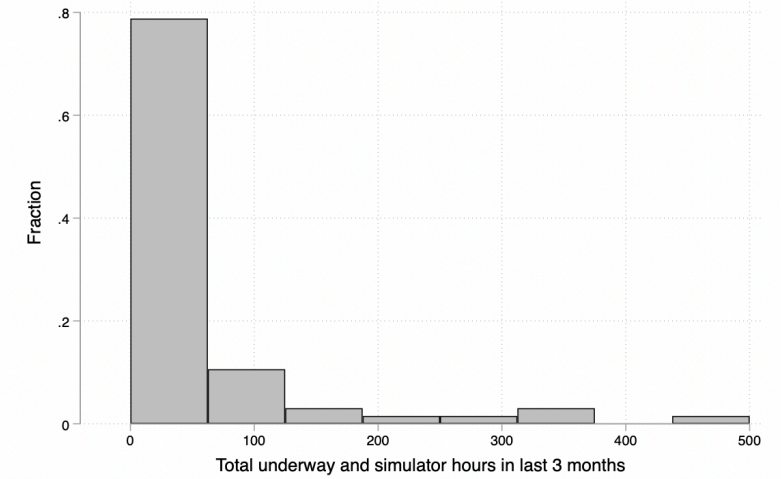
Given that the Surface Community plans to hold career-defining "go/no-go" assessments at fixed times in officers' careers which may come after leaving a deployed ship, it is important to stress that we still do have statistical evidence concerning how shiphandling skills degrade over time. As we discuss in our regression analysis below, more data is required to answer that question empirically.

Table 3. OOD watchstanding experience in the past 90 days.

	Mean	s.d.	Median	5th percentile	95th percentile
Underway hours in past 90 days					
as CONN	3.1	16.3	0	0	130
as JOOD	14.6	49.2	0	0	100
as OOD UI	2.7	13.8	0	0	12
as OOD	41.8	96.5	0	0	264
in any position in past 90 days	62.29	109.33	0	0	330
Simulator hours in past 90 days					
as CONN	0.7	3.4	0	0	2
as JOOD	0.2	1.2	0	0	0
as OOD	2.1	6.2	0	0	10
in any position in past 90 days	2.95	8.41	0	0	18
Overall hours (underway and simulator) in past 90 days					
as CONN	3.8	17.0	0	0	10
as JOOD	14.8	49.2	0	0	100
as OOD	43.9	95.9	1	0	264
in any position	65.2	108.8	3.5	0	338

Note: Data were self-reported on surveys administered during the 2019/20 round of assessments at the end of their first tours.

Figure 4. Watchstanding experience in the past 90 days in any position.



Special Evolutions

The survey asked officers to record how many of five types of special evolutions (underway replenishment, anchoring, entering/exiting port, high-density watches, navigating a Traffic Separation Scheme) were performed over their career and in the past 90 days in various positions. In total, this is 60 data points, and it became clear to us that this was too much detail, possibly leading to inaccurate data which is not particularly necessary for our research questions. Our recommendation for future surveys is to considerably reduce the number variables collected concerning special evolutions.

Table 4 summarizes the total number of special evolutions performed underway and in simulators in various positions. A similar pattern emerges as for hours of experience: the median is lower than the mean and distributions are skewed to the left, with a significant fraction of officers not having performed any of these special evolutions. This is an important consideration: the assessment scenario involves a complicated Straits Transit, yet most officers have never conducted one before. (Appendix Table 1 summarizes specific evolutions completed overall and Appendix Table 2 summarizes evolutions conducted in the past 90 days.)

Table 4. Special evolutions completed in the first tour, either underway or in a simulator.

	Mean	s.d.	Median	5th percentile	95th percentile
Total underway evolutions					
as CONN	34.6	26.9	25	8	81
as JOOD	31.0	66.2	16	1	71
as OOD	25.6	42.4	12	2	88
Total simulator evolutions					
as CONN	19.5	19.3	14	2	71
as JOOD	10.4	16.0	3.5	0	48
as OOD	11.7	14.7	7	0	39
Total underway or simulator evolutions					
as CONN	54.0	34.2	45.5	13	112
as JOOD	41.4	72.6	20.5	7	95
as OOD	37.3	48.1	22.0	4	97
in any position	132.2	133.6	98	33	282

Note: Data were self-reported on surveys administered during the 2019/20 round of assessments at the end of their first tours.

4.2 Examinations and assessments

RoR and NSS Exam Performance

Table 5 summarizes performance on the RoR and NSS exams. The mean RoR test score in 2019 was 90%, which happens to also be the official passing grade. However, there is a large variation, with many students doing very well and many failing: only 72% of officers passed the RoR test (i.e., scored above a 90%). A similar pattern emerges for the 2019 NSS exam: the mean score was 75%, but with a passing grade of 80% only 42.4% of officers passed.

Table 5 also summarizes test scores from the 2018 data collection round, as analyzed in Cunha and Dearth (2018). It shows that mean test scores and passing rates are not too dissimilar from one another. (While not shown, mean test scores across years are also not statistically different from one another, mainly due to the small sample sizes.) The salient differences between the data collection rounds are that: on average the 2019 sample had more experience (because all officers are at the end of their first tour), the 2019 sample had less recent experience (due to finishing their tours and being in a training pipeline), and the 2019 officers had an additional year or more working under the new post-collision environment of heightened operational scrutiny. Given these differences, we do not believe it is possible to draw meaningful conclusions about differences in scores between the collection rounds. Such overtime comparisons over time should be feasible in the future when assessments are conducted at regular intervals on the population of OODs, so long as we ensure that assessments are conducted in a systematic and comparable manner - one of the key recommendations of this study.

Table 5. Comparison of 2019 and 2018 OOD proficiency assessment performance.

2019	Mean	s.d.	min	max	% pass
Rules of the Road test	0.90	0.08	0.70	1	69.7%
NSS test	0.75	0.12	0.55	0.95	42.4%

2018	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: 90% is a passing score on the Rules of the Road test; 80% is a passing score on the NSS test. The 2018 data is from a random sample of 164 OODs while on their first tour (Cunha and Dearth, 2019). The 2019 data is from a random sample of 66 OODs at the end of their first tour.

Assessment Performance

Table 6 displays officer's assessed performance on the simulator exercise, overall and in the four specific areas. It is clear that assessments are normally distributed – that is, the distributions are bell-shaped with a large mass around "average" and smaller symmetric masses in the upper and lower tails. As we demonstrate in the regression analysis below, there is a very strong correlation between performance in the specific categories and overall performance; indeed, a correlation is to be expected as the assessment rubric guides instructors to use performance in the categories for the overall score. We also include the number of observations below percentages for overall performance to highlight how few observations there are in some of the categories. For example, only one OOD (out of 66) scored an "excellent" and only four scored "unsatisfactory."

While this data is informative of OOD performance, there are many unanswered questions about what this distribution of proficiency means for the surface community:

- Are assessors assigning scores in a consistent manner? Assessors likely have similar views on what makes a proficient, competent OOD, but subtle variations in those views can manifest in shifted distributions of scores and obscure true difference in OOD ability.
- What level of performance constitutes enough "proficiency" to confidently serve as an OOD? The scenario was particularly difficult and – as evidenced from the survey – a large majority of officers had either never conducted a straits transit, or had not done one recently.
- How should we think about differences between scores, such as comparing "average" to "above average" or "unsatisfactory" to "marginal"? Or do we only care about a binary assessment which can inform a "go/no-go" decision for career advancement?

Future assessments and associated policies must address these issues.

Table 6. OOD proficiency assessment performance.

Assessment categories	Unsatisfactory	Marginal	Below average	Average	Above average	Exceptional	Excellent
Management of bridge team	7.6%		22.7%	37.9%	25.8%		6.0%
Formality / Presence / Leadership	7.6%		24.2%	39.4%	24.2%		4.5%
Decision making	10.6%		28.8%	25.8%	30.3%		4.5%
Performance under stress	9.1%		34.8%	24.2%	22.7%		7.6%
Overall performance	6%	14%	23%	32%	12%	12%	2%
Number of observations	4	9	15	21	8	8	1

Note: Data are from the 2019/20 assessments conducted at the end of OOD's first tour.

Table 7 compares the 2019 and 2018 overall assessment scores. When we collapse the 2019 data from seven to three categories to match the 2018 data, the score distributions are remarkably similar, with about two-thirds of the samples scoring in the middle ("complete with concerns" or "below average, average, or above average") and one-sixth of the samples scoring in each of the tails. As with the comparison of RoR and NSS exam scores across survey rounds, we caution using this data to make inferences about the effect of policy changes between 2018 and 2019 due to differences in the samples and our inability to abstract from contemporaneous factors which could have influenced proficiency.

Table 7. Comparison of 2019 and 2018 OOD proficiency assessments.

	% Significant concerns	% Complete with concerns	% Complete no concerns
2019 Overall Performance	20%	67%	14%

	% Unsatisfactory and % Marginal	% Average, Below and Above Average	% Exceptional and % Excellent
2018 Overall Performance	18%	66%	16%

Note: 2019 data are from the 2019/20 assessments conducted at the end of OOD's first tour; 2018 data is from the 2018 assessments conducted during OOD's first tours (Cunha and Dearth, 2019).

4.3 Correlates of Proficiency

Methodology

Our final set of analyses estimate the statistical relationships between experience, knowledge, and proficiency using multivariate regression models. The outcome is the overall subjective assessment of OODs and the explanatory variables are the demographic and experience-related variables captured in the survey, the NSS and RoR exam scores, and the subjective assessment category scores. A multivariate regression framework is crucial in this context because the explanatory variables are likely to be highly correlated with one another—for example, prior-enlisted officers are generally older, or those in high-traffic home-ports will likely have more experience in dense traffic settings. A multivariate regression model also allows us to estimate

partial correlations between independent variables and OOD proficiency (for example, the partial correlation of commissioning source with proficiency), which are the statistics that identify the impacts of specific policies.

Our data is particularly rich, containing multiple measures of proficiency and experience which were collected as either continuous variables or categorical variables with many possible responses. While detailed data such as this is generally preferable, it poses problems for regression analysis when combined with the small sample size (N=66). We mention this because the tension between detailed data and a small sample guides our choice of empirical regression models. In particular, we reduce the variability of most proficiency and experience measures by transforming them into binary indicators, combining certain variables, and only including explanatory variables that are necessary to demonstrate the relevant statistical relationships. While not included in this report, we have thoroughly checked that our conclusions are robust to different modelling choices.

Regression Results

We report models that use two parameterizations of the overall assessment score: a binary indicator of scoring "average or above" with a score of three or more on the 7-point scale, and the linear score ranging from one to seven. Due to the small sample size and large number of covariates, it is not feasible to include all covariates in one regression model. However, all of our reported models include fixed effects for individual instructors. While only some of the instructor fixed effects are statistically significant, they are generally large in magnitude and demonstrate the substantial variation in mean assessment scores across assessors. Assessors were effectively randomly assigned to subjects, and while these large magnitudes could be reflecting natural variation across subjects, we believe that the main source of variation is differences in assessors' beliefs of what constitutes proficient ship handling. The large instructor fixed effects reiterate our recommendation to train assessors to assign scores in a consistent manner based on objective criteria.

Table 8 first presents regression results from models which contain, in order, area category scores, RoR and NSS exam scores, and demographics. Columns 1 and 4 of Table 8 include only indicators of receiving a score of "average or above" (>=3) in the assessment categories. All coefficients are large and many are statistically significant, confirming that the overall assessment is highly correlated with subcategory scores. The largest coefficients are for the categories of "performance under stress" and "decision making," attributes that Commanding Officers look for when qualifying and trusting an OOD. For example, the coefficient of 0.514 for "Average or above decision making" in Column 1 implies that the likelihood of receiving a 3 or higher on the overall assessment is 51.4 percentage points higher if the OOD received a 3 or higher in the decision making category.

Columns 2 and 5 include only the RoR and NSS exam scores. Only the NSS exam appears to be strongly correlated with the subjective assessment score, which suggests that the material covered in the NSS test is more closely aligned with the characteristics and abilities that assessors believe make a proficient OOD. Columns 3 and 6 include only demographic variables. Few coefficients are statistically significant, but the magnitudes of some are large. In particular, OCS graduates outperform Naval Academy and ROTC graduates, West Coast homeported OODs perform the best, males perform better than females, and younger OODs perform better than older OODs.

Table 8. Regression estimates, the correlation between proficiency and knowledge/ demographics.

	Average or above (score >= 3)			Assessment score (scale of 1-7)		
	(1)	(2)	(3)	(4)	(5)	(6)
Assessment categories						
Average or above management of bridge	0.117			0.437		
	(0.108)			(0.320)		
Average or above leadership	0.219**			0.391		
	(0.098)			(0.290)		
Average or above performance under stress	0.227***			1.055***		
	(0.080)			(0.239)		
Average or above decision making	0.514***			1.146***		
	(0.090)			(0.269)		
Exam performance						
Rules of the Road score		0.69			1.44	
		(0.88)			(2.49)	
Navigation, Seamanship & Ship handling score		1.22**			3.81**	
		(0.59)			(1.66)	
Platform						
Amphibious ship			-0.049			-0.167
			(0.147)			(0.395)
Commissioning source						
U.S. Naval Academy			-0.194			0.020
			(0.228)			(0.613)
Naval ROTC			-0.129			-0.140
			(0.206)			(0.553)
Home port location						
Forward deployed			-0.115			-0.465
			(0.210)			(0.563)
West coast			0.079			0.106
			(0.170)			(0.456)
Gender						
Female			-0.202			-1.145***
			(0.154)			(0.413)
Age						
25 to 29 years old			-0.118			-0.448
			(0.164)			(0.441)
Greater than 29 years old			-0.362			-0.534
			(0.297)			(0.799)
OOD assessor fixed effects						
Assessor 1	-0.084	-0.30	-0.417	-0.006	-0.68	-0.845
	(0.129)	(0.25)	(0.322)	(0.383)	(0.71)	(0.864)
Assessor 2	-0.056	-0.05	-0.239	-0.050	-0.08	-0.548
	(0.108)	(0.22)	(0.230)	(0.321)	(0.61)	(0.619)
Assessor 3	-0.029	0.01	-0.165	-0.134	-0.07	-0.440
	(0.093)	(0.20)	(0.203)	(0.276)	(0.56)	(0.546)
Assessor 4	0.083	0.23	0.364	0.433	0.62	1.153
	(0.196)	(0.38)	(0.470)	(0.581)	(1.07)	(1.264)
Assessor 5	-0.101	0.08	-0.072	-0.276	0.09	-0.162
	(0.200)	(0.40)	(0.453)	(0.594)	(1.14)	(1.216)
Assessor 6	-0.066	0.13	0.155	-0.935**	-0.46	0.037
	(0.150)	(0.29)	(0.333)	(0.445)	(0.82)	(0.894)
Assessor 7	-0.236	-0.05	-0.225	-1.349***	-0.74	-0.974
	(0.147)	(0.30)	(0.333)	(0.438)	(0.84)	(0.894)
R-squared	0.794	0.16	0.167	0.777	0.17	0.260

*** p<0.01, ** p<0.05, * p<0.1 N=66. Reference groups: OCS commissioning source; CRUDES ship types; age under 25 years; assessor 8. CRUDES ships include destroyers, cruisers. Forward deployed ships are in Rota, Bahrain or Japan. West Coast ships include San Diego CA, Everett WA or Pearl Harbor HI.

Table 9 presents the correlations between experience and proficiency, controlling for demographics but not the exam scores or the assessment categories (as these can be considered outcomes). To further preserve degrees of freedom, we estimate separate models for different measures of experience. Columns 1 and 5 consider overall OOD watchstanding hours, underway or in a simulator; columns 2 and 6 include overall OOD underway and simulator hours separately; columns 3 and 7 include all watch hours in any position, whether underway or in a simulator; and columns 4 and 8 again separate out underway and simulator hours. Across all of the models, even with our most parsimonious specifications, there is no significant correlation between experience and proficiency and coefficients are generally small. This lack of correlation could reflect a lack of correlation in the population, but we strongly suspect that such a correlation does exist yet we are not able to detect it because of the small size of the sample. For example, Cunha and Dearth (2019) estimated similar models using the 2018 proficiency assessment data which had 2.5 times the sample size (164 vs 66) and found statistically significant correlations. The causal relationship between experience and proficiency is clearly an important policy parameter for the Surface Warfare Community and something we hope to study in the future as more data becomes available.

Table 9. Regression estimates, the correlation between proficiency and experience.

	Average or above (score >= 3)				Assessment score (scale of 1-7)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Overall Experience								
Above median OOD hours, underway or simulator	0.007				0.590			
	(0.141)				(0.392)			
Above median OOD hours, underway		-0.050				0.513		
		(0.150)				(0.422)		
Above median OOD hours, simulator		-0.076				-0.139		
		(0.135)				(0.377)		
Above median watch hours, underway or simulator in any position			-0.017				0.192	
			(0.144)				(0.409)	
Above median watch hours, underway in any position				-0.090				0.224
				(0.145)				(0.413)
Above median watch hours, simulator in any position				-0.026				0.087
				(0.141)				(0.402)
Covariates included	yes	yes	yes	yes	yes	yes	yes	yes
Assessor FE	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.112	0.119	0.119	0.112	0.155	0.147	0.124	0.122

*** p<0.01, ** p<0.05, * p<0.1 N=66. Covariates include amphibious ships, U.S. Naval Academy, NROTC, forward deployed ships, West Coast ships, age categories, and gender.

4. Discussion and Recommendations

This report has documented the continuation of our research efforts in 2019 to analyze data collected by SWSC on the proficiency and experience of first-tour OODs. The 2019 data included 66 assessments, NSS and RoR exams, and self-reported experience surveys. Our main findings include:

- On average, only about 40% (10 out of 27 months) of OOD's first tours were spent underway, resulting in significant periods of time during which they were not able to gain operational watchstanding experience other than in simulators. In addition, there was considerable variation in underway time across OODs, with about 40% of the sample having less than 10 months underway experience.

- Resulting partly from the amount of time spent underway, there was large variation in the number of hours OODs spent on the bridge with a significant fraction of OODs having very few hours.
- Simulator hours account for only about 10% of total hours of experience (simulator plus underway). Combined with the amount of time spent in-port, we believe there is scope to increase simulator training.
- At the end of their first tour, the majority of OODs had no underway or simulator experience in the past 90 days. Given that we know skills degrade over time, this lack of recent experience is an important consideration when interpreting measured OOD proficiency in the assessment exercise.
- Subjective proficiency assessment scores were normally distributed around "average," and were almost identical to the distribution of assessments which were conducted in 2018 on a cross-section of first-tour OODs.
- Proficiency is positively correlated with knowledge (that is, RoR and NSS exam scores and assessment category scores), but we do not find any significant relationship between proficiency and experience. We believe that this lack of correlation is in part due to the small sample size.

Recent changes to SWO training and assessments as codified in the *Surface Warfare Officer Career Manual* (COMNAVSURFOR, 2019) offer opportunities to use future assessment and experience data to inform policy. In particular, the manual (and instructions preceding it) institute:

- the establishment of 10 assessments throughout an OOD's career which will constitute a population-level, longitudinal database of ship-handling proficiency, and
- the requirement to track experience in the Mariner Skills Logbook which will contain a complete, validated record of OODs' operational experience throughout their careers.

However, in order for this data to be useful for analysis, several issues must be addressed:

- The experience survey should be continually assessed to make sure that it is collecting as detailed data as possible while minimizing reporting errors that result from survey fatigue. In particular, we suggest removing some of the sub-categories in hours of experience and numbers of special evolutions completed.
- Experience data is a necessary component of any future research effort. While we can continue to collect data from surveys at the time of assessment, it may be easier and less subject to recall bias to create a system by which the Mariner Skills Logbook data is routinely entered into an electronic format. Until logbooks are made fully electronic, this function could be performed by having SWSC assessors "audit" the logbooks of officers undergoing OOD assessments and having the "auditor" enter the required experience data.
- The assessment scenario will need to be updated regularly so that officers always see a novel scenario in each assessment. When it is, we must ensure that assessments are reflecting a consistent measure of "proficiency" across cohorts and across different career milestones.
- The assessment scenario should also be evaluated to ensure it is testing the correct set of skills. Would an easier, a harder, a more complex, or a simpler scenario better assess the proficiency of an OOD?

- Assessors must be trained to ensure consistency and comparability across the population. We suggest creating a detailed "Assessment Guide" which codifies how to assign OODs to various proficiency scores, which will be especially useful to prevent inconsistencies when assessment personnel rotate positions.
- Future research and analysis can only be as good as the data that is available. We strongly recommend the creation of a data repository for all future assessments and experience surveys (or Mariners Skills Logbook data) which can track individuals and cohorts over time.

References

Commander Naval Surface Forces (COMNAVSURFOR). (2019, Dec 16) *Surface Warfare Officer Career Manual* (COMNAVSURFORINST 1412.7). Department of the Navy.

Cunha, Jesse and Robert Dearth. (2019) "Measurement and Analysis of Officer of the Deck Competency." NPS Technical Report NPS-GSDM-20-001.

Salazar, Vincent. (2020) "U.S. Naval Officers of the Deck: Continuing Assessment of the Measurement and Determinants of Proficiency." Naval Postgraduate School Master's Thesis.

U.S. Government Accountability Office (GAO). (2019) *Navy readiness actions needed to evaluate effectiveness of changes to Surface Warfare Officer training*. (GAO-20-154). Government Accountability Office.

Acronyms

Advanced Division Officer Course	ADOC
Basic Division Officer Course	BDOC
Conning Officer	CONN
Junior Officer of the Deck	JOOD
Officer of the Deck	OOD
Surface Warfare Officers School	SWSC
Rules of the Road	RoR
Reserve Officer Training Corps	ROTC
Navigation, Seamanship, and Ship-handling	NSS
U.S. Government Accountability Office	GAO

Appendix Figure 1. The 2019 OOD assessment survey.

Officer of the Deck Survey

Name: _____ **Date:** _____

Home port: _____ **Age:** ____ **Years of Service:** _____

Prior enlisted? Yes / No If yes, rating/MOS: _____ **Sex:** Male / Female

Commissioning Source: Naval Academy / ROTC / OCS / Maritime Academies

If not the Academy, which school? _____
Did you receive ship-handling training in school on a simulator? Yes/No

Which ship did you serve on (e.g., USS Princeton CG-59)? _____

When did you check onboard? (Month/year) _____
When did you check-out of your first ship? (Month/year) _____
When did you complete BDOC? (Month/year) _____

Since you checked onboard, how long did you spend…

In a shipyard or pier-side maintenance availability? _____months
In-port and not underway? _____months
Underway, not on deployment (including any cross-deck periods)? _____months
Underway, on deployment? (including any cross-deck periods)? _____months
**The total sum should equal the length of your first tour.* _____months

When did you qualify OOD? (Month/year) _____

When was your last formal BRM training with the Navigation, Seamanship, and Shiphandling Training (NSST) center? (Month/year) _____

Are you SWO qualified? Yes / No If so, when did you qualify? (Month/year)_____

Bridge Experience

During normal underway steaming, were you in a three-person bridge team (OOD, JOOD, & CONN) or a four-person bridge team (OOD, OOD (U/I), JOOD, & CONN)? 3-person / 4-person

Please provide the number of hours you have spent on the bridge and in simulators in various roles.
(A 4 section watch for 6 months is approximately 1000 hours)

	On the bridge		In a simulator	
	Total career hours	Hours in last 3 months	Total career hours	Hours in last 3 months
As Conning Officer (CONN)				
As Junior Officer of the Deck (JOOD)				
As Officer of the Deck (OOD)				
As OOD Under Instruction (U/I)				

Special evolutions *Please provide the number of evolutions that you have conducted on the bridge and in simulators in various roles.*

As a Conning Officer (CONN):

	On the bridge		In a simulator	
	Total career evolutions	Evolutions in last 3 months	Total career evolutions	Evolutions in last 3 months
Sea & anchor details (Entering/exiting port)				
UNREPs				
Anchorings				
Straits or Traffic Separation Scheme transits				
High-traffic density watches				

As a Junior Officer of the Deck (JOOD):

	On the bridge		In a simulator	
	Total career evolutions	Evolutions in last 3 months	Total career evolutions	Evolutions in last 3 months
Sea & anchor details (Entering/exiting port)				
UNREPs				
Anchorings				
Straits or Traffic Separation Scheme transits				
High-traffic density watches				

As an Officer of the Deck (OOD) or OOD (U/I):

	On the bridge		In a simulator	
	Total career evolutions	Evolutions in last 3 months	Total career evolutions	Evolutions in last 3 months
Sea & anchor details (Entering/exiting port)				
UNREPs				
Anchorings				
Straits or Traffic Separation Scheme transits				
High-traffic density watches				

Appendix Table 1. Special evolutions completed in the first tour, either underway or in a simulator.

	Mean	s.d.	Median	5th percentile	95th percentile
Special Evolutions Underway					
Pierwork					
as CONN	7.9	9.7	5	1	15
as JOOD	9.5	31.1	5	0	20
as OOD	7.5	20.0	4	0	18
Underway Replenishment					
as CONN	8.8	8.3	6	2	30
as JOOD	4.6	7.7	3	0	10
as OOD	3.6	7.6	1.5	0	15
Anchorings					
as CONN	1.2	1.7	1	0	5
as JOOD	1.4	3.5	1	0	4
as OOD	1.3	2.4	0	0	4
TSS and Strait Transits					
as CONN	4.4	5.7	2	0	16
as JOOD	5.6	12.8	2	0	18
as OOD	4.2	10.4	1	0	15
High traffic situations					
as CONN	12.3	15.7	5	0	50
as JOOD	9.9	16.1	5	0	30
as OOD	9.0	15.2	3	0	43
Special Evolutions Simulator					
Pierwork					
as CONN	6.0	6.8	4	0	22
as JOOD	3.2	6.2	1	0	12
as OOD	3.0	4.1	2	0	10
Underway Replenishment					
as CONN	3.9	5.0	2.5	0	15
as JOOD	1.5	3.1	0	0	8
as OOD	2.1	3.6	1	0	10
Anchorings					
as CONN	1.2	2.1	0	0	5
as JOOD	0.5	0.9	0	0	2
as OOD	0.9	1.9	0	0	3
TSS and Strait Transits					
as CONN	4.6	6.7	2	0	24
as JOOD	2.4	3.9	1	0	10
as OOD	2.7	3.8	1	0	10
High traffic situations					
as CONN	3.8	3.8	3	0	10
as JOOD	2.8	4.7	0.5	0	10
as OOD	2.9	3.6	2	0	10

Appendix Table 2. Special evolutions completed in the past 90 days.

	Mean	s.d.	Median	5th percentile	95th percentile
Underway hours in past 90 days					
as CONN	3.1	16.3	0	0	130
as JOOD	14.6	49.2	0	0	100
as OOD UI	2.7	13.8	0	0	12
as OOD	41.8	96.5	0	0	264
in any position in past 90 days	62.29	109.33	0	0	330
Simulator hours in past 90 days					
as CONN	0.7	3.4	0	0	2
as JOOD	0.2	1.2	0	0	0
as OOD	2.1	6.2	0	0	10
in any position in past 90 days	2.95	8.41	0	0	18
Overall hours (underway and simulator) in past 90 days					
as CONN	3.8	17.0	0	0	10
as JOOD	14.8	49.2	0	0	100
as OOD	43.9	95.9	1	0	264
in any position	65.2	108.8	3.5	0	338